PRIFYSGOL ABERYSTWYTH UNIVERSITY

Aberystwyth University

Kernel-based fuzzy-rough nearest-neighbour classification for mammographic risk analysis

Qu, Yanpeng; Shang, Changjing; Shen, Qiang; MacParthalain, Neil; Wu, Wei

Published in:
International Journal of Fuzzy Systems

DOI:
10.1007/s40815-015-0044-1

Publication date:
2015

Citation for published version (APA):
Qu, Y., Shang, C., Shen, Q., MacParthalain, N., & Wu, W. (2015). Kernel-based fuzzy-rough nearest-neighbour classification for mammographic risk analysis. *International Journal of Fuzzy Systems*, *17*(3), 471-483. https://doi.org/10.1007/s40815-015-0044-1

General rights
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Download date: 26. Apr. 2024

Kernel-based Fuzzy-rough Nearest-neighbour Classification for Mammographic Risk Analysis

Yanpeng Qu, Changjing Shang, Qiang Shen, Neil Mac Parthalán and Wei Wu

Abstract

Mammographic risk analysis is an important task for assessing the likelihood of a woman developing breast cancer. It has attracted much attention in recent years as it can be used as an early risk indicator when screening patients. In this paper, a kernel-based fuzzy-rough nearest-neighbour approach to classification is employed to address the issue of the assessment of mammographic risk. Four different breast tissue density assessment metrics are employed to support this study, and the performance of the proposed approach is compared with alternative nearest-neighbour-based classifiers and other popular learning classification techniques. Systematic experimental results show that the work employed here generally improves the classification performance over the others, measured using criteria such as classification accuracy rate, root mean squared error, and the kappa statistics. This demonstrates the potential of kernel-based fuzzy-rough nearest-neighbour classification as a robust and reliable tool for mammographic risk analysis.

Keywords: Mammographic Risk Analysis, Kernel-based Fuzzy-rough Sets, Nearest-neighbour Algorithms, Classification.

1. Introduction

Breast cancer is the most common cancer and the most common cause of death due to cancer amongst women in 140 of 184 countries worldwide, with nearly 1.7 million new cases diagnosed and 522,000 deaths in 2012. This represents about 12% of all new cancer cases and 25% of all cancers in women [1], [2], [3]. There were 6.3 million women alive who had been diagnosed with breast cancer in the previous five years. Since 2008 breast cancer incidence has increased by more than 20%, while mortality has increased by 14%. Although increased levels of the occurrence of breast cancer have been recorded, so too has the level of early detection by screening using mammographic imaging and expert opinion. However, even expert radiologists sometimes fail to detect a significant proportion of mammographic abnormalities. A large number of detected abnormalities are usually discovered to be benign following medical investigation.

Existing mammographic Computer Aided Diagnosis (CAD) systems [4] concentrate on the detection and classification of mammographic abnormalities. As breast tissue density increases however, the effectiveness of these systems in detecting such abnormalities is considerably reduced. In [5], it is claimed that mammographic density is the most indicative risk factor for breast cancer. Also, density estimation can be used to evaluate the likelihood of hidden abnormalities [6]. Techniques for automatic classification to support the consideration of tissue density and the minimisation of human bias when searching for mammographic abnormalities are therefore highly desirable. Given the fact that the estimation of mammographic risk currently remains a major challenge in modern medical science, much research effort is needed in order to improve the accuracy of such classification [7], [8], [9].

Generally, a classification problem can be solved from a variety of perspectives, such as probability theory [10] (e.g., Bayesian networks [11]), function approximation (e.g., SVM [12]), decision tree learning [13] (e.g., C4.5/J48 [14]), rule induction-based classifier (e.g., JRip [15]) and instance-based learning [16] (e.g., k nearest-neighbours or kNN [17]). This paper, based on the initial ideas proposed in [18], presents a kernel-based fuzzy-rough nearest-neighbour approach for instance-based learning. It implements classification tasks using kernel-based fuzzy-rough sets (KFRS), which are a flexible hybridisation of kernel methods [12] and fuzzy-rough sets [19]. Similar to other kernel-based methods [20] the proposed approach addresses the classification problem by exploiting the properties of the

Corresponding Author: Qiang Shen is with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Wales, UK.
Email: qqs@aber.ac.uk
Yanpeng Qu is with the Information Science and Technology College, Dalian Maritime University, Dalian 116026, P. R. China.
Email: yanpengqu@dlmu.edu.cn
Changjing Shang and Neil Mac Parthalán are with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Wales, UK.
Email: {cns, ncm}@aber.ac.uk
Wei Wu is with the School of Mathematical Sciences, Dalian University of Technology, Dalian, China.
Email: wuweiw@dlut.edu.cn

© 2004 TFSA (*only used in first page*)

kernel functions. Experimental evaluations show that the algorithms established following this approach perform well on benchmark datasets, in application to mammographic risk analysis.

In order to have a comprehensive analysis, four different breast tissue density assessment metrics: Wolfe [21], Boyd [22], Tabár [23], and BI-RADS [24] are employed. To evaluate the performance of the kernel-based fuzzy-rough nearest-neighbour algorithms, comparative studies with alternative nearest-neighbour-based classifiers and other popular learning classification techniques are carried out, through systematic experimental investigation. The results, which are measured using criteria such as classification accuracy rate, root mean squared error, and the kappa statistic, demonstrate that the kernel-based fuzzy-rough nearest-neighbour approach offers improved and robust performance over others.

The remainder of this paper is structured as follows. The theoretical foundation of the kernel-based fuzzy-rough nearest-neighbour algorithm is described in Section 2. The kernel-based fuzzy-rough sets and kernel-based fuzzy-rough nearest-neighbour approach are then presented in Section 3. The data used and its labelling schemes for the experimental evaluation are described in Section 4, and the experimental results are discussed in Section 5. The paper is concluded in Section 6, with an outline of proposed further work.

2. Theoretical Background

A. *Hybridisation of Rough Sets and Fuzzy Sets*

The work on rough set theory (RST) [25] provides a methodology that can be employed to extract knowledge from data in a concise way. It is able to minimise information loss during the extraction process while also limiting the amount of human intervention. Central to rough set theory is the concept of indiscernibility. Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe) and $\mathbb{A}$ is a non-empty finite set of attributes so that $a: \mathbb{U} \longrightarrow V_a$ for every $a \in \mathbb{A}$. V_a is the set of values that attribute a may take. For any $P \subseteq \mathbb{A}$, there exists an associated equivalence relation $IND(P)$:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 | \forall a \in P, a(x) = a(y)\}. \quad (1)$$

The partition generated by $IND(P)$ is denoted $\mathbb{U}/IND(P)$ or abbreviated to $\mathbb{U}/P$ and is calculated as follows:

$$\mathbb{U}/IND(P) = \bigotimes \{a \in P: \mathbb{U}/IND(\{a\})\} \quad (2)$$

where,

$$\mathbb{U}/IND(\{a\}) = \{\{x | a(x) = b,\ x \in \mathbb{U}\} | b \in V_a\} \quad (3)$$

and,

$$A \otimes B = \{X \cap Y | \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\}. \quad (4)$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_P$. Let $X \subseteq \mathbb{U}$. X can be approximated using only the information contained in P by constructing the *P-lower* and *P-upper* approximations of X:

$$\underline{P}X = \{x | [x]_P \subseteq X\} \quad (5)$$

$$\overline{P}X = \{x | [x]_P \cap X \neq \emptyset\}. \quad (6)$$

The tuple $< \underline{P}X, \overline{P}X >$ is called a rough set.

Although useful, rough sets only operate effectively on datasets containing discrete values. As most datasets contain real-valued attributes, a subjective judgement or threshold must therefore be employed in order for RST to operate on such data. The imposition of such a subjective threshold is however, contrary to the concept of domain independence of RST. An appropriate way of handling the problem of real-valued data is the use of fuzzy-rough sets (FRS) [26]. FRS offers a high degree of flexibility in enabling the vagueness and imprecision present in real-valued data to be modelled effectively.

Definitions for the fuzzy lower and upper approximations can be found in [19], [27], where a T-transitive fuzzy similarity relation is used to approximate a fuzzy concept X:

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in \mathbb{U}} I\left(\mu_{R_P}(x, y), \mu_X(y)\right) \quad (7)$$

$$\mu_{\overline{R_P}X}(x) = \sup_{y \in \mathbb{U}} T\left(\mu_{R_P}(x, y), \mu_X(y)\right) \quad (8)$$

Here, I is a fuzzy implicator and T is a T-norm. R_P is the fuzzy similarity relation induced by the subset of features P:

$$\mu_{R_P}(x, y) = T_{a \in P}\{\mu_{R_a}(x, y)\} \quad (9)$$

$\mu_{R_a}(x, y)$ is the degree to which objects x and y are deemed similar with respect to feature a.

Note that formulae (7) and (8) are sensitive to noisy values, as with their crisp counterparts. To tackle this problem, vaguely-quantified rough sets (VQRS) have been introduced in [28]. Following this approach, given a pair of fuzzy quantifiers (Q_u, Q_l), with each quantifier being an increasing $[0,1] \rightarrow [0,1]$ mapping, the lower and upper approximations of X by R can be respectively (re-)defined by

$$\mu_{\underline{R_P}X}^{Q_u}(x) = Q_u\left(\frac{|R_P(x, y) \cap X|}{|R_P(x, y)|}\right) = Q_u\left(\frac{\sum_{y \in \mathbb{U}} min\left(\mu_{R_P}(x, y), \mu_X(y)\right)}{\sum_{y \in \mathbb{U}} \mu_{R_P}(x, y)}\right) \quad (10)$$

$$\mu_{\overline{R_P}X}^{Q_l}(x) = Q_l\left(\frac{|R_P(x, y) \cap X|}{|R_P(x, y)|}\right) = Q_l\left(\frac{\sum_{y \in \mathbb{U}} min\left(\mu_{R_P}(x, y), \mu_X(y)\right)}{\sum_{y \in \mathbb{U}} \mu_{R_P}(x, y)}\right) \quad (11)$$

The fuzzy set intersection is interpreted by the T-norm *min* and the fuzzy set cardinality by the sigma-count op-

eration in this work (although other interpretations may similarly be used in principle). As an important difference to (7) and (8), the VQRS approximations do not directly extend the classical rough set approximations, in a sense that when X and R are crisp, (10) and (11) may still be fuzzy.

B. *Fuzzy-rough Nearest Neighbours*

Initial attempts to combine the fuzzy nearest-neighbour (FNN) [29] algorithm with concepts from fuzzy rough set theory were presented in [30] and improved in [31]. In such work, a fuzzy-rough ownership function is constructed that attempts to handle both fuzzy uncertainty (caused by overlapping classes) and rough uncertainty (caused by insufficient knowledge about the objects). The resultant fuzzy-rough classification algorithm from the improved approach is termed FRNN-O in this paper for easy reference.

It should be noted that this algorithm does not use fuzzy lower or upper approximations to determine class membership. A very preliminary attempt to do so was described in [32]. However, the authors did not state how to use the upper and lower approximations to derive classifications. Also, in [33], a rough-fuzzy weighted k-nearest leader classifier was proposed; however, the concepts of lower and upper approximations were redefined for this purpose and have no overlap with the traditional definitions.

In [34], a fuzzy-rough nearest-neighbour (FRNN) algorithm is proposed. It works by examining each of the decision classes in the training data in turn. It computes the membership of a test object to the fuzzy lower and upper approximations of each class. These values are then compared with the highest existing values: $\mu_l(y)$ and $\mu_u(y)$. If the approximation membership values for the currently considered class are higher, then $\mu_l(y)$ and $\mu_u(y)$ are assigned these values and the class label is assigned to this test object. If not, the algorithm continues to iterate through all remaining decision classes. Classification accuracy is calculated by comparing the output with the actual class labels of the test objects.

An extension of FRNN is vaguely quantified nearest neighbour (VQNN) [28] which employs (10) and (11) in order to determine class membership of test objects. The underlying learning mechanism is very similar to that of FRNN and hence omitted here.

C. *Classes of Kernels in Statistics*

In a kernel-based algorithm, a mapping ϕ from a given data space on to a possibly high-dimensional space is employed to change the distribution of the data from the original nonlinear problem to a linearly separable problem. By replacing the inner product with an appropriate kernel function, a nonlinear mapping can be implicitly performed on a high dimensional feature space without increasing the number of parameters. Consider the case of mapping an n-dimensional feature space to an m-dimensional feature space:

$$\phi: x \to \phi(x), \quad x \in \mathbb{R}^n, \phi(x) \in \mathbb{R}^m. \tag{12}$$

A kernel denotes a function K such that for all $x, y \in \mathbb{R}^n$:

$$K(x, y) = \phi(x) \cdot \phi(y). \tag{13}$$

In statistics, symmetric positive definite functions are termed covariances. Hence, kernels are covariance-based in essence. From a statistics perspective, generally, two important classes of kernels are: stationary kernels and non-stationary kernels [35]. The work in this paper focuses on stationary kernels.

Stationary kernels $K(x, y) = K_S(x - y)$ do not depend on the data object values themselves, but only on the lag vector separating the two objects x and y. Isotropic stationary kernels, which depend only on the norm of the lag vector, are most commonly used. For isotropic stationary kernels, the covariance form is:

$$K_{cov}(x, y) = K_I(\| x - y \|), \tag{14}$$

and the correlation form is

$$K_{cor}(x, y) = K_I(\| x - y \|)/K_I(0). \tag{15}$$

A *non-stationary* kernel $K(x, y)$ is one which depends explicitly on the two data objects x and y. Note that a special kind of non-stationary kernel, called a *reducible kernel*, can be reduced to a stationary kernel.

3. Kernel-based Fuzzy-rough Nearest-neighbour Classification

A. *Need for Kernel-based Fuzzy-rough Sets*

The relationship between T-transitivity and kernels has been explored recently [36]. It has been shown that any kernel: $X \times X \to [0,1]$, $K(x, x) = 1$, $\forall x \in X$, is T_{cos} -transitive, where $T_{cos}(a, b) = max(ab - \sqrt{1 - a^2}\sqrt{1 - b^2}, 0)$. As an initial attempt, kernelised fuzzy-rough sets, which combine kernel methods with concepts from fuzzy-rough set theory, have also been proposed [37]. In this approach, kernels $K(x, y)$ are constrained such that they impose: a) reflexivity, b) symmetry, and c) T_{cos}-transitivity. Such kernels are employed to calculate the degree to which objects x and y are similar for every feature. The fuzzy lower and upper approximations in kernelised fuzzy-rough sets are defined by:

$$\mu_{\underline{K_P X}}(x) = \inf_{y \in \mathbb{U}} I_{cos}\left(K_P(x, y), \mu_X(y)\right) \tag{16}$$

$$\mu_{\overline{K_P X}}(x) = \sup_{y \in \mathbb{U}} T_{cos}\left(K_P(x, y), \mu_X(y)\right). \tag{17}$$

where, the implicator

$$I_{cos} = \begin{cases} 1, & a \le b \\ ab + \sqrt{(1 - a^2)(1 - b^2)}, & a > b \end{cases}. \tag{18}$$

In general, however, for fuzzy-rough sets, T-transitivity is not necessarily displayed, and fuzzy tolerance relations may be sufficient [38]. Yet, as reflected by (9), the fuzzy similarity relation induced by the subset of features P is computed through the combination that is implemented by a T-norm. Specifically, for kernelised fuzzy-rough sets, it is:

$$K_P(x,y) = T_{a\in P}\{\mu_{R_a}(x,y)\}. \tag{19}$$

In this case, the choice of a kernel function becomes restricted. This is due to the fact that not many kernel functions can be denoted by a T-norm-based combination of reflexive functions. The Gaussian kernel employed in [37] is workable, because for: $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,\ y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$

$$\exp\left(-\frac{\|x-y\|^2}{\theta}\right) = \prod_{i=1}^{n} \exp\left(-\frac{(x_i-y_i)^2}{\theta}\right) \tag{20}$$

and because its product is still a T-norm. However, for most kernels, such as the rational quadratic kernel and the wave kernel (see below), this property may not hold. In order to address these problems, kernel-based fuzzy-rough sets are proposed in this paper as follows.

B. *Kernel-based Fuzzy-rough Sets*

In geometry, the inner product of two vectors is the projection of one onto another. Indeed, the square of the norm distance in a Hilbert space can be expressed by the inner product. In this case, the inner product can measure the similarity between the images of two features by mapping them onto the Hilbert space. Therefore, given a non-empty set $\mathbb{U}$ and a kernel function K being reflexive (that is $K(x,x) = 1$), for an arbitrary fuzzy concept X, the lower and upper approximations of a kernel-based fuzzy-rough set can be defined as:

$$\mu_{\underline{R_P^K X}}(x) = \inf_{y\in\mathbb{U}} I\left(\mu_{K_P}(x,y), \mu_X(y)\right) \tag{21}$$

$$\mu_{\overline{R_P^K X}}(x) = \sup_{y\in\mathbb{U}} T\left(\mu_{K_P}(x,y), \mu_X(y)\right). \tag{22}$$

It is important to note that the general framework of fuzzy-rough sets remains intact using the definition described in this paper. Indeed, the kernel methods play a special role in calculating the fuzzy tolerance relations. It is because of this fact that the term *kernel-based fuzzy-rough sets* (KFRS) is employed here rather than *kernelised fuzzy-rough sets*.

As well as fuzzy-rough sets, the corresponding lower and upper approximations of the kernel-based vaguely quantified rough set (KVQRS) can also be defined such that

$$\mu_{\underline{R_P^K X}}^{Q_u}(x) = Q_u\left(\frac{|R_P^K(x,y)\cap X|}{|R_P^K(x,y)|}\right) = Q_u\left(\frac{\sum_{y\in\mathbb{U}} min\left(\mu_{K_P}(x,y), \mu_X(y)\right)}{\sum_{y\in\mathbb{U}} \mu_{K_P}(x,y)}\right) \tag{23}$$

$$\mu_{\overline{R_P^K X}}^{Q_l}(x) = Q_l\left(\frac{|R_P^K(x,y)\cap X|}{|R_P^K(x,y)|}\right) = Q_l\left(\frac{\sum_{y\in\mathbb{U}} min\left(\mu_{K_P}(x,y), \mu_X(y)\right)}{\sum_{y\in\mathbb{U}} \mu_{K_P}(x,y)}\right) \tag{24}$$

where, $\mu_{K_P}(x,y)$ is induced by the subset of features P and kernel function K:

$$\begin{aligned}\mu_{K_P}(x,y) &= T_{a\in P}\{\phi(a(x))\cdot\phi(a(y))\}\\ &= T_{a\in P}\{K(a(x),a(y))\}\\ &= T_{a\in P}\{K_a(x,y)\}.\end{aligned} \tag{25}$$

As indicated previously, all isotropic stationary kernels in the correlation form of (15) are suitable for being integrated into such KFRS. A collection of certain commonly used isotropic stationary kernels in this correlation form are listed as follows:

- Gaussian kernel: $K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\theta}\right)$
- Exponential kernel: $K(x,y) = \exp\left(-\frac{\|x-y\|}{\theta}\right)$
- Rational quadratic kernel: $K(x,y) = 1 - \frac{\|x-y\|^2}{\|x-y\|^2+\theta}$
- Wave kernel: $K(x,y) = \frac{\theta}{\|x-y\|}\sin\left(\frac{\|x-y\|}{\theta}\right)$.

Note that for specific non-stationary kernels [35], the reflexivity holds also. Thus, such non-stationary kernel functions are also available for constructing KFRS. For instance, the following non-stationary kernel is reflexive:

$$K(x,y) = \frac{\|x\| + \|y\| - \|x-y\|}{2\sqrt{\|x\|\|y\|}}. \tag{26}$$

C. *Kernel-based Fuzzy-rough Nearest-neighbour Algorithms*

Having introduced kernel-based fuzzy-rough sets, a combination of kernel methods with the conventional fuzzy-rough nearest neighbour approaches [30], [32] can be readily established. This is straightforward as with the work of [34]. The resulting combined learning algorithm is termed kernel-based fuzzy-rough nearest-neighbour algorithm (KFRNN), and is outlined in Algorithm 1.

The rationale behind this algorithm is basically the same as that adopted by any nearest-neighbour approaches. In KFRNN, the k nearest-neighbours are examined according to the kernel-based fuzzy tolerance relations. The average of the lower and upper approximations of the kernel-based fuzzy-rough sets are employed as the decision qualifier. After an iteration through all the decision classes, the class label of the test object will be assigned the same as that of the sample which has the highest value of the average of the lower and upper approximations. The complexity of this algorithm is: $O(|\mathcal{C}|\cdot 2|\mathbb{U}|)$.

Algorithm 1 Kernel-based fuzzy-rough nearest-neighbour algorithm

KFRNN($\mathbb{U}$, $\mathcal{C}$, y, k, K)
$\mathbb{U}$: the training set;
$\mathcal{C}$: the set of decision classes;
y: the object to be classified;
k: the number of nearest neighbours;
K: the chosen kernel function.

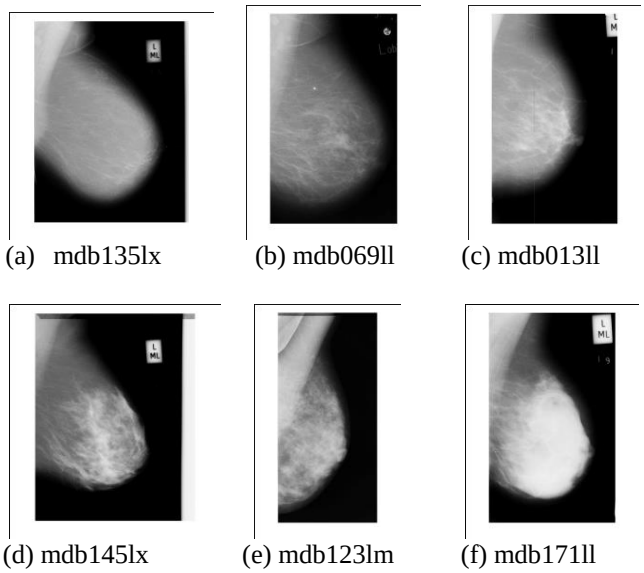
(1) $N \leftarrow$ get Nearest-Neighbour(y, k)
(2) $\tau \leftarrow 0, Class \leftarrow \emptyset$
(3) $\forall X \in \mathcal{C}$
(4) **if** $((\mu_{\underline{R_P^K}X}(y) + \mu_{\overline{R_P^K}X}(y))/2 \geq \tau)$
(5) $Class \leftarrow X$
(6) $\tau \leftarrow (\mu_{\underline{R_P^K}X}(y) + \mu_{\overline{R_P^K}X}(y))/2$
(7) **output** $Class$

It is interesting to note that the accuracy of FRNN is decided only by the greatest similarity between the objects in the training datasets and the test object [39]. Also, for VQNN, the classification only depends on the highest summation of the similarities for each class within the *k* nearest neighbours. As the mechanisms of FRNN and VQNN are respectively retained in KFRNN and KVQNN, such effects of the number of the nearest neighbours can also be observed in the empirical results (as to be presented later). That is, only the single nearest neighbour is needed for classification by KFRNN and the choice of *k* has indeed a major impact upon the classification for KVQNN. Although this can be considered as an advantage with regard to the issue of parameter selection for KFRNN, KVQNN enjoys a more robust performance in the presence of noisy data.

4. Experimental Data

The data employed for the experimental evaluation in this paper is derived from features extracted from images in the Mammographic Image Analysis Society (MIAS) database [40] (see [41] for the feature extraction process). It involves a complete set of Medio-Lateral-Oblique (MLO) left and right mammograms of 161 women (322 objects). Each mammogram object is represented by 280 features, 10 derived from morphological characteristics, and the remaining 270 derived from the extracted texture information. The spatial resolution of the images is $50\mu m \times 50\mu m$ and quantised to 8 bits with a linear optical density in the range 0–3.2.

Mammographic risk assessment metrics which are commonly used are those based on the Wolfe [21], Boyd [22], Tabár [23], or BI-RADS [24] labelling schemes (see Figure 1 for examples). These four metrics can be grouped into two approaches of assessment. Boyd measures the percentage area of dense breast tissue. In contrast, Wolfe, BI-RADS, and Tabár all include patterns and texture information in estimating the classification.



(a) mdb135lx (b) mdb069ll (c) mdb013ll

(d) mdb145lx (e) mdb123lm (f) mdb171ll

Figure 1. Example mammograms, where: (a) SCC: 0%, W: N1, T: Pattern II, B: I (b) SCC: 0 - 10%, W: N1, T: Pattern III, B: I (c) SCC: 11 - 25%, W: P1, T: Pattern III, B: II (d) SCC: 26 - 50%, W: P2, T: Pattern I, B: III (e) SCC: 51 - 75%, W: P2, T: Pattern IV, B: III and (f) SCC: > 75%, W: DY, T: Pattern V, B: IV.

Wolfe proposed four categories of mammographic risk; these four groups have an incidence of developing breast cancer of 0.1, 0.4, 1.7 and 2.2, respectively [21]:

- N1 is defined as a mammogram that is composed mainly of fatty tissue and a few fibrous tissue strands;
- P1 shows a prominent duct pattern, and a beaded appearance can be found either in the subareolar area or the upper axillary quadrant;
- P2 indicates severe involvement of a prominent duct pattern which may occupy from one-half up to all of the volume of the parenchyma, and often the connective tissue hyperplasia produces coalescence of ducts in some areas;
- DY features a general increase in density of the parenchyma (which may be homogeneous) and there may, or may not, be a minor component of prominent ducts.

Boyd et al. [22] introduced a quantitative classification of mammographic densities. It is based on the proportion of dense breast tissue relative to the overall breast area. The classification is known as Six-Class-Categories (SCC) where the density proportions are: Class1: 0%, Class2: $(0\% - 10\%)$, Class3:

$[10\% - 25\%)$, Class4: $[25\% - 50\%)$, Class5: $[50\% - 75\%)$, and Class6: $[75\% - 100\%]$. The increase in the level of breast tissue density has been associated with an increase in the risk of developing breast cancer, specifically the relative risk for SCC 3 to 6 are 1.9, 2.2, 4.6, and 7.1, respectively [22].

Tabár et al. [23] describe breast composition of four building blocks: nodular density, linear density, homogeneous fibrous tissue, and radiolucent adipose tissues which also define mammographic risk classification. In particular, the following patterns are defined, with Patterns I-III corresponding to lower breast cancer risk, and Patterns IV-V relating to higher risk [23]:

- Pattern I: mammograms are composed of 25%, 16%, 35%, and 24% of the four building blocks, respectively;
- Pattern II has approximate compositions as: 2%, 14%, 2%, and 82%;
- Pattern III is quite similar in composition to Pattern II, except that the retroareolar prominent ducts are often associated with periductal fibrosis;
- Pattern IV is dominated by prominent nodular and linear densities, with compositions of 49%, 19%, 15%, and 17%;
- Pattern V is dominated by extensive fibrosis and is composed as 2%, 2%, 89%, and 7% of the building blocks.

For BI-RADS [24] there are four classification categories. BI-RADS I: the breast is almost entirely fatty; BI-RADS II: there is some fibroglandular tissue; BI-RADS III: the breast is heterogeneously dense; BI-RADS IV: the breast is extremely dense. Lam et al. reported associations between BI-RADS II-IV and breast carcinoma (adjusted for weight) in postmenopausal women of which the risks are 1.6, 2.3, and 4.5, respectively [42].

For denoising the mammographic datasets, the correlation-based feature selection (CFS) technique introduced in [43] is employed to weaken the impact of noisy data in mammographic datasets in this paper. CFS is a filter-based approach to feature selection and uses a search algorithm along with an evaluation metric to decide on the 'goodness' or merit of potential feature subsets. Rather than scoring (and ranking) individual features, the method scores (and ranks) the worth of subsets of features. As the feature subset space is usually large, CFS employs a best-first-search heuristic. This heuristic algorithm takes into account the usefulness of individual features for predicting the class along with the level of intercorrelation amongst features. It assumes that good feature subsets contain features that are highly correlated to the class, yet not correlated to each other. CFS calculates a matrix of feature-to-class and feature-to-feature correlations from the training data. Table 1 shows the reduced sizes of the mammographic datasets with 4 strategies of class labels by CFS. The reduced 4 datasets will be applied in the following experiments as well.

Table 1. Reduct size of MIAS datasets

Original feature	Reduct size			
	BI-RADS	Boyd	Tabár	Wolfe
280	32	32	31	30

5. Experimental Results

A. *Experimental Set-up*

For experimental evaluation, both KFRNN and KVQNN employ the Wave kernel function as a similarity metric[1]. For KFRNN, the Kleene-Dienes *T*-norm [44], [45] is used to implement the implicator, which is defined by $I(x,y) = \max(1-x, y)$. The quantifiers used to implement KVQNN are $Q_l = Q_{(0.1,0.6)}$ and $Q_u = Q_{(0.2,1.0)}$ [34], empirically chosen according to the general formula:

$$Q_{(\alpha,\beta)}(x) = \begin{cases} 0, & x \le \alpha \\ \frac{2(x-\alpha)^2}{(\beta-\alpha)^2}, & \alpha \le x \le \frac{\alpha+\beta}{2} \\ 1 - \frac{2(x-\beta)^2}{(\beta-\alpha)^2}, & \frac{\alpha+\beta}{2} \le x \le \beta \\ 1, & \beta \le x \end{cases}. \quad (27)$$

Stratified 10 × 10-fold cross-validation (10-FCV) is employed for data validation. In 10-FCV, a given dataset is partitioned into 10 subsets. Of these 10 subsets, a single subset is retained as the testing data for the learned classifier, and the remaining 9 subsets are used for training. The cross-validation process is then repeated 10 times (the number of folds). The 10 sets of results are then aggregated to produce a single estimation of accuracy. The advantage of 10-FCV over random sub-sampling is that all objects are used for both training and testing, and each object is used for testing only once per fold. The stratification of the data prior to its division into different folds ensures that each class label (as far as possible) has equal representation in all folds, thereby helping to alleviate bias/variance problems [46].

In order to investigate the level of 'fit' of these models, the root mean squared error (RMSE) measure is used. The RMSE is the squared root of the variance of the residuals. It indicates the absolute fit of a model to the data and how close the observed data objects are to the model predicted values. Note that RMSE is an absolute measure. As the squared root of a variance, RMSE can be viewed as the standard deviation of the unexplained variance. Lower values of RMSE indicate better fit. RMSE

[1] The relevant WEKA software can be download from https://dl.dropboxusercontent.com/u/2043486/weka.jar

is a good measure of how accurately the model predicts the response, and is a generally accepted criterion for assessing fit, if the purpose of the resulting model is for prediction. In addition, conventional classification accuracy is also used to assess the performance of learnt classifiers.

To compare with the existing work, in this paper, the kappa statistics [47] is employed to evaluate the experimental results also. The kappa statistics is generally thought to be a more robust measure than simple percent agreement calculation since it summarises the level of agreement between observers after agreement by chance has been removed. It tests how well observers agree with themselves (repeatability) and with each other (reproducibility). It should be noted that the kappa statistic only tends to make sense when a comparison involves an equal number of classes, but for presentational completeness the figures for all cases are provided here.

B. *Performance Evaluation*

In this section, a comparison with other nearest-neighbour-based methods is systematically presented first. Then, a comparative study with four other popular learning classifier algorithms is also reported. These experiments make use of both the original and reduced MIAS datasets.

1) Comparison with Alternative Nearest-neighbour Classifiers: Here, KFRNN and KVQNN are compared with three nearest-neighbour classification methods: IBk (k nearest-neighbours) [17], FNN (standard fuzzy nearest-neighbours) [29], and the fuzzy ownership algorithm, FRNN-O [30], which is the state-of-the-art version of fuzzy rough nearest-neighbour algorithms. In order to comprehensively evaluate the performance of the kernel-based nearest-neighbour algorithms on the MIAS datasets, k is increasingly set to all the odd numbers between 0 and 322 (the number of the objects in a dataset) in different runs. With an extra round for the case when $k = 322$, this results in 162 sets of runs for each dataset as for each value of k, 10×10 -fold cross-validation is performed. The results can be seen in Figure 2.

For all of the MIAS datasets, the denoising process using CFS improves the performance for most classifiers. Occasionally, by IBk, the result for reduced BI-RADS datasets is worse than that of using the original. Compared with the other nearest-neighbour algorithms, generally, the kernel-based nearest-neighbour algorithms result in the best performance. As with the proposed approach, the state-of-the-art algorithm FRNN-O also obtains consistent performance for all MIAS datasets and hence, has a robust performance. However, for all original MIAS datasets and reduced Wolfe datasets, the results of FRNN-O are not so good as those of KFRNN and KVQNN.

The highest classification accuracies achieved by each nearest-neighbour algorithm for the original and reduced MIAS datasets are summarised in Table 2. It is noted that very occasionally, namely for the original BI-RADS dataset and the reduced Tabár labelled dataset, KVQNN slightly underperforms than IBk and FRNN-O, respectively. However, in general, KVQNN provides the best performances across all the original and reduced MIAS datasets.

2) Comparison with other Popular Techniques: As shown above, the KVQNN algorithm offers the best performance amongst the nearest-neighbour approaches. In order to further evaluate the performance of kernel-based fuzzy-rough nearest-neighbour algorithms on the mammographic dataset, the experimental results are also compared against those obtained by the use of other types of popular learning classifier. The algorithms compared are briefly outlined below, which are each a commonly adopted representative of the underlying methodology shared by their corresponding type of learning classifiers:

Table 2. Highest classification accuracies of different nearest-neighbour algorithms

	BI-RADS		Boyd		Tabár		Wolfe	
Algorithm	Original	Reduced	Original	Reduced	Original	Reduced	Original	Reduced
KVQNN	71.45	74.72	60.22	61.79	63.55	67.89	67.10	70.54
KFRNN	69.74	71.34	54.91	59.50	59.78	66.32	65.68	67.57
FNN	64.24	62.17	50.93	51.98	56.84	60.72	59.99	59.40
FRNN-O	66.00	72.17	53.48	60.71	57.18	67.92	62.08	64.02
IBk	71.89	73.69	58.83	60.32	63.09	65.83	66.17	68.39

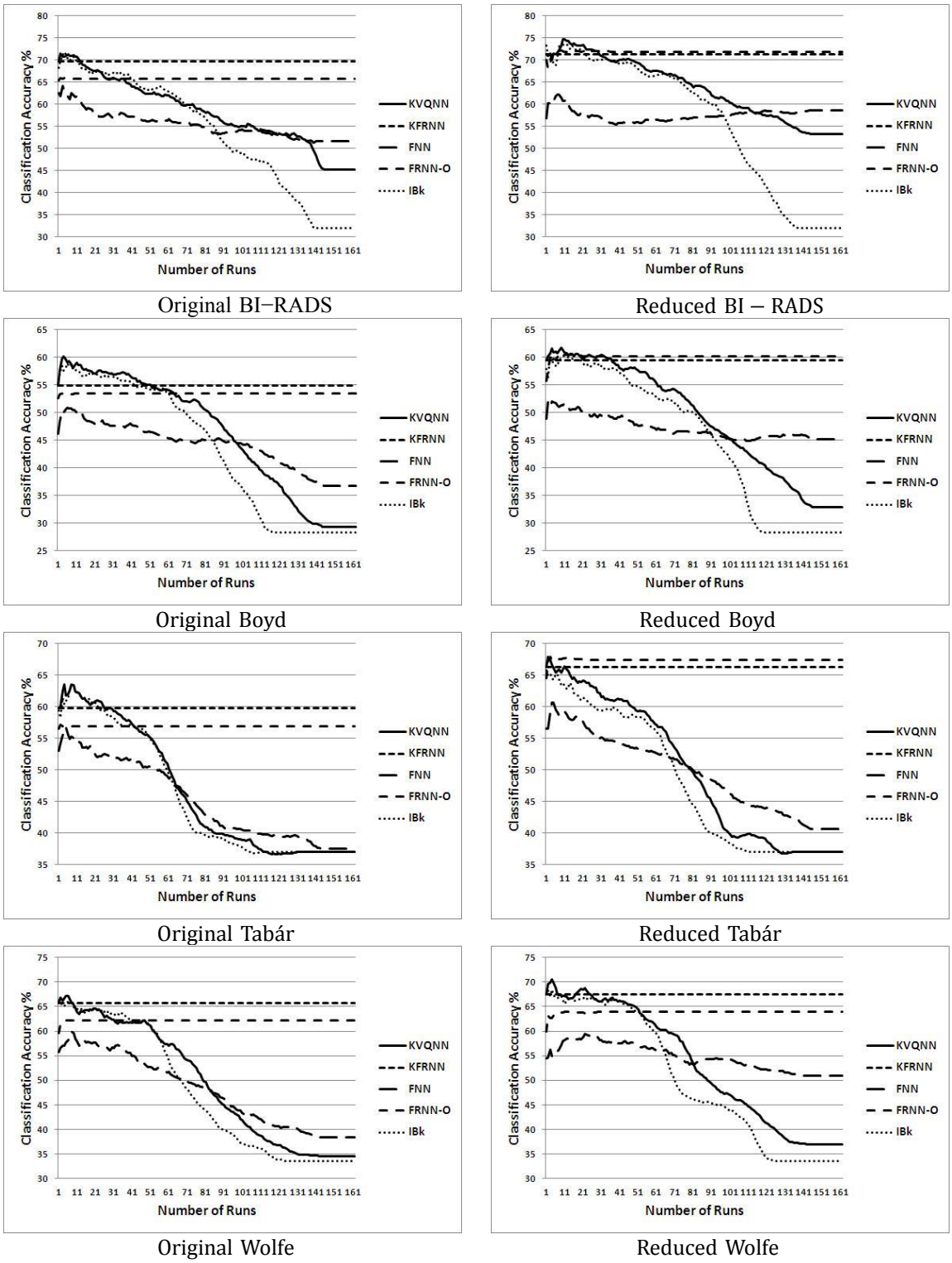


Figure 2. Comparison with alternative nearest-neighbour algorithms

- NB (naive Bayesian classifier) [11] is a simple probabilistic classifier based on applying Bayes' theorem [10] with strong (naive) independence assumptions. Depending on the precise nature of the probability model, naive Bayesian classifiers can be trained very efficiently in a supervised learning setting. The learning only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.
- SVM (as function based classifier) [12] is a kernel-based algorithm. It solves the non-linear sep-

arable problems by mapping them into a higher dimensional feature space. In such a space, the points mapped from the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped onto that same space and predicted to belong to a category based on which side of the gap they fall on.

- J48 (as decision tree based classifier) is based on ID3 [13] and creates decision trees by choosing the most informative features and recursively partitioning a training data table into subtables based on the values of such features. Each node in the tree represents a feature, with the subsequent nodes branching from the possible values of this node according to the current subtable. Partitioning stops when all data items in the subtable have the same classification. A leaf node is then created to represent this classification.
- JRip (as rule induction-based classifier) [15] learns propositional rules by repeatedly growing rules and pruning them. During the growth phase, features are added greedily to fit training samples. Once the ruleset is generated, a further optimisation is performed where rules are evaluated and poor quality ones deleted, based on their performance on randomised data.

Classification accuracy rate, RMSE and the kappa statistic are again used to support the comparative study of applying the aforementioned approaches for mammographic risk analysis. In comparing the experimental results below, only those of KVQNN are used (as KVQNN consistently beats KFRNN as shown in Table 2). Once again, 10×10-fold cross-validation based on all original and reduced MIAS datasets is performed.

The average classification accuracies for the original and reduced MIAS datasets are respectively recorded in Tables 3 and 4, with standard deviations in brackets. Generally, KVQNN gains the best and most stable results. Although for the original BI-RADS, Boyd and Tabár datasets, KVQNN occasionally underperforms as compared to SVM, for the four reduced MIAS datasets KVQNN achieves the best results compared to the other approaches. In particular, for reduced Tabár, KVQNN may have a slightly larger value of standard deviation in certain cases, but it always has considerably high classification accuracy rates. This demonstrates that overall, the KVQNN algorithm provides a better performance than the other classifiers.

Table 3: Classification accuracy: Original MIAS datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	71.45	(7.40)	69.75	(7.93)	69.09	(7.71)	67.12	(8.67)	64.42	(8.18)
Boyd	60.22	(8.72)	57.02	(9.49)	60.49	(8.94)	52.16	(8.45)	50.87	(8.70)
Tabár	63.55	(7.69)	60.28	(8.29)	66.50	(7.64)	56.85	(7.80)	56.09	(7.78)
Wolfe	67.10	(7.12)	66.21	(6.88)	67.50	(8.86)	60.12	(7.91)	60.16	(7.89)

Table 4: Classification accuracy: Reduced MIAS datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	74.72	(7.25)	72.66	(7.41)	74.12	(7.93)	65.79	(8.20)	67.26	(8.15)
Boyd	61.79	(8.29)	58.92	(9.16)	59.63	(8.48)	52.27	(7.78)	50.93	(9.00)
Tabár	67.89	(8.39)	62.49	(8.49)	65.67	(7.30)	59.71	(7.61)	58.35	(8.65)
Wolfe	70.54	(7.05)	68.18	(7.80)	70.38	(7.56)	61.18	(8.06)	62.21	(8.17)

In Tables 5 and 6, the results are presented for kappa statistic on the original and reduced MIAS datasets. High values of kappa statistic are indicative of high agreement between the comparators. Thus, KVQNN provides the best performance consistently for most MIAS datasets except the original Boyd and Tabár labelling. In particular, for the BI-RADS, reduced Wolfe, and reduced BI-RADS datasets, the values of kappa statistic of KVQNN are equal to or higher than 0.60, which means that the results gained by KVQNN indicate highly moderate or substantial agreements between the comparators. Additionally, given the low standard deviations, KVQNN results in a considerably stable performance as well.

Tables 7 and 8 demonstrate the comparison between the RMSE values of KVQNN and those of the other classifiers for the original and reduced MIAS datasets. Note that RMSE should not be assessed in isolation of classification accuracy or other metrics. Because RMSE is a metric associated with the 'fit' of the model to the actual data, very low RMSE values are possibly indicative of overfitting. It is therefore important to view it in terms of other metrics, namely, classification accuracy rate and the kappa statistic.

Table 5: Kappa statistic: Original MIAS Datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	0.60	(0.10)	0.59	(0.11)	0.57	(0.11)	0.55	(0.12)	0.51	(0.11)
Boyd	0.49	(0.11)	0.46	(0.12)	0.50	(0.11)	0.40	(0.11)	0.36	(0.11)
Tabár	0.50	(0.11)	0.48	(0.11)	0.55	(0.10)	0.42	(0.11)	0.40	(0.11)
Wolfe	0.55	(0.10)	0.54	(0.09)	0.55	(0.12)	0.45	(0.11)	0.46	(0.11)

Table 6: Kappa statistic: Reduced MIAS datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	0.65	(0.10)	0.63	(0.10)	0.64	(0.11)	0.53	(0.11)	0.55	(0.11)
Boyd	0.51	(0.11)	0.48	(0.11)	0.49	(0.11)	0.40	(0.10)	0.37	(0.12)
Tabár	0.57	(0.12)	0.50	(0.11)	0.53	(0.10)	0.46	(0.10)	0.43	(0.12)
Wolfe	0.60	(0.10)	0.56	(0.11)	0.59	(0.10)	0.47	(0.11)	0.49	(0.11)

Table 7: RMSE: Original MIAS datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	0.34	(0.04)	0.38	(0.05)	0.39	(0.05)	0.39	(0.05)	0.37	(0.04)
Boyd	0.32	(0.03)	0.37	(0.04)	0.36	(0.04)	0.38	(0.03)	0.34	(0.03)
Tabár	0.32	(0.03)	0.39	(0.04)	0.36	(0.04)	0.40	(0.04)	0.36	(0.03)
Wolfe	0.35	(0.03)	0.40	(0.04)	0.40	(0.06)	0.43	(0.04)	0.39	(0.04)

Table 8: RMSE: Reduced MIAS datasets

	KVQNN		NB		SVM		J48		JRip	
BI-RADS	0.31	(0.04)	0.35	(0.05)	0.36	(0.06)	0.39	(0.05)	0.36	(0.04)
Boyd	0.31	(0.03)	0.34	(0.04)	0.36	(0.04)	0.38	(0.03)	0.34	(0.03)
Tabár	0.32	(0.04)	0.37	(0.04)	0.37	(0.04)	0.38	(0.04)	0.36	(0.03)
Wolfe	0.34	(0.04)	0.38	(0.05)	0.38	(0.05)	0.42	(0.05)	0.38	(0.04)

As can be observed, KVQNN results in the lowest but reasonable values of RMSE for all the original and reduced MIAS datasets. Given the lowest standard deviations achieved by KVQNN as well, it means that the results obtained by KVQNN algorithm consistently fit the underlying models of MIAS datasets well.

Overall, as a representative of the kernel-based fuzzy rough nearest-neighbour algorithm, KVQNN performs consistently well and robustly for all of the 4 popular classification methods for MIAS datasets. Such performance includes not only a high classification accuracy, but also a rationally good fitness to the underlying model of MIAS datasets. This ensures that the proposed approach provides a good generalisation for mammographic risk assessment, with reduced potential overfit on the MIAS datasets.

The proposed work gives the best accuracy. However, the classification rates achieved by any of the compared approaches are relatively low (in terms of the usual applications of such classification techniques), having been able to obtain a value of just around 70%. Nevertheless, as pointed out previously, this is a very difficult application domain. There is no actual ground-truth to ensure which classification is to be ultimately correct in the first place. Therefore, these results help to provide a useful reference aid for human decision making. The eventual task of deciding on the actual mammographic risk is up to human radiologists. The present approach is shown to be a good candidate for playing such a supportive role.

6. Conclusions

In this paper, an effective classification approach, kernel-based fuzzy-rough nearest-neighbour (KFRNN) and a direct extension of it, kernel-based vaguely quantified nearest-neighbour (KVQNN), have been proposed. These algorithms have been utilised to support performing the task of mammographic risk analysis. Compared with alternative nearest-neighbour-based methods and other popular classifiers learning techniques, KFRNN and KVQNN achieve a better and more robust performance for original and reduced MIAS datasets.

Whilst promising, further work remains. For instance, as an initial implementation, the wave kernel function is adopted in this work to construct fuzzy similarities. It would be interesting to investigate how other types of kernel function may be used as the alternative and what might be the effects. Additionally, in [39], two algorithms, namely: similarity nearest-neighbour (SNN) and aggregated-similarity nearest-neighbour (ASNN), are also proposed as the fuzzy set-based equivalence of FRNN and that of VQNN, respectively. Clearly, how the proposed kernel-based techniques will perform for the fuzzy parallels of FRNN and VQNN is worth further

investigation. Furthermore, in this research, the CFS method is utilised for feature selection. With such a technique, the results of the assessment of mammographic risk are improved considerably. A further extension to this work would therefore be to explore how KFRNN or KVQNN may perform using different feature selection methods [48], [49].

A more complete comparison of KFRNN and KVQNN with other techniques over different datasets from other application domains, e.g., Mars terrain images [50], would form the basis for a wider series of topics for future investigation. Given the existing use of multiple criteria for assessing classifiers performance, the recent development of fuzzy complex numbers may be utilised to support such evaluation [51]. Alternatively, different performance criteria may be integrated using information aggregation techniques [52], [53] to simplify the evaluation process.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (Grant No. 61272171), the Fundamental Research Funds for the Central Universities (Grant No. 3132014094, 3132013335, 3132013325), and the China Postdoctoral Science Foundation (Grant No. 2013M541213). The authors would like to thank the support provided by Aberystwyth University and by the colleagues in the Advanced Reasoning Group with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science at Aberystwyth University, UK. The authors are also grateful to the reviewers for their constructive comments which have helped improve this work.

References

[1] World Health Organization, "Breast cancer: prevention and control," from: http://www.who.int/cancer/detection/breastcancer/en/, accessed: Aug. 2014.

[2] World Cancer Research Fund International, "Breast cancer," from: http://www.wcrf.org/cancer_statistics/data_specific_cancers/breast_cancer_statistics.php, accessed: Aug. 2014.

[3] World Health Organization, International Agency for Research on Cancer. "Latest world cancer statistics: Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed," from: www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf, accessed: Aug. 2014

[4] M. Tortajada, A. Oliver, R. Martí, S. Ganau, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwiggelaar, "Breast peripheral area correction in digital mammograms. Computers in Biology and Medicine," vol. 50, pp. 32-40, 2014.

[5] V. Tesic, B. Kolaric, A. Znaor, S. K. Kuna, and B. Brkljacic, "Mammographic density and estimation of breast cancer risk in intermediate risk population," *The Breast Journal*, vol. 19, no. 1, pp. 71–78, 2013.

[6] A. Wang, C. Vachon, K. Brandt, and K. Ghosh, "Breast Density and Breast Cancer Risk: A Practical Review," *Mayo Clinic Proceedings*, vol. 89, no. 4, pp. 548–557, 2014.

[7] Y. Qu, C. Shang, W. Wu, and Q. Shen, "Evolutionary Fuzzy Extreme Learning Machine for Mammographic Risk Analysis," *International Journal of Fuzzy Systems*, vol. 13, no.4, pp. 282–291, 2011.

[8] H. Strange, Z. Chen, E. Denton, and R. Zwiggelaar, "Modelling mammographic microcalcification clusters using persistent mereotopology," *Pattern Recognition Letters*, vol. 47, pp. 157-163, 2014.

[9] W. He, E. Denton, and R. Zwiggelaar, "A study on mammographic image modelling and classification using multiple databases," in *Breast Imaging*, pp. 696-701, 2014.

[10] P. Pfeiffer, *Concepts of probability theory*. Courier Dover Publications, U.S., 2013.

[11] R. Neapolitan, *Probabilistic reasoning in expert systems: theory and algorithms*, Create Space Independent Publishing Platform, 2012.

[12] B Han and L. Davis, "Density-based multifeature background subtraction with support vector machine," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1071–1023, 2012.

[13] C. Aggarwal and C. Zhai, *Mining text data*, The Springer Publishing Co., 2012.

[14] K. Karimi and H. Hamilton, "Finding temporal relations: Causal Bayesian networks vs. C4.5," *Foundations of Intelligent Systems*, pp. 266–273, 2010.

[15] B. Qin, Y. Xia, and S. Prabhakar, "Rule induction for uncertain data," *Knowledge and Information Systems*, vol. 29, no. 1, pp. 103–130, 2011.

[16] T. Lejarraga, V. Dutt, and C. Gonzalez, "Instance-based learning: A general model of repeated binary choice," *Journal of Behavioral Decision Making*, vol. 25, no. 2, pp. 143–153, 2012.

[17] S. Jiang, G. Pang, M. Wu, and L. Kuang "An improved K-nearest-neighbor algorithm for text categorization," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1503–1509, 2012.

[18] Y. Qu, C. Shang, Q. Shen, N. Mac Parthaláin, and W. Wu, "Kernel-based fuzzy-rough nearest neighbour classification," in *Proceedings of the 20th In-*

ternational Conference on Fuzzy Systems, pp. 1523–1529, 2011.

[19] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[20] S. Maji, A. Berg, and J. Malik, "Efficient classification for additive kernel SVMs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66 – 77, 2013.

[21] J. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer*, vol. 37, pp. 2486–2492, 1976.

[22] N. Boyd, J. Byng, R. Jong, E. Fishell, L. Little, A. Miller, G. Lockwood, D. Tritchler, and M. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study," *Journal Of The National Cancer Institute*, vol. 87, no. 9, pp. 670–675, 1995.

[23] L. Tabár, T. Tot, and P. Dean, *The Art and Science of Early Detection with Mammography*. Georg Thieme Verlag, 2005.

[24] American College of Radiology, *Illustrated Breast Imaging Reporting and Data System BIRADS*, 3rd ed., 1998.

[25] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishing, 1991.

[26] D. Dubois and H. Prade, "Putting rough sets and fuzzy sets together," *Intelligent Decision Support*, pp. 203–232, 1992.

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

[28] C. Cornelis, M. De Cock, and A. Radzikowska, "Vaguely quantified rough sets," *Lecture Notes in Artificial Intelligence*, vol. 4482, pp. 87–94, 2007.

[29] J. Keller, M. Gray, and J. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, And Cybernetics*, vol. 15, no. 2, pp. 580–588, 1985.

[30] M. Sarkar, "Fuzzy-rough nearest neighbors algorithm," *Fuzzy Sets and Systems*, vol. 158, pp. 2123–2152, 2007.

[31] L. Sun and C. Li, "A fast and scalable fuzzy-rough nearest neighbor algorithm," *WRI Global Congress on Intelligent Systems*, vol. 4, pp. 311–314, 2009.

[32] H. Bian and L. Mazlack, "Fuzzy-rough nearest-neighbor classification approach," in *Proceeding of the 22nd International Conference of the North American Fuzzy Information Processing Society*, pp. 500–505, 2003.

[33] V. Babu and P. Viswanath, "Rough-fuzzy weighted k-nearest leader classifier for large data sets," *Pattern Recognition*, vol. 42, no. 9, pp. 1719–1731, 2009.

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

[35] M. Genton, "Classes of kernels for machine learning: a statistics perspective," *Journal of Machine Learning Research*, vol. 2, pp. 299–312, 2001.

[36] B. Moser, "On the T-transitivity of kernels," *Fuzzy Sets and Systems*, vol. 157, pp. 1787–1796, 2006.

[37] Q. Hu, D. Chen, D. Yu, and W. Pedrycz, "Kernelised fuzzy rough sets," in *th International Conference, Rough Sets and Knowledge Technology*, pp. 304–311, 2009.

[38] R. Jensen and C. Cornelis, "Fuzzy-rough instance selection," in *Proceedings of the 19th International Conference on Fuzzy Systems*, pp. 1776–1782, 2010.

[39] Y. Qu, Q. Shen, N. Mac Parthaláin, C. Shang, and W. Wu, "Fuzzy similarity-based nearest-neighbour classification as alternatives to their fuzzy-rough parallels," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 184-195, 2013.

[40] J. Suckling, J. Partner, D. Dance, S. Astley, I.Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, D.Betal, P.Taylor, and J. Savage, "The mammographic image analysis society digital mammogram database," in *International Workshop on Digital Mammography*, pp. 211–221, 1994.

[41] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Perez, E. Denton, and R. Zwiggelaar, "A novel breast tissue density classification methodology," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 55–65, 2008.

[42] P. Lam, P. Vacek, B. Geller, and H. Muss, "The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in vermont," *Cancer*, vol. 89, pp. 369–375, 2000.

[43] M. Hall, *Correlation-based feature selection for discrete and numeric class machine learning*. PhD thesis, University of Waikato, 2000.

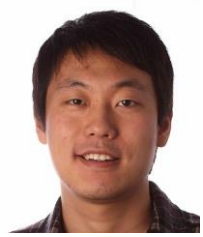
[44] S. Kleene, *Introduction to Metamathematics*. New York: Van Nostrand, 1952.

[45] S. Dienes, "On an implication function in many-valued systems of logic," *Journal of Symbolic Logic*, vol. 14, no. 2, pp. 95–97, 1949.

[46] D. Rajnarayan and D. Wolpert, "Bias-variance trade-offs: novel applications," *Encyclopedia of Machine Learning*, pp. 101–110, 2010.

[47] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[48] N. Mac Parthaláin and R. Jensen, "Unsupervised fuzzy-rough set-based dimensionality reduction," *Information Sciences*, vol. 229, pp. 106-121, 2013.
[49] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*. IEEE Press and Wiley & Sons, 2008.
[50] C. Shang and D. Barnes, "Fuzzy-rough feature selection aided support vector machines for Mars image classification," *Computer Vision and Image Understanding*, vol. 117, no. 3, pp. 202–213, 2013.
[51] X. Fu and Q. Shen, "Fuzzy complex numbers and their application for classifiers performance evaluation," *Pattern Recognition*, vol. 44, no. 7, pp. 1403–1417, 2011.
[52] T. Boongoen, C. Shang, N. Iam-On, and Q. Shen, "Extending data reliability measure to a filter approach for soft subspace clustering," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 40, no. 6, 2011.
[53] T. Boongoen and Q. Shen, "Nearest-neighbor guided evaluation of data reliability and its applications," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 40, no. 6, pp. 1622–1633, 2010.



Yanpeng Qu is a Lecturer with the Information Science and Technology College at Dalian Maritime University, China. He received a PhD degree in Computational Mathematics from Dalian University of Technology, China. His current research interests include rough and fuzzy set theory, pattern recognition, neural networks, classification and feature selection.

Changjing Shang is a University Research Fellow with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science at Aberystwyth University, UK. She obtained a PhD in Computing and Electrical Engineering from Herriot-Watt University, UK and worked for Herriot-Watt, Loughborough and Glasgow Universities prior to joining Aberystwyth. Her research interests include pattern recognition, data mining and analysis, feature extraction and selection, space robotics, and image processing and classification.

Qiang Shen holds the established Chair in Computer Science and is the Director of the Institute of Mathematics, Physics and Computer Science at Aberystwyth University, UK. He has a PhD in Knowledge-Based Systems and a DSc in Computational Intelligence. His research interests include: computational intelligence, reasoning under uncertainty, pattern recognition, data mining, and their applications for intelligent decision support (e.g., crime detection, consumer profiling, systems monitoring, and medical diagnosis). Prof. Shen has been a long-serving associate editor of two premier IEEE Transactions (Cybernetics and Fuzzy Systems), and an editorial board member for several other leading international journals. He has authored 2 research monographs, and over 340 peer-reviewed papers, including one receiving an IEEE Outstanding Transactions Paper award.

Neil Mac Parthaláin is a University Research Fellow with Department of Computer Science, Institute of Mathematics, Physics and Computer Science at Aberystwyth University, Wales, UK. He obtained a PhD degree from Aberystwyth University. His areas of research include rough set theory, fuzzy set theory, pattern recognition, feature selection, classification and medical imaging and applications. He has published around 30 peer-refereed conference papers and academic journal articles in these and related areas. He was a member of the organising committee for 16th International Conference on Fuzzy Systems (FUZZ-IEEE 2007), London and has been involved with the organization of a number of special sessions at the IEEE series of International Conferences on Fuzzy Systems.

Wei Wu received a MSc degree from Jilin University, Changchun, China and a D.Phil. degree from Oxford University, Oxford, UK. He is a Professor with the School of Mathematical Sciences at Dalian University of Technology, Dalian, China. He serves as an associate editor for the Journal of Information and Computational Science, Numerical Mathematics - A Journal of Chinese Universities, and the Journal of Mathematical Research and Exposition. His research interests include numerical analysis and neural network computation.