World Scientific
www.worldscientific.com

Quantum-inspired minimum distance classification in a biomedical context

Giuseppe Sergioli[*,¶], Giorgio Russo[†,‖], Enrica Santucci[*,**],
Alessandro Stefano[†,††], Sebastiano Emanuele Torrisi[‡,‡‡],
Stefano Palmucci[§,§§], Carlo Vancheri[‡,¶¶] and Roberto Giuntini[*,‖‖‖]

[*]University of Cagliari, Via Is Mirrionis 1,
I-09123 Cagliari, Italy

[†]Institute of Molecular Bioimaging and Physiology,
National Research Council (IBFM-CNR),
Contrada Pietropollastra-Pisciotta, I-90015 Cefalù, Italy

[‡]Regional Referral Centre for Rare Lung Diseases,
A.O.U. Policlinico-Vittorio Emanuele,
University of Catania, Catania, Italy

[§]Radiodiagnostics and Oncological Radiotherapy Unit,
University Hospital "Policlinico-Vittorio Emanuele",
Catania, Italy
[¶]giuseppe.sergioli@gmail.com
[‖]giorgio.russo@ibfm.cnr.it
[**]enrica.santucci@gmail.com
[††]alessandro.stefano@ibfm.cnr.it
[‡‡]torrisiseby@hotmail.it
[§§]spalmucci@sirm.org
[¶¶]vancheri@unict.it
[‖‖‖]giuntini@unica.it

We propose an application of a *quantum-inspired* version of the Nearest Mean Classifier (NMC) (G. Sergioli, E. Santucci, L. Didaci, J. A. Miszczak and R. Giuntini, A quantum inspired version of the NMC classifier, *Soft Comput.* **22**(3) (2018) 691. G. Sergioli, G. M. Bosyk, E. Santucci and R. Giuntini, A quantum-inspired version of the classification problem, *Int. J. Theo. Phys.* **56**(12) (2017) 3880. E. Santucci and G. Sergioli, Classification problem in a quantum framework, in quantum foundations, probability and information, *Proc. Quantum and Beyond Conf.*, 13–16 June 2016, Vaxjo, Sweden, A. Khrennikov and T. Bourama, Springer-Berlin, Germany, 2018 (in press, 2018). E. Santucci, Quantum minimum distance

classifier, *Entropy* **19**(12) (2017) 659.) to a biomedical context. In particular, we benchmark the performances of such a quantum-variant of NMC against NMC and other (nonlinear) classifiers with respect to the problem of classifying the probability of survival for patients affected by idiopathic pulmonary fibrosis (IPF).

Keywords: Nearest mean classifier; quantum theory; idiopathic pulmonary fibrosis.

1. Introduction

Quantum mechanics is a probabilistic theory that turns out to be particularly suitable to describe different kinds of stöchastic processes that — in principle — are not confined to the microscopic domain. In recent years, for example, quantum formalism has been widely exploited in nonstandard contexts such as game theory, economic processes, cognitive sciences and so on.[1–7]

Along this perspective, another nonstandard application of quantum formalism concerns the solution of some classification problems which are typical of signal processing[8] and of pattern recognition.[9,10] Exhaustive surveys concerning the applications of quantum computing in computational intelligence and machine learning can be found in Refs. 11 and 12. The main aim of most of these approaches is to speed up the computational processes involved in machine learning by "quantizing" some key-algorithms that are used in classical pattern recognition (Refs. 13 and 14). The approach we have proposed instead in, Refs. 15–18 is based on a different ground that essentially consists in using the quantum formalism to reach remarkable benefits also in a classical context. Our model, called *Quantum Nearest Mean Classifier* (QNMC), allows us to process any kind of classical dataset in a supervised system by (i) translating each element of the dataset (pattern) into a *density pattern*, i.e. a density operator (the mathematical tool to formally describe a quantum state) of the real vector space associated to the dataset at issue; (ii) defining, for any class of density patterns, a *quantum centroid* that has no counterpart in the initial classical dataset; (iii) using the standard minimum distance procedure to classify an unlabeled density pattern; (iv) decoding the result of the classification process in the classical pattern space.

The aim of the construction we have previously outlined is to show that quantum formalism can be profitably used also outside its natural domain of application. In particular, we will show that the expressive power of quantum formalism allows us to obtain remarkable advantages in the accuracy of standard machine learning processes like pattern recognition.

In Refs. 15 and 16 we have benchmarked the performances of QNMC and of the standard *Nearest Mean Classifier* (NMC) against both some artificial and real-world datasets that are typically used in machine learning to conclude that QNMC significantly outperforms NMC.

In the present work, we propose a particular application of our model to a real dataset (IPF dataset) that is obtained from a group of 126 patients possibly suffering from *Idiopathic pulmonary fibrosis* (IPF). IPF is a disease characterized by the development of fibrotic areas within the parenchyma of lungs causing a progressive

reduction of the respiratory function. The prognosis of IPF patient is very poor with a median survival of 3–5 years from diagnosis; the dataset includes baseline variables with an established relation to patient's survival. In this paper, we refer to the IPF dataset to compare the performances of two different variants of QNMC not only with NMC but also with other well-known standard classifiers: *Linear Discriminant Analysis* (LDA) and *Quadratic Discriminant Analysis* (QDA).

The paper is organized as follows: in the first section, we briefly describe the formal structure of both NMC and QNMC. In the second section, we summarize some interesting results presented in Refs. 15–18 by comparing NMC and QNMC on different datasets and we show the advantages of QNMC in terms of pattern classification accuracy. In the third section, we first introduce an alternative encoding from the real vector (pattern) space to the density operator space that will turn out to be particularly useful when applied to the IPF dataset. Second, we introduce the IPF dataset and we provide a detailed description of the dataset features. Finally, we show and discuss the promising results arising from the application of two different QNMC variants on the IPF dataset, showing an improvement of the accuracy with respect to some standard classifiers (NMC, LDA, QDA). In the last section, we propose some possible developments and strategies that could be used to improve the benchmark of our model.

2. Classical and Quantum Version of the Nearest Mean Classifier

In this section, we briefly describe the quantum version of the standard Nearest Mean Classifier, which is a particularly simple and fast algorithm of supervised learning, i.e. learning from a training dataset of correctly labeled objects.

In machine learning, any object is characterized by a given set of d features. A d-feature object is naturally represented by a d-dimensional real vector $\mathbf{x} = [x^1, \ldots, x^d] \in \mathbb{R}^d$.[a] Formally, a pattern can be represented as a pair $(\mathbf{x}_i, \lambda_i)$, where $\mathbf{x}_i$ is the d-dimensional vector associated with a given object and λ_i is the label that refers to the class which the object belongs to. We can simply consider a class as a set of objects. For the sake of simplicity, we confine ourselves to the special (but very common) case, where each object belongs to one and only one class of objects. Let $\Lambda = \{\lambda_1, \ldots, \lambda_N\}$ be the set of labels that is in one-to-one correspondence with the set of all classes. The aim of the classification process is to design a *classifier* that attributes (in the most accurate way) a label (class) to any unlabeled object. In supervised learning, such a classifier is obtained by getting information from a *training set* $\mathcal{S}_{\text{tr}}$, i.e. a set of correctly labeled objects. Formally,

$$\mathcal{S}_{\text{tr}} = \{(\mathbf{x}_1, \lambda_1), \ldots, (\mathbf{x}_M, \lambda_M)\},$$

where $\mathbf{x}_i \in \mathbb{R}^d$ and λ_i is the label associated to the class which x_i belongs to.

---

[a] For the sake of clarity regarding the indexes, we use the superscription to indicate the different components of the vector and subscript to indicate different vectors.

Given a training dataset $\mathcal{S}_{\mathrm{tr}} = \{(\mathbf{x}_1, \lambda_1), \ldots, (\mathbf{x}_M, \lambda_M)\}$, we can define the jth class $\mathcal{S}_{\mathrm{tr}}^j$ in the following way:

$$\mathcal{S}_{\mathrm{tr}}^j = \{(\mathbf{x}_i, \lambda_i) \in \mathcal{S}_{\mathrm{tr}} : \lambda_i = \lambda_j\}.$$

The jth class $\mathcal{S}_{\mathrm{tr}}^j$ represents the set of all patterns of $\mathcal{S}_{\mathrm{tr}}$ that have been label λ_j. Finally, by M_j we will denote the number of elements of $\mathcal{S}_{\mathrm{tr}}^j$.

One of the simplest classification methods in pattern recognition is the so-called NMC. The NMC algorithm consists of the following steps:

(1) Training: one has to compute the centroid μ_j for each class $\mathcal{S}_{\mathrm{tr}}^j$ in the following way:

$$\boldsymbol{\mu}_j = \frac{1}{M_j} \sum_{i \in \{m \in M : \lambda_m = \lambda_j\}} \mathbf{x}_i. \tag{1}$$

(2) Classification: the associated classifier is a function $Cl : \mathbb{R}^d \to \Lambda$ such that $\forall \mathbf{x} \in \mathbb{R}^d$:

$$Cl(\mathbf{x}) = \lambda_j \quad \text{iff } \forall\, k \in \{1, \ldots, M\} : d(\mathbf{x}, \boldsymbol{\mu}_j) \leq d(\boldsymbol{x}, \boldsymbol{\mu}_k),$$

where $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$ is the Euclidean distance.

Intuitively, this classifier associates to a d-feature object $\mathbf{x}$ the label of the closest centroid to $\mathbf{x}$.

In order to evaluate the performance of a classifier, one introduces another set of patterns (called *test set*) that does not belong to the training set.[19] Formally, the test set is a set $\mathcal{S}_{\mathrm{ts}} = \{\{\mathbf{y}_1, \beta_1\}, \ldots, \{\mathbf{y}_{M'}, \beta_{M'}\}\}$, such that $\mathcal{S}_{\mathrm{tr}} \cap \mathcal{S}_{\mathrm{ts}} = \emptyset$.

Then, by applying NMC to the test set, it is possible to evaluate the semi-supervised classifier performance by considering the accuracy (ACC) of the classification process as the ratio between the number of all correctly classified test patterns and the cardinality of the test set.[b]

Let us notice that the values of such quantities are obviously related to the training/test datasets; as a natural consequence, also the classifier performance is strictly dataset-dependent.

In order to provide a quantum counterpart of NMC (called *Quantum Nearest Mean Classifier* (QNMC)) we need the following steps:

(1) For each pattern, one has to provide a suitable encoding into a quantum object (i.e. a density operator) that we will call *density pattern*;
(2) For each class of density patterns, one has to define the quantum counterpart of the classical centroid, that we will call *quantum centroid*;

---

[b] We recall that the classification accuracy is defined as $\mathrm{ACC} = 1 - \mathrm{ERR}$, where ERR is the classification error.

(3) Finally, one has to provide a suitable notion of *quantum distance* between density patterns.

Even though there are infinite ways to transform a real vector into a density operator, in Ref. 16 we have proposed a promising encoding called *stereographic encoding* (SE). In order to introduce the stereographic encoding, it will be expedient to introduce the notion of stereographic projection.

Let us consider the $(d+1)$-dimensional vector space $\mathbb{R}^{d+1}$. The stereographic projection SP is a map $\text{SP} : \mathbb{R}^{d+1} \to \mathbb{R}^d$ such that for any vector $\mathbf{x} = [x^1, \ldots, x^{d+1}] \in \mathbb{R}^{d+1}$,

$$\text{SP}(\mathbf{x}) = \left(\frac{x^1}{1 - x^{d+1}}, \ldots, \frac{x^d}{1 - x^{d+1}} \right).$$

The inverse of SP is a map $\text{SP}^{-1} : \mathbb{R}^d \to \mathbb{R}^{d+1}$ such that for any vector $\mathbf{x} = [x^1, \ldots, x^d] \in \mathbb{R}^d$,

$$\text{SP}^{-1}(\mathbf{x}) = \frac{1}{\sum_{i=1}^d (x^i)^2 + 1} \left[2x^1, \ldots, 2x^d, \sum_{i=1}^d (x^i)^2 - 1 \right], \tag{2}$$

where $\frac{1}{\sum_{i=1}^d (x^i)^2 + 1}$ is a normalization factor.

Definition 1 (Density pattern by SE). The density pattern $\rho_{\mathbf{x}}$ associated to the d-feature object $\mathbf{x} \in \mathbb{R}^d$ is defined as follows:

$$\rho_{\mathbf{x}} := \text{SP}(\mathbf{x})^\dagger \cdot \text{SP}(\mathbf{x}). \tag{3}$$

Clearly, every density pattern is a quantum pure state, i.e. $\rho_{\mathbf{x}}^2 = \rho_{\mathbf{x}}$. Therefore, SE allows us to encode any real vector $\mathbf{x} \in \mathbb{R}^d$ into a density operator $\rho_{\mathbf{x}}$. On this basis, we can define the quantum training dataset as follows:

$$\mathcal{S}_{\text{tr}}^{\text{q}} := \{\{\rho_{\mathbf{x}_1}, \lambda_1\}, \ldots, \{\rho_{\mathbf{x}_M}, \lambda_M\}\}.$$

In other terms, the quantum training dataset is the original training dataset, where in each pattern the object (the vector $\mathbf{x}$) is replaced by its quantum counterpart (the density pattern $\rho_{\mathbf{x}}$).

On this basis, we can define the notion of *quantum centroid*.

Definition 2 (Quantum centroids). Let $\mathcal{S}_{\text{tr}}^{\text{q}} = \{\{\rho_{\mathbf{x}_1}, \alpha_1\}, \ldots, \{\rho_{\mathbf{x}_M}, \alpha_M\}\}$ be a quantum training dataset. The quantum centroid ρ_j of the jth class $\mathcal{S}_{\text{tr}}^j$ is defined as follows:

$$\rho_j := \frac{1}{M_j} \sum_{i \in \{m \in M : \lambda = \lambda_j\}} \rho_{\mathbf{x}_i}. \tag{4}$$

Notice that the quantum centroids ρ_j are mixed states and that they are generally different from the encoding of their respective classical centroids $\boldsymbol{\mu}_j$. Accordingly, the

definition of quantum centroid leads to a new object that does not have any classical counterpart.

Finally, as to the distance function between density patterns, we will use the *trace distance* which is frequently used in quantum information as a measure of distinguishability between two states (see, e.g. Ref. [20]).

Definition 3 (Trace distance). Let ρ and σ be two quantum density operators belonging to the same Hilbert space. The trace distance (d_{tr}) between ρ and σ is given by

$$d_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2} \mathrm{Tr}|\rho - \sigma|, \tag{5}$$

where $|A| = \sqrt{A^\dagger A}$.

Notice that the trace distance is a metric; hence, it satisfies: (i) $d_{\mathrm{tr}}(\rho, \sigma) \geq 0$ with equality iff $\rho = \sigma$ (positivity), (ii) $d_{\mathrm{tr}}(\rho, \sigma) = d_{\mathrm{tr}}(\sigma, \rho)$ (symmetry) and (iii) $d_{\mathrm{tr}}(\rho, \omega) + d_{\mathrm{tr}}(\omega, \sigma) \geq d_{\mathrm{tr}}(\rho, \sigma)$ (triangle inequality).

We have introduced all the necessary ingredients to describe in detail the QNMC algorithm, which, similarly to the classical case, consists of the following steps:

— obtaining the quantum training dataset $\mathcal{S}_{\mathrm{tr}}^{\mathrm{q}}$ by applying the encoding given in Definition 1 to each pattern of the classical training set $\mathcal{S}_{\mathrm{tr}}$;
— calculating the quantum centroids ρ_j according to Definition 2;
— classifying an arbitrary pattern $\mathbf{x}$ by means of the quantum classifier QCL that is a function $\mathrm{QCL} : \mathbb{R}^d \to \Lambda$ such that $\forall \mathbf{x} \in \mathbb{R}^d : \mathrm{QCL}(\rho_{\mathbf{x}}) = \lambda_i$ iff $\forall k \in \{1, \ldots, M\}$ we have $d_{\mathrm{tr}}(\rho_{\mathbf{x}}, \rho_j) \leq d_{\mathrm{tr}}(\rho_{\mathbf{x}}, \rho_i)$.

2.1. *Experimental results*

In what follows, we summarize some preliminary results that are obtained by comparing the performances of NMC and QNMC on different (artificial and real) different datasets. In particular, we consider three artificial (two-feature) datasets (Moon, Banana and Gaussian) and four real (many-feature) datasets (Diabetes, Cancer, Liver and Ionosphere) extracted from the UCI Irvine Machine Learning Repository.[c]

In our experiment, we follow the standard methodology of randomly splitting each dataset into a training set ($\mathcal{S}_{\mathrm{tr}}$) and a test set ($\mathcal{S}_{\mathrm{ts}}$) that are populated with 80% and 20% of the original dataset, respectively. Moreover, in order to obtain statistically significant results, we perform 100 experiments for each dataset by randomly splitting the dataset each time according the method we have described above.

We summarize our results in Table 1.

---

[c] https://archive.ics.uci.edu/ml/index.php.

Table 1. Average results for NMC and QNMC (in %) and their standard deviations. $\#\mathcal{S}_{\text{tr}}$ is the cardinality of the training dataset; $\#\mathcal{S}_{\text{ts}}$ is the cardinality of the test set; (d) is the number of the features of the elements of the dataset.

Dataset	$\#\mathcal{S}_{\text{tr}}/\#\mathcal{S}_{\text{ts}}$	ACC (NMC)	ACC (QNMC)
Banana	4240/1060(2)	55.0 ± 1.8	$\mathbf{71.0 \pm 1.2}$
Gaussian	160/40(2)	55.5 ± 7.7	$\mathbf{76.2 \pm 5.6}$
Moon	160/40(2)	77.9 ± 5.7	$\mathbf{88.9 \pm 4.4}$
Diabetes	614/154(8)	63.4 ± 3.9	$\mathbf{68.7 \pm 3.2}$
Cancer	546/137 (10)	$\mathbf{96.4 \pm 1.4}$	93.7 ± 1.9
Liver	463/116(10)	53.8 ± 4.2	$\mathbf{59.6 \pm 4.2}$
Ionosphere	280/71(34)	72.9 ± 4.5	$\mathbf{83.7 \pm 4.3}$

Let us note that the accuracy of QNMC is significantly greater (especially for 2-feature datasets) than the accuracy of NMC for all the datasets, except for the Cancer dataset.

One key difference between NMC and QNMC concerns the invariance under rescaling. Let us suppose that each pattern of the training and test sets is multiplied by the same *rescaling factor* t, i.e. $\mathbf{x}_m \mapsto t\mathbf{x}_m$ and $\mathbf{y}_{m'} \mapsto t\mathbf{y}_{m'}$ for any m and m'. Then, the (classical) centroids change according to $\boldsymbol{\mu}_j \mapsto t\boldsymbol{\mu}_j$ and the classification problem of each pattern of the rescaled test set becomes

$$\operatorname*{argmin}_{i} d(t\mathbf{y}_{m'}, t\boldsymbol{\mu}_i) = t \operatorname*{argmin}_{i} d(\mathbf{y}_{m'}, \boldsymbol{\mu}_i), \tag{6}$$

which has the same solution of the unrescaled problem, i.e. $t = 1$.

On the contrary, QNMC turns out to be not invariant under rescaling. Far from being a shortcoming, this allows us to introduce a "free" parameter, i.e. the *rescaling factor*, that proves useful to get a further improvement of the classification performance for 2-feature datasets (see Ref. 16, Fig. 1, p. 7).

The method that we have introduced in this section allows us to get a relevant improvements of the standard NMC when we have an *a priori* knowledge about the distribution of the dataset we have to deal with. Indeed, if we need to classify an unknown pattern, looking at the distribution of the training dataset, we can guess *a priori* if: (i) for that kind of distribution QNMC performs better than NMC and (ii) what is the suitable rescaling has to be applied to the original dataset in order to get a further improvement of the accuracy.

3. Applying QNMC to the IPF Dataset

As mentioned at the beginning of the previous section, there are different ways to encode a d-dimensional feature vector into a density operator.[9] Indeed, finding the "optimal" encoding, if any, that outperforms for any dataset all possible encodings is still an open and intricate problem. This fact is not so surprising because in pattern recognition is not possible to establish an absolute and *a priori* superiority of a given classification method with respect to the other ones, the main reason being the fact

that every dataset has unique and specific characteristics (according to the well-known *No Free Lunch* Theorem[19]).

Another kind of encoding (called *informative encoding* (IE)) has been introduced in Refs. 17 and 18.

Let $\mathbf{x} = (x^1, \ldots, x^d) \in \mathbb{R}^d$ be a d-dimensional vector.

(1) We map the vector $\mathbf{x} \in \mathbb{R}^d$ into a vector $\mathbf{x}' \in \mathbb{R}^{d+1}$, whose first d features are the components of the vector $\mathbf{x}$ and the $(d+1)$th feature is the norm of $\mathbf{x}$. Formally,

$$\mathbf{x} = (x^1, \ldots, x^d) \mapsto \mathbf{x}' = (x^1, \ldots, x^d, |\mathbf{x}|). \tag{7}$$

(2) We obtain the vector $\mathbf{x}''$ by dividing the first d components of the vector $\mathbf{x}'$ by $|\mathbf{x}|$ as

$$\mathbf{x}' \mapsto \mathbf{x}'' = \left(\frac{x^1}{|\mathbf{x}|}, \ldots, \frac{x^d}{|\mathbf{x}|}, |\mathbf{x}|\right). \tag{8}$$

Now, similarly to Definition 1, we can define the notion of density pattern by informative encoding.

Definition 4 (Density pattern by IE).

$$\rho_{\mathbf{x}} := (\mathbf{x}'')^\dagger \cdot \mathbf{x}'', \tag{9}$$

where the vector $\mathbf{x}''$ is given by Eq. (8).

Accordingly, this encoding maps real d-dimensional vectors $\mathbf{x}$ into a $(d+1)$-dimensional pure state $\rho_{\mathbf{x}}$.

According to recent debates on quantum machine learning,[9] in order to avoid loss of information it is crucial that, in the transition from the classical to the quantum feature-space, the norm of the original feature-vector is explicitly incorporated in the resulting pure quantum state. It is not hard to see that both the SE and the IE keep track, albeit in a different way, of the vector norm of the original feature-vector and Eqs. (2) and (8) have the same limit when the rescaling factor t tends to infinity.

However, unlike the SE, the IE has the property to *explicitily* incorporates the information about the norm of the initial vector as one of its component (specifically, the last component).

As we have seen in the previous section, QNMC is the quantum counterpart of the standard NMC that is one of the more basic standard classifiers. Other well-known standard models that will be taken into account in the following are the so called *Linear Discriminant Analysis* (LDA) and *Quadratic Discriminant Analysis* (QDA) classifiers[19] that belong to the class of minimum distance classifiers. The main feature of such classifiers consists in classifying patterns by using a distance measure which involves not only the centroids of the classes but also the class distribution (by means of the *covariance matrix*[19]). The difference between LDA and QDA can be summarized as follows: (i) in the LDA case the distance measure depends on the average covariance matrix (over all the covariance matrices related to each class) and the

discriminant function (i.e. the surface which separates classes in the optimal way) is linear; (ii) in the QDA case, the distance measure depends on all the covariance matrices simultaneously and the discriminant function is quadratic. In what follows, we compare different variants of QNMC with the aforementioned classifiers (NMC, LDA, QDA) by referring to a very special real-world dataset obtained from a biomedical context: the *Idiopatic Pulmonary Fibrosis* (IPF) dataset.[d]

3.1. *The IPF dataset*

In detail, the IPF dataset includes a group of 126 consecutive patients (the patterns) retrospectively extracted from databases of the Regional Referral Centre for Interstitial and Rare lung diseases of Catania. These patients are divided into three different classes (with different cardinality), where each class corresponds to a different degree of survival (that is named GAP stage). All patients were required to have received a Multidisciplinary team diagnosis of IPF according to 2011 American Thoracic Society (ATS)/European Respiratory Society (ERS)/Japanese Respiratory Society (JRS)/Latin American Thoracic Association (ALAT) IPF guidelines.[21] A minimum follow-up time of three years from diagnosis was also required in order to assess survival. For this reason, only patients diagnosed between July 2010 and December 2014 were considered. The dataset includes a series of baseline variables (the features) with an established relation to survival (the classes, where three different survival "degrees" are considered).[22,23]

The dataset is organized in the following way: the patterns are numbered in the column A (we also indicate in the column B the dates of birth of each patient). We distinguish between two different blocks of features; the first block (from column C to column I, highlighted in light grey) contains features that allow to perfectly classify each patient; indeed, by using the features introduced in the columns C ... I, it is possible to exactly evaluate the "GAP stage" of each patient (each feature adds a score to the calculation of the GAP stage). Indeed, the features introduced from C to I are all it takes in order to assign to each patient the class to which he belongs to; in other words, these features are useful to have an *a priori* classification of each patient. The second block of features are introduced (in light green) from column J to column U; even if these features should allow to classify the patients, anyway — unlike the first block — there is not a systematic method to classify each patient by involving this set of features only. Finally, the column W contains the labels associate to each different class (the column V is only used as a support to calculate W). The rest of the paper will be devoted to use the introduced quantum-inspired algorithm to classify the IPF dataset, only involving the second block of features. But before let us briefly provide a medical description of the meaning of each feature. Regarding the first block, the feature "Forced Vital Capacity" (FVC) represents the amount of air which can be forcibly exhaled from the lungs after taking the deepest breath possible.[24] This

---

[d]The dataset is downloadable from http://people.unica.it/giuseppesergioli/files/2018/02/IPFDataset. xlsx.

value, measured with a spirometer, was reported in the dataset as percent of predicted value (FVC%), resulting from the comparison between a list of normal reference values and the measured ones.[24] In the context of IPF, both baseline FVC% value and its change over the time, represent strong predictors of mortality.[25,26] The feature "Diffusing Capacity for Carbon Monoxide" (DLCO), measures the ability of the lungs to transfer gas from inhaled air to the red blood cells in pulmonary capillaries.[27] As in the case of FVC, also DLCO is expressed as percent of predicted value. Interestingly in IPF, DLCO is frequently reduced since early stages of the disease, making this variable more sensitive than FVC to assess interstitial lung damage.[28] Another feature collected which significantly impacts on survival, as in IPF as in other diseases, is the "Age at first diagnosis".[21] Dataset also included the variable "Sex". Incidence and prevalence of IPF are higher in males than in females with a ratio ranging from 1.6:1 to 2:1. Moreover, male sex was demonstrated to be related with a worse prognosis.[21,29] All of these four features were recently included in a single multidimensional index, known as GAP (gender [G], age [A] and lung physiology variables [P]). This index assigns a point to each variable in order to obtain a single value, in the dataset "GAP point", which resumes the weight of each variable. Points ranging from 0 to 3, 4–5 and 6–8 compose respectively "GAP stage 1, 2 and 3",[30] that we consider as the label of our dataset. Simply speaking, the columns from F to I indicate the contribute in the calculation of the GAP stage provided by the features "Sex", "FVC", "DLCO" and "Age", respectively. Regarding the second block of features, Oxygen saturation (SpO2%) reflects blood oxygenation, and heart rate were indirectly measured with a pulse oximeter. Reduced levels of SpO2, which are frequently associated with high levels of heart rate, are usually related to a worse survival.[21] Information regarding smoking habit was also collected and reported as follows: never smoker = 0, ex-smoker = 1 and current smoker = 2. Dataset included also a description of high resolution computed tomography (HRCT) features which, according to 2011 IPF guidelines, describe three scenarios: "definite UIP", "possible UIP" and "inconsistent with UIP".[21] Recent studies demonstrated that also this evaluation at baseline is related with prognosis.[31] Other variables regarding information on lung transplantation, duration of follow-up (days), status at the end of follow-up (alive = 0 or died = 1), confirmation of diagnosis through biopsy and family history of the Interstitial Lung Disease (ILD) were also included in the dataset.

3.2. *Applying QNMC to the IPF dataset*

It is not hard to believe that each feature described above does not have the same impact on the evaluation of the GAP stage (i.e. in the classification process). As an example (confining to the second block of features only), it is possible to say that "Sex" and "Oxygen Saturation" have more impact in the classification process with respect to the rest of the considered features. In general, it is possible to recognize for each feature a different impact on the classification process.

As previously noticed, QNMC is not invariant under rescaling.[15,16,18] This seeming shortcoming can be beneficially used to model the different incidence of each feature. The strategy we adopt is to assign to each feature a rescaling factor that is proportional to its degree of incidence. Unlike the previous section, where all the dataset features were multiplied by the same rescaling real factor, here we multiply each feature by a different weight that depends on the incidence of the feature itself in the evaluation of the GAP stage. On this basis, the classification process will be remarkably sensitive to the introduction of this rescaling factor. Anyway, the method to assign the most suitable rescaling factor to each feature is essentially empirical and it can be re-arranged *in itinere* during the semi-supervised classification process.

After rescaling, the original dataset is transformed into the following (rescaled) dataset:

$$\mathcal{S}^{(r)} = \mathcal{S}^{(r)}_{\mathrm{tr}} \cup \mathcal{S}^{(r)}_{\mathrm{ts}}, \tag{10}$$

where

$$\mathcal{S}^{(r)}_{\mathrm{tr}} = \{(\gamma_1 \mathbf{x}_1, \lambda_1), \ldots, (\gamma_M \mathbf{x}_M, \lambda_M)\}, \quad \gamma_i \in \mathbb{R}, \ i = 1, \ldots, M,$$
$$\mathcal{S}^{(r)}_{\mathrm{ts}} = \{\{\delta_1 \mathbf{y}_1, \beta_1\}, \ldots, \{\delta_{M'} \mathbf{y}_{M'}, \beta_{M'}\}\}, \quad \delta_j \in \mathbb{R}, \ j = 1, \ldots, M'.$$

Finally, the quantum version $\mathcal{S}^{\mathrm{q}(r)} = \mathcal{S}^{\mathrm{q}(r)}_{tr} \cup \mathcal{S}^{\mathrm{q}(r)}_{ts}$ of the rescaled dataset is obtained by replacing (in Eq. (10)) all vectors $\mathbf{x}_i$, $\mathbf{y}_j$ with their corresponding quantum analogues $\rho_{\mathbf{x}_i}, \rho_{\mathbf{y}_j}$.

Table 2 allows us to compare the performances — in terms of classification error — of three standard classifiers — NMC, LDA and QDA — with the two variants of the quantum-inspired classifier. In detail, for each classifier we have evaluated the total error (with its respective standard deviation) obtained by running the algorithm 50 times for each different choice of the rescaling factor (each of them in accord with the different survival degree of the features of the dataset).

Table 2. Average classification error for NMC, QNMC (with difference encodings and different rescaling), LDA and QDA classifiers (in %) over 50 runs with related standard deviations.

Classifier	Total error
QNMC (SE)	0.455 ± 0.093
QNMC (IE)	0.378 ± 0.092
QNMC (IE) Resc 1	0.334 ± 0.097
QNMC (IE) Resc 2	0.341 ± 0.071
QNMC (IE) Resc 3	0.344 ± 0.076
QNMC (IE) Resc 4	$\mathbf{0.314 \pm 0.081}$
NMC	0.495 ± 0.085
LDA	0.393 ± 0.082
QDA	0.568 ± 0.119

As shown in Table 2, QNMC gives rise generally to a significant improvement of the accuracy in the classification process with respect to the three standard classifiers. Interestingly enough, the accuracy-values obtained for the third class are remarkable. In particular, the QNMC-variant based on the informative encoding exhibits better performance than NMC (about 12%) and QDA (where the difference is very high, about 24%). On the other hand, this version of QNMC performs in a similar way to LDA (the difference is about 2%). Since LDA is a classifier which takes into account the class distribution by means of the covariance matrix (i.e. we can say it is more "informative"), this result suggests that this version of QNMC is sensitive to the dataset distribution and, consequently, it gives a more accurate classification with respect to NMC, which does not take into account the data distribution.

Let us note that the "stereographic" QNMC provides a classification accuracy worse than the "informative" QNMC (about 8%). This result seems to suggests that the choice of the specific encoding is fundamental and strongly affects the performance of these (and possibly other) variants of QNMC.

The final result we would like to discuss concerns the use of the informative encoding together with different rescaling parameters for different features (according to the real different incidence of these features on the probability of survival). In particular, we have rescaled the feature columns "Follow Up Time (days)", "Oxygen saturation %" and "Heart rate" first by a rescaling parameter equal to 0.1 ("QNMC (IE) Resc 1"), after by a rescaling factor equal to 10 ("QNMC (IE) Resc 2") and finally by a rescaling factor equal to 20 ("QNMC (IE) Resc 3").[e] In this regard, we can observe a further improvement in terms of accuracy, up to a classification error equals to 0.33. The most interesting result is obtained by concurrently rescaling the feature columns "HRCT Pattern", "Smoking", "Smoking Status" by a parameter equal to 600 and the columns "Sex" and "Oxygen saturation %" by a parameter equal to 10. In this case, we reach a classification error equals to 31% ("QNMC (IE) Resc 4"), which is much lower than the NMC classification error (indeed, they differ approximately by 20%).

Let us remark that in the proposed approach, which consists in rescaling the feature columns by a real parameter in order to reach some computational benefits, we have adopted a systematic empirical procedure in order to get favorable rescaling parameters. Nevertheless, by the preliminary results shown in Table 2, it is possible to note that — in accord with the *a priori* assignment of the incidence of each feature — we obtain advantages in terms of classification performance by multiplying the more significant features by a higher rescaling parameter and the less significant ones by a lower rescaling parameter. Consequently, the rescaling factor can be thought of as a "weight" which somehow reflects the relevance of a specific feature column. All this suggests, as a future work, a theoretical analysis in order to systematically obtain the most convenient rescaling for each feature of a given dataset.

---

[e] Let us remark that we do not consider to apply the nonuniform rescaling of the dataset in case of classical classifiers.

We conclude the experimental sections with the following two remarks:

(1) Even if it is possible to establish whether a classifier is "good" or "bad" for a given dataset by the evaluation of some *a priori* data characteristics, generally it is no possible to establish an *absolute* superiority of a given classifier for any dataset, due to the No Free Lunch Theorem.[19] Anyway, QNMC seems to be particularly convenient when the data distribution is difficult to treat with the standard NMC;

(2) Clearly, there are more sophisticated (classical) classifiers than NMC, LDA and QDA for our IPF dataset. However, the preliminary results we have presented in this paper show that our quantum-inspired minimum distance model outperforms not only its natural classical counterpart (NMC) but also other more performing and sophisticated minimum distance methods.

4. Concluding Remarks

This paper is mostly devoted to show the potentialities of quantum formalism in the context of classification problems related to biomedical contexts. In particular, we have shown that some kinds of quantum-inspired classifiers remarkably outperform some standard classifiers (NMC, LDA, QDA) in the classification accuracy both for artificial and real-world datasets. In particular, in the second part of the paper we have focused on a very special dataset obtained by a real biomedical context.

As it is well-known, techniques used in biometrics are much more sophisticated than those presented in this work. Some biomedical contexts need, for example, classifiers that also include qualitative analysis, strictly depending on the dataset under investigation. Anyway, we think that the results of this paper open the way to new investigations of quantum formalism to biometric classification problems. In particular, in our future investigation we will consider three main directions: (i) the *quantization* of some standard classifiers that are more sophisticated and performing than the NMC; (ii) as we have remarked in the paper, the choice of the *optimal* encoding is strongly dataset-dependent. Anyway, this point deserves a further investigation. For example, it should be remarkable to identify some classes of datasets that, because of their internal properties, are more suitable to be treated with some encodings instead of others; (iii) finally, as we have seen, the quantum-inspired classification process we have considered is strongly based on the distribution of patterns. Hence, the role of the distance function is crucial. However, some datasets (like the IPF dataset) also contain features whose values cannot be ordered (for example *Sex* or *Smoking status*). Therefore, it should be useful to modify such datasets by replacing this unordered values with some ordered values preserving the statistical properties of the dataset themselves.

Acknowledgments

References

1. J. Eisert, M. Wilkens and M. Lewenstein, *Phys. Rev. Lett.* **83**(15) (1999) 3077.
2. E. W. Piotrowski and J. Sladkowski, *Int. J. Theo. Phys.* **42**(5) (2003) 1089.
3. D. Aerts and S. Sozzo, Quantum structure in cognition: Why and how concepts are entangled, in *Quantum Interaction QI 2011*, eds. D. Song, M. Melucci, I. Frommholz, P. Zhang, L. Wang and S. Arafat, Lecture Notes in Computer Science, Vol. 7052 (Springer, Berlin, Heidelberg, 2011), pp. 116–127.
4. D. Aerts, S. Sozzo, L. Gabora and T. Veloz, Quantum structure in cognition: Fundamentals and applications, *The Fifth International Conference on Quantum, Nano and Micro Technologies ICQNM 2011*, eds. V. Privman and V. Ovchinnikov (2011), arXiv:1104.3344.
5. E. Beltrametti, M. L. Dalla Chiara, R. Giuntini, R. Leporini and G. Sergioli, *Int. J. Theo. Phys.* **53**(10) (2014) 3293.
6. M. L. Dalla Chiara, R. Giuntini, R. Leporini and G. Sergioli, *Math. Slovaca* **66**(2) (2016) 313–334.
7. M. L. Dalla Chiara, R. Giuntini, R. Leporini, E. Negri and G. Sergioli, *Front. Psychol.* **6** (2015) 1583.
8. Y. C. Eldar and A. V. Oppenheim, *IEEE Signal Process. Magazine* **19**(6) (2002) 12.
9. M. Schuld, I. Sinayskiy and F. Petruccione, *Contemp. Phys.* **56**(2) (2014) 172.
10. M. Schuld, I. Sinayskiy and F. Petruccione, *Quantum Inf. Process.* **13**(11) (2014) 2567.
11. A. Manju and M. J. Nigam, *Artific. Intel. Rev.* **42**(1) (2014) 79.
12. P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining* (Academic Press, 2014).
13. S. Lloyd, M. Mohseni and P. Rebentrost, arXiv:1307.0411.
14. S. Lloyd, M. Mohseni and P. Rebentrost, *Nature Phys.* **10**(9) (2014) 631.
15. G. Sergioli, E. Santucci, L. Didaci, J. A. Miszczak and R. Giuntini, *Soft Comput.* **22**(3) (2018) 691.
16. G. Sergioli, G. M. Bosyk, E. Santucci and R. Giuntini, *Int. J. Theo. Phys.* **56**(12) (2017) 3880.
17. E. Santucci and G. Sergioli, Classification problem in a quantum framework, in *Quantum Foundations, Probability and Information*, eds. A. Khrennikov and B. Toni (Springer, 2018), pp. 215–229.
18. E. Santucci, *Entropy* **19**(12) (2017) 659.
19. R. O. Duda, P. E. Hart and D. G. Stork (eds.), *Pattern Classification*, 2nd edn. (Wiley Interscience, 2000).
20. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information — 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).
21. G. Raghu, H. R. Collard, J. J. Egan *et al.*, *Am. J. Respir. Crit. Care Med.* **183** (2011) 788.
22. S. Puglisi, S. E. Torrisi, R. Giuliano, V. Vindigni and C. Vancheri, *Semin. Respir. Crit. Care Med.* **37** (2016) 358.
23. S. Puglisi, S. E. Torrisi, V. Vindigni, R. Giuliano, S. Palmucci, M. Mule and C. Vancheri, *Ther. Adv. Chronic Dis.* **7** (2016) 108.

24. M. R. Miller, J. Hankinson, V. Brusasco *et al.*, *Eur. Respir. J.* **26** (2005) 319.
25. R. M. du Bois *et al.*, *Am. J. Respir. Crit. Care Med.* **184** (2011) 1382.
26. S. E. Torrisi, M. Pavone, A. Vancheri and C. Vancheri, *Eur Respir. Rev.* **26** (2017).
27. N. Macintyre, R. O. Crapo, G. Viegi *et al.*, *Eur. Respir. J.* **26** (2005) 720.
28. R. M. du Bois *et al.*, *Am. J. Respir. Crit. Care Care Med.* **184** (2011) 459.
29. M. K. Han, S. Murray, C. D. Fell, K. R. Flaherty, G. B. Toews, J. Myers, T. V. Colby, W. D. Travis, E. A. Kazerooni, B. H. Gross and F. J. Martinez, *Eur. Respir. J.* **31** (2008) 1183.
30. B. Ley, C. J. Ryerson, E. Vittinghoff, J. H. Ryu, S. Tomassetti and J. S. Lee, *Annal. Inter. Med.* (2013).
31. M. L. Salisbury, L. B. Tolle, M. Xia, S. Murray, N. Tayob, A. M. Nambiar, S. L. Schmidt, A. Lagstein, J. L. Myers, B. H. Gross, E. A. Kazerooni, B. Sundaram, A. R. Chughtai, F. J. Martinez and K. R. Flaherty, *Respir. Med.* **131** (2017) 229.
32. S. Gambs, arXiv:0809.0444v2 [quant-ph].
33. C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, 1976).
34. F. Holik, G. Sergioli, H. Freytes and A. Plastino, *Foundat. Sci.* **23**(1) (2018) 119.
35. E. Aïmeur, G. Brassard and S. Gambs, Machine learning in a quantum world, *Conf. Canadian Society for Computational Studies of Intelligence* (Springer, Berlin, Heidelberg, 2006), pp. 431–442.