Scientific Project

Optical Implementation of Quantum-Inspired and Classical Classifiers with a Joint Transform Correlator

João Guilherme
(105939)

Supervisors:

Prof. Emmanuel Zambrini Cruzeiro

Prof. Hugo Terças

June 23, 2025

Abstract

We present optical implementations of classical and quantum-inspired nearest-mean classifiers using a Joint Transform Correlator (JTC). The quantum-inspired classifiers encode classical feature vectors into quantum states, enabling the use of trace distance as a similarity measure. We propose and evaluate several quantum encodings—standard amplitude, informative, and inverse stereographic—and demonstrate a method to estimate trace distance optically via fidelity measurements from the JTC. Numerical simulations on the MNIST dataset confirm that quantum-inspired methods outperform classical baselines, achieving balanced accuracies up to 82.4% with the informative encoding. Experimental validation using a simple, phase-only SLM and single-lens optical setup confirms the feasibility of implementing this quantum-inspired classifier in hardware, highlighting its potential for fast and energy-efficient analog computation.

I. Introduction

Quantum-inspired algorithms are classical routines that use the mathematical formalism and physical intuitions from quantum mechanics — such as superposition-style probability amplitudes, interference or quantum state discrimination — yet run entirely on conventional hardware. These algorithms have multiple applications, including machine learning, where they promise faster convergence or reduced dimensional dependence compared with traditional methods.

Quantum-inspired algorithms often outperform classical ones in software, but optical implementations can offer further advantages — particularly for high-dimensional data streams or large-scale datasets—by leveraging speed, parallelism, and energy efficiency ([8],[4]).

These advantages arise not simply because light travels fast, but because optical systems can naturally perform certain operations, such as the Fourier transform, in constant time via light propagation and using lenses. Optical setups can also exploit massive spatial parallelism, allowing many operations to be carried out simultaneously. In contrast to digital processors that are limited by clocking, thermal dissipation, and wiring constraints, optical systems offer nearly dissipationless computation through physical propagation, supporting high processing power with minimal energy loss. This makes them especially well-suited for scalable, real-time machine learning tasks ([7]).

Many machine learning algorithms rely on distance-based decision rules, where a query sample is compared to a set of class prototypes using similarity measures such as dot products or Euclidean distance. A growing line of research explores how such similarity computations can be accelerated or parallelized using optical systems ([3]).

Among these, the Joint Transform Correlator (JTC) stands out as a particularly attractive architecture for practical implementation ([5], [9]). Its appeal lies in its minimal hardware requirements: the system can be realized with only a phase-modulating Spatial Light Modulator (SLM) and a single lens to perform the optical Fourier transform. This makes it compact, cost-effective, and relatively easy to align compared to more complex systems like VanderLugt correlators ([11]), which require pre-fabricated matched filters and precise placement.

The JTC also offers valuable computational properties. Because it performs correlation in the Fourier domain, it naturally exhibits translation invariance - shifting the input patterns does not significantly affect the output peak amplitude. This makes it well-suited for comparing data with spatial variations, such as digit images or other structured patterns.

In this work we propose optical implementations of the Quantum-Inspired and Classical Nearest Mean Classifier ([10], [2]) that use a single SLM and a 1-f lens system ([5]).

We begin by presenting the classical and quantum-inspired classifiers considered in this work (Section II). We then describe the Joint Transform Correlator (JTC) (Section III), establishing its theoretical connection to distance-based similarity measures. In Section IV, we show explicitly how the JTC can optically compute both classical distances and quantum-inspired measures, such as the trace distance. In Section V we detail our experimental implementation, demonstrating a realization of the quantum-inspired nearest-mean classifier (QNM-C) using a single phase-only spatial light modulator (SLM) and a 1-f lens system. Finally, we summarise our findings and discuss future directions of work in Section VI and Section VII.

II. Supervised learning and quantum-inspired classification

i. Supervised learning

Machine learning (ML) provides algorithmic tools that *learn* directly from data rather than relying on hand-crafted rules. In the *supervised* setting, learning is guided by *examples* whose correct answers are known in advance. We review the key concepts used throughout this work.

Input data. We start from N *raw samples* $\mathbf{u}_i \in \mathbb{R}^p$, $i = 1, \ldots, N$, each accompanied by a *label* $y_i \in \{1, \ldots, K\}$. The labels partition the data into

$$C_k = \{\mathbf{u}_i : y_i = k\}, \qquad N_k = |C_k|, \qquad p_k = \frac{N_k}{N},$$

where N_k and p_k are the size and empirical probability of class k, respectively. It is convenient to organise the entire data set as a matrix $\mathbf{M} = [\mathbf{u}_1 \; \mathbf{u}_2 \; \cdots \; \mathbf{u}_N] \in \mathbb{R}^{p \times N}$ and the labels as a vector $\mathbf{y} = [y_1, \ldots, y_N]^\top$.

Training and test sets. A *training set* $(\mathbf{M}_{\text{train}}, \mathbf{y}_{\text{train}})$ is used to tune the parameters of a classifier, while a disjoint *test set* $(\mathbf{M}_{\text{test}}, \mathbf{y}_{\text{test}})$ gauges its ability to generalise to unseen data. The algorithm is trained on the training set and evaluated on the test set. In this work we will

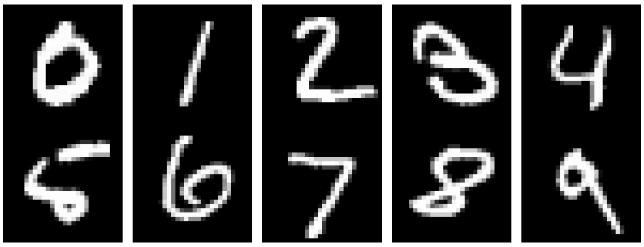


Figure 1: One representative example per class from the MNIST dataset. Each 28×28 grayscale image is flattened to a $d = 784$-dimensional feature vector before the pre-processing (centering, standardisation, PCA).

be using the MNIST dataset (Fig.1), with 60000 training samples and 10000 testing samples.

Pre-processing. Real-world data rarely arrives in a form that is optimal for learning. We can apply three light transformations that are standard in pattern recognition:

(a) *Centring* removes global offsets, $\mathbf{u}_i \longmapsto \mathbf{u}_i - \bar{\mathbf{u}}$, where $\bar{\mathbf{u}} = N^{-1} \sum_{j=1}^{N} \mathbf{u}_j$.

(b) *Standardisation* rescales each feature so that every dimension has unit variance, preventing attributes with large numerical ranges from dominating the learning process.

(c) *Principal-component analysis* (PCA) projects the centered, standardised vectors onto the $p'(< p)$ directions of greatest variance, $\mathbf{x}_i = \Pi_{\mathrm{PCA}} \mathbf{u}_i \in \mathbb{R}^{p'}$, thereby discarding redundant noise and shrinking the effective dimensionality. In the case of the MNIST dataset, the PCA will mostly remove the border pixels, that are always black, as they do not contribute to the variance of the data.

We apply this pre-processing pipeline to the MNIST dataset in order to reduce the dimensionality of the data and improve the performance of the classifiers.

Balanced accuracy. Performance is summarised through the *confusion matrix* $V_{k'k}$, whose entry counts how many test samples belonging to class k are classified as k'. The *balanced accuracy* (BA)

$$\mathrm{BA} = \frac{1}{K} \sum_{k=1}^{K} \frac{V_{kk}}{N_k} \tag{1}$$

computes the mean of the per-class accuracies and is robust to class imbalance. When classes are homogeneous in size, BA reduces to the usual accuracy.

Classical Nearest-Mean Classifier (CNM-C). The nearest-mean classifier (NMC) is a simple, interpretable classifier that assigns a label to a test sample based on the *centroid* of the training samples in each class. The *centroid* of class k is defined as the mean of the training samples in that class:

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{u}_i, \qquad \text{where } \mathbf{u}_i \in C_k. \tag{2}$$

The CNM classifier assigns a test sample $\mathbf{x}$ to the class j if:

$$d(\mathbf{x}, \boldsymbol{\mu}_j) = \min_{k=1,\ldots,K} d(\mathbf{x}, \boldsymbol{\mu}_k) \tag{3}$$

where the distance $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ is the Euclidean distance. The CNM classifier is simple to implement, as it only requires the computation of the centroids and K distance calculations, one for each centroid.

ii. Quantum-inspired classification

Quantum-inspired machine learning uses the mathematical formalism native to quantum mechanics - density operators, trace distance, state discrimination — but applies them on ordinary CPUs/GPUs. By embedding classical feature vectors in a Hilbert space, one can exploit the geometry of quantum states to obtain alternative similarity measures that sometimes improve accuracy or robustness compared with purely classical methods ([2], [10]). Every quantum-inspired algorithm must start with a function that maps each classical data vector to a density operator ρ_x. The choice of this encoding is very important, as it determines the properties of the resulting quantum-inspired algorithm.

Density–pattern encodings. We must define a a map that sends a real feature vector $\mathbf{x} \in \mathbb{R}^d$ to a *density operator* $\rho_{\mathbf{x}} = \tilde{\mathbf{x}}\, \tilde{\mathbf{x}}^\top$. There are many different ways of doing this mapping, for example:

(a) Standard amplitude (SA) encoding.
Normalise the input vector to unit length:

$$\tilde{\mathbf{x}}^{(\mathrm{SA})} = \begin{cases} \dfrac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ (1, 0, \ldots, 0)^\top, & \mathbf{x} = \mathbf{0}. \end{cases}$$

This encoding maps the input vector to a unit vector, which is then used to construct the density operator. We lose information about the norm of the input vector, as it is always set to 1, but the direction of the vector is preserved.

(b) Inverse stereographic (IS) encoding.
Embed $\mathbf{x}$ into $\mathbb{R}^{d+1}$ through the inverse stereographic projection onto the unit sphere:

$$\tilde{\mathbf{x}}^{(\mathrm{IS})} = \frac{1}{1 + \|\mathbf{x}\|^2} \Big(2x_1, \ldots, 2x_d, \|\mathbf{x}\|^2 - 1 \Big)^\top.$$

In contrast to the standard amplitude map, the inverse-stereographic encoding retains both radial (norm) and directional information. The first d coordinates are proportional to the original components x_k, so the orientation of $\mathbf{x}$ is still recognisable, while the $(d+1)$-st coordinate $\|\mathbf{x}\|^2 - 1$ embeds the vector's magnitude. Because the whole $(d+1)$-vector is subsequently normalised, every input is sent to a point on the unit sphere S^{d+1}; the origin maps to the "south pole" $(0, \ldots, 0, -1)^\top$, whereas vectors with large norm accumulate near the "north pole", hence preserving norm information in a non-linear way. The mapping is conformal (it preserves local angles) but non-linear, compressing large norms and expanding small ones, which can improve cluster separability in the quantum feature space while delivering valid, unit-norm quantum states for all classical inputs.

(c) Informative (INF) encoding [10].

Attach the norm as an extra amplitude so that both the direction and *length* of a real vector $\mathbf{x} \in \mathbb{R}^d$ affect its quantum representation:

$$\tilde{\mathbf{x}}^{(\mathrm{INF})} = \begin{cases} \left(\frac{x^1}{\|\mathbf{x}\|\sqrt{\|\mathbf{x}\|^2+1}}, \ldots, \frac{\|\mathbf{x}\|}{\sqrt{\|\mathbf{x}\|^2+1}}\right)^\top, & \mathbf{x} \neq \mathbf{0}, \\ (1, 0, \ldots, 0)^\top, & \mathbf{x} = \mathbf{0} \end{cases}$$

Hence, this encoding maps real d-dimensional vectors $\mathbf{x}$ into a $(d+1)$-dimensional vector. In this way we obtain an embedding that retains the information about the original vector norm and also preserves the information about the direction.

For any of the above encodings, the corresponding *density operator* is obtained by taking the outer product of the encoded vector with itself:

$$\rho_{\mathbf{x}} = \tilde{\mathbf{x}}\,\tilde{\mathbf{x}}^\top \tag{4}$$

This guarantees $\rho_{\mathbf{x}}$ is positive semidefinite, unit-trace, and idempotent ($\rho_{\mathbf{x}}^2 = \rho_{\mathbf{x}}$), hence a valid pure quantum state. They differ in how much geometric information (direction, norm) is retained, a factor that can significantly affect classification performance.

Quantum centroid. For each class k with training subset $\mathcal{S}_k^{\mathrm{tr}}$, define its *quantum centroid* as the sum of the density operators from the class, divided by the number of elements in the class:

$$\varrho_k = \frac{1}{|\mathcal{S}_k^{\mathrm{tr}}|} \sum_{(\mathbf{x}_i) \in \mathcal{S}_k^{\mathrm{tr}}} \rho_{\mathbf{x}_i} \tag{5}$$

Unlike the classical centroid, ϱ_k is usually a *mixed* state and has no direct counterpart in the original feature space.

Quantum distance measure. Similarity between two density patterns is quantified by the *trace distance*,

$$d_{\mathrm{tr}}(\rho, \sigma) = \tfrac{1}{2}\,\mathrm{Tr}|\rho - \sigma|, \tag{6}$$

where $|A| = \sqrt{A^\dagger A}$. Note that the trace distance defines a metric; it satisfies the triangle inequality, $d_{tr}(\rho, \sigma) \geq 0$ and $d_{tr}(\rho, \sigma) = d_{tr}(\sigma, \rho)$, and is equal to zero if and only if $\rho = \sigma$.

Having defined the quantum centroid and quantum distance, we can now define the quantum-inspired version of the nearest-mean classifier.

Quantum-inspired nearest-mean classifier (QNM-C). The QNM-C is the exact same as the CNM-C, but uses the quantum centroid and quantum distance instead of the classical centroid and Euclidean distance. The algorithm is as follows:

(a) Encode every training sample $\mathbf{x}_i$ as a density operator $\rho_{\mathbf{x}_i}$ using one of the encodings described in ii.

(b) Compute the quantum centroid for each class k:

$$\varrho_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \rho_{\mathbf{x}_i} \tag{7}$$

(c) For each test density operator $\rho_{\mathbf{x}}$, assign the class k if

$$d_{\mathrm{tr}}(\rho_{\mathbf{x}}, \varrho_k) = \min_{j=1,\ldots,K} d_{\mathrm{tr}}(\rho_{\mathbf{x}}, \varrho_j) \tag{8}$$

Having defined both the classical and quantum-inspired nearest-mean classifiers, we can now describe the joint transform correlator (JTC) and how it can be used to implement these classifiers optically.

III. Joint Transform Correlator

i. Classical Joint Transform Correlator

A *joint transform correlator* (JTC) is an optical architecture that performs the cross– and auto–correlation of two 2-D patterns by means of only *free-space propagation* and *Fourier transform* lenses. Unlike classical Vander Lugt correlators, the JTC does not need a pre-fabricated matched filter: both the *reference* pattern $s_1(x, y)$ and the *query* pattern $s_2(x, y)$ are displayed side-by-side in the same input plane [3].

Operational principle. Let the two patterns be centered on the SLM display (phase modulation only) at $\pm x_0$ along the horizontal axis, i.e. $s_1(x + x_0, y)$ and $s_2(x - x_0, y)$. A $1f$-configuration lens (focal length f) produces at its back focal plane the joint Fourier transform, given by:

$$\begin{aligned} G(\alpha, \beta) = {} & S_1\!\left(\tfrac{\alpha}{\lambda f}, \tfrac{\beta}{\lambda f}\right) e^{-i2\pi x_0 \alpha/(\lambda f)} \\ & + S_2\!\left(\tfrac{\alpha}{\lambda f}, \tfrac{\beta}{\lambda f}\right) e^{+i2\pi x_0 \alpha/(\lambda f)} \end{aligned} \tag{9}$$

where (α, β) are spatial-frequency coordinates and S_j denotes the Fourier transform of s_j. A CMOS camera records and stores the *joint power spectrum* $|G|^2$; its intensity contains four terms:

$$\begin{aligned} |G|^2 = {} & |S_1|^2 + |S_2|^2 \\ & + S_1 S_2^* \, e^{-i4\pi x_0 \alpha/(\lambda f)} \\ & + S_2 S_1^* \, e^{+i4\pi x_0 \alpha/(\lambda f)} \end{aligned} \tag{10}$$

The SLM displays the joint power spectrum and the same lens takes the inverse Fourier transform of this intensity pattern. This produces at the fourier plane the correlation signals of the two images, given by:

$$\begin{aligned} g(x', y') &= R_{11}(x', y') + R_{22}(x', y') \\ &\quad + R_{12}(x' - 2x_0, y') + R_{21}(x' + 2x_0, y') \end{aligned} \tag{11}$$

where (x', y') are the coordinates at the correlation plane and the correlation signals are defined as:

$$R_{ij}(x', y') = \iint s_i(x' - x,\ y' - y)\, s_j(x, y)\, dx\, dy \tag{12}$$

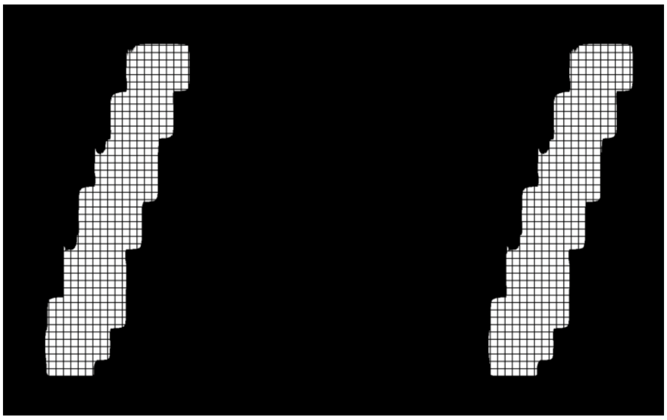
The constant terms yield *autocorrelation* peaks near the optic axis, whereas the two cross-terms produce a pair of off-axis peaks located at $(\pm 2x_0, 0)$ whose heights are proportional to the cross-correlation $s_1 \star s_2$. Therefore, we just need to measure the intensity of the cross-correlation peaks to determine the similarity between the two patterns. To obtain a normalized value between 0 and 1, we can divide the intensity of the cross-correlation peaks by the intensity of the auto-correlation peak. This normalization ensures that the resulting value is independent of the absolute intensity levels and only reflects the relative similarity between the two patterns. This gives us a value of 1 when the two images are identical and 0 when they are completely different.

This is the classical version of the JTC. There have been proposals of introducing some non-linearity both at the fourier plane and input plane, and that improves the performance of the JTC ([9], [5], [3], [1]).

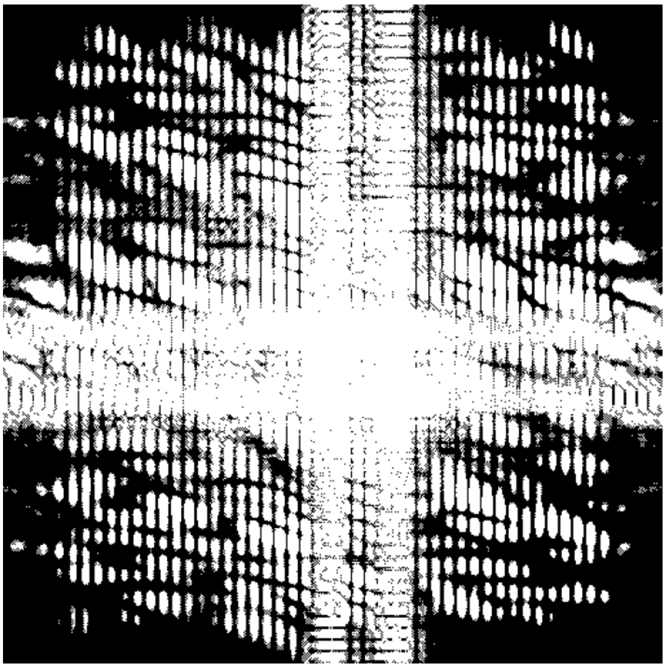
ii. Binary Joint Transform Correlator

Javidi & Horner's single–SLM Binary JTC [5] applies a hard threshold at two points of the optical cycle and works entirely with a bipolar alphabet $\{-1, +1\}$. At the input plane the grey-scale reference and query patterns are turned into a $-1/+1$ map by comparing each pixel with the median of its histogram (Fig. 2a). On a phase-only SLM those two levels translate into a 0 or π phase delay, so no optical power is wasted. After a first Fourier lens the camera records the joint power spectrum. That spectrum is thresholded in exactly the same fashion and written back to the same SLM during a second exposure (Fig. 2b), where a second Fourier lens produces the correlation plane (Fig. 2c).

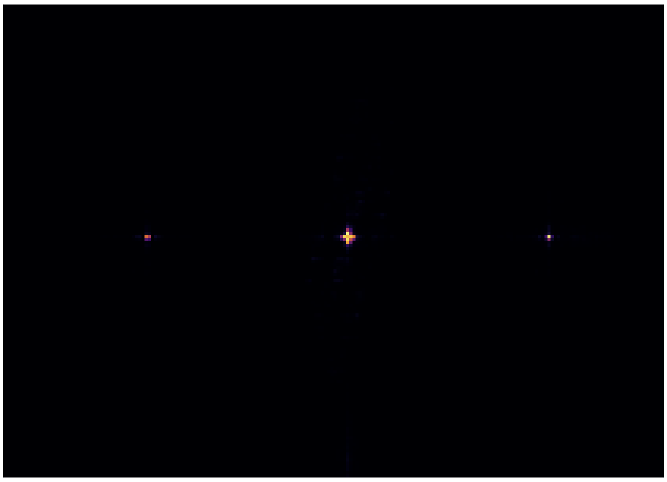
The double threshold approach has two immediate benefits. First, it transforms autocorrelation terms into sharp, delta-like spikes (Fig. 2c), which clears space on the detector for larger reference windows and removes the placement constraints typically imposed by the classical JTC. Second, it ensures the correct phase of the first harmonic is preserved through the non-linearity, while higher harmonics become both weaker and spatially separated; a simple spatial filter can then isolate the desired term. Besides those, the binary JTC has more benefits: it substantially increases the correlation-peak intensity, enhances the peak-to-sidelobe ratio, sharpens the correlation peak, and improves the signal-to-noise ratio compared to the classical JTC under the same illumination conditions, without requiring any additional hardware or optical components.



(a) Binary input images of the JTC, notice the grid introduced by the fill-factor of the SLM.



(b) Binary joint power spectrum produced after tresholding the joint power spectrum measured by the camera.



(c) Correlation plane intensity, showing the cross-correlation peaks at the off-axis positions.

Figure 2: Python-based optical simulation of a binary-input Joint Transform Correlator, implemented with LightPipes [6] for realistic wave-propagation and Fourier-transform modeling. The fill factor of the SLM is being simulated by an added grid pattern.

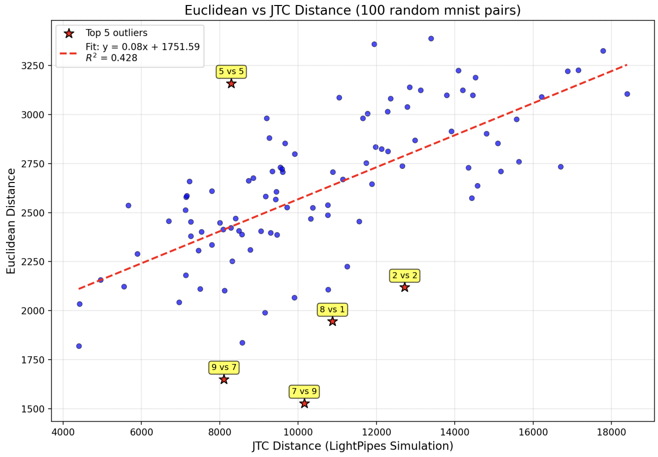


Figure 3: Scatter plot of the JTC dissimilarity versus the Euclidean distance for 100 random pairs of MNIST digits. The coefficient of determination is $R^2 = 0.428$. While the R^2 value is relatively low, multiple outliers contribute to this reduction. However, the general trend of the data is well captured by the fitted line, indicating a meaningful correlation between JTC dissimilarity and Euclidean distance.

iii. JTC-based dissimilarity

Let $\mathrm{sim}(s_1, s_2) \in [0, 1]$ denote the normalised height of the cross-correlation peak produced by the JTC when the input patterns are s_1 and s_2. We quantify dissimilarity through its reciprocal,

$$d_{\mathrm{JTC}}(s_1, s_2) = \frac{1}{\mathrm{sim}(s_1, s_2)} \tag{13}$$

Because sim = 1 for identical patterns and decreases towards 0 as the patterns diverge, d_{JTC} is non-negative and symmetric, but it is *not* a metric in the strict mathematical sense: it does not satisfy the triangle inequality and it is not zero for equal images. Nonetheless it serves the same practical purpose as a distance — patterns that are more alike yield a smaller d_{JTC}, and that is the most important property for the classification task.

We can see in Figure 3 that the JTC dissimilarity is moderately correlated with the Euclidean distance, with a coefficient of $R^2 = 0.428$. There are some outliers that contribute to this low correlation, but the general trend of the data is well captured by the fitted line. This indicates that the JTC dissimilarity is a meaningful measure of dissimilarity between patterns, and it can be used as a distance measure for the CNM-C.

IV. Optical implementations of the classifiers

i. Optical implementation of the CNM-C

The optical version of the classical nearest-mean classifier keeps all training-time work in software: after the usual pre-processing pipeline (PCA, centring, standardisation) each class centroid μ_k is computed numerically on the CPU and stored once. At inference time only the similarity calculation is off-loaded to the JTC. A query vector $\mathbf{x}$ is encoded on the left half of the SLM, while a stored centroid pattern μ_k is written on the right. The two-pass JTC then produces a single normalised correlation peak whose inverse, $d_k = 1/\,\mathrm{sim}(\mathbf{x}, \mu_k)$, plays the role of the Euclidean distance.

By cycling through the K stored centroids—two camera exposures per centroid—the system obtains all d_k values optically and selects the class with the smallest distance. The result is a hardware-minimal classifier: training remains purely digital, while the costly dot- product stage is performed optically.

ii. Optical implementation of the QNM–C

In Sec. ii we defined the QNM-C. There, the decision rule was formulated in terms of the *trace distance*, yet this quantity is not directly accessible in a classical–optical setup: a JTC yields a scalar similarity signal rather than the operator norm $\frac{1}{2}\|\rho - \sigma\|_1$. So in order to implement the QNM-C we need to find a method to use the JTC to obtain an estimate of the trace distance, which we can then use to classify the test samples. We will start by defining the purity of a quantum state, then the fidelity between two quantum states, and finally relate the fidelity to the trace distance. This will allow us to use the JTC to estimate the trace distance and implement the QNM-C classifier.

Purity. For a single state the purity

$$P(\rho) = \mathrm{Tr}(\rho^2) \tag{14}$$

ranges from 1 (pure state) to $1/d$ (maximally mixed state in Hilbert space of dimension d).

Fidelity. For two states ρ, σ the fidelity is defined as[1]

$$F(\rho, \sigma) = \left[\mathrm{Tr}\left(\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\right)\right]^2 \tag{15}$$

When both states are pure, $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$, this simplifies to

$$F(\rho, \sigma) = |\langle\psi \mid \phi\rangle|^2 \tag{16}$$

Because the JTC similarity is proportional to $\langle\psi \mid \phi\rangle$, its squared magnitude is directly proportional to the fidelity of two pure optical encodings. This is only true for pure states, for mixed states the fidelity cannot be calculated directly from the JTC similarity, which can be a problem if the quantum centroids have a low purity.

Trace distance from fidelity. The **Fuchs–van de Graaf inequalities** relate fidelity and trace distance D_{tr}:

$$1 - \sqrt{F(\rho, \sigma)} \;\le\; D_{\mathrm{tr}}(\rho, \sigma) \;\le\; \sqrt{1 - F(\rho, \sigma)} \tag{17}$$

If one of the states is pure (in our case, the input vectors are always pure, by construction), the trace distance can be expressed in terms of the fidelity:

$$D_{\mathrm{tr}}(\rho, \sigma) = \sqrt{1 - F(\rho, \sigma)} \tag{18}$$

[1] See, e.g. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th anniv. ed., 2010.

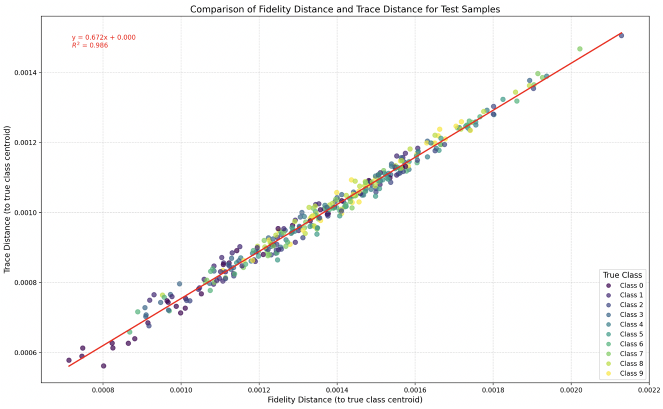


Figure 4: Trace distance vs. fidelity distance computed with the JTC using the inverse stereographic encoding. The R^2 value of 0.986 indicates a strong linear relationship between the two measures, confirming that the JTC can effectively estimate the trace distance.

Denoting the JTC similarity output by $\mathrm{sim}(\rho, \sigma) = |\langle \psi \mid \phi \rangle|$, the trace distance between a quantum centroid ϱ and an input state ρ can be expressed as

$$D_{\mathrm{tr}}(\varrho, \rho) = \sqrt{1 - \mathrm{sim}(\varrho, \rho)^2} \tag{19}$$

It can be seen in Figure 4 that, for the inverse stereographic encoding, the trace distance and fidelity are strongly correlated, with an R^2 value of 0.986. This confirms that the JTC can effectively estimate the trace distance from the fidelity, allowing us to use the JTC to implement the QNM-C. This calculation of the trace distance using the JTC is not exclusive to the QNM-C classifier, it can be used by any quantum-inspired classifier that uses the trace distance to perform classification.

Uhlmann's theorem. The main difficulty in using this approach to compute the trace distance is that we can only evaluate the Fidelity (and therefore the trace distance) directly with the JTC for pure states, whereas quantum centroids are generally mixed states (Table 1). However, Uhlmann's theorem allows us to calculate the fidelity between two mixed states by using their purifications. A purification is the process of representing a mixed state as a pure quantum state of higher-dimensional Hilbert space. Uhlmann's theorem states that for any two density operators ρ and σ, there exist purifications $|\psi_p\rangle$ of ρ and $|\phi_p\rangle$ of σ such that

$$F(\rho, \sigma) = \max_{|\psi_p\rangle, |\phi_p\rangle} |\langle \psi_p \mid \phi_p \rangle|^2 \tag{20}$$

where the maximum is taken over all purifications of ρ and σ. This means that we can compute the trace distance between two mixed states by computing the fidelity between their purifications, which can be done using the JTC.

Although each quantum centroid is, in general, a mixed state, we can still evaluate its trace-distance to a (pure) input state by first selecting an appropriate purification for the centroid. With that purification, the JTC can be used to calculate the trace distance between the centroid and the input, and that distance then serves as our classification metric.

Implementing this calculation based on Uhlmann's theorem lies beyond the scope of the present study; we mention it only as a promising direction for future improvements to this implementation of the QNM-C.

Table 1: Centroid purity for each MNIST class under the three encodings described in Sec. ii.

Class	Standard	Informative	Stereographic
0	0.1776	0.4623	1.0000
1	0.3186	0.5714	1.0000
2	0.0901	0.3682	1.0000
3	0.1052	0.3842	1.0000
4	0.0990	0.3863	1.0000
5	0.0830	0.3270	1.0000
6	0.1321	0.4133	1.0000
7	0.1493	0.4320	1.0000
8	0.0791	0.3454	1.0000
9	0.1065	0.3758	1.0000

Choosing the correct encoding. The encoding used has an impact on the performance of the QNM-C classifier, as can be seen in Table 2. For the optical version this is especially true, since the choice of the encoding impacts the purity of the quantum centroids (Table 1), which in turn affects the fidelity and trace distance calculations. For our case, we are doing PCA to reduce the number of features to $d = 50$, so the maximally-mixed state would have a purity of $P = 1/50 = 0.02$. That is much lower than the purity of the quantum centroids obtained with the three encodings, which are generally all above $P = 0.1$, so we expect that the JTC can effectively estimate the trace distance for all three encodings.

iii. Computational simulations

We started by creating a simulation of the JTC using Python and the LighPipes library [6], which allows us to simulate the propagation of light through optical systems. The simulation was used to validate the performance of the JTC and the classifiers before implementing them in hardware.

We ran all the classifiers using the MNIST dataset using the pre-processing pipeline described in Sec. II. The PCA reduced the number of features to $d = 50$. The results are presented in Table 2.

From the results in Table 2 we can see that all quantum-inspired variants surpass the classical baselines, especially when using the JTC, and two patterns emerge:

(i) Direction matters more than norm for the trace metric. The *standard-amplitude* (SA) map - which throws away the vector length and keeps only its direction - delivers the highest balanced accuracy when the *exact* trace distance is used (85.8 %). Augmenting the state with explicit norm information, either by the extra coordinate of the *informative* (INF) map or by the quadratic term of the *inverse-stereographic* (IS) map, does not improve performance.

Table 2: Balanced accuracy (% mean ± std.) on MNIST. For the CNM-C the encoding is not applicable, so we leave it empty. For the QNM-C we show the results for the three encodings described in Sec. ii. The Fidelity distance is being calculated using the software implementation of the JTC. The two best results are in **bold**.

Classifier	Metric	Encoding	Balanced Acc.
CNM-C	Euclidean	–	80.38 ± 0.38
CNM-C	JTC	–	72.42 ± 0.55
QNM-C	Trace	Standard	$\mathbf{85.84 \pm 0.38}$
QNM-C	Fidelity	Standard	78.77 ± 0.34
QNM-C	Trace	Stereographic	80.41 ± 0.41
QNM-C	Fidelity	Stereographic	80.74 ± 0.44
QNM-C	Trace	Informative	81.26 ± 0.37
QNM-C	Fidelity	Informative	$\mathbf{82.39 \pm 0.40}$

Table 3: Average centroid purity and absolute BA gap between the trace and JTC-fidelity scores (values from Table 2).

Encoding	Mean purity P	$\lvert \mathrm{BA}_{\mathrm{trace}} - \mathrm{BA}_{\mathrm{fid}} \rvert$
Standard (SA)	0.18	7.1 pp
Informative (INF)	0.41	1.1 pp
Stereographic (IS)	1.00	0.3 pp

(ii) Higher centroid purity narrows the trace–fidelity gap. Fidelity is what the JTC can estimate directly, whereas the trace distance is computed in software. The closer a centroid is to a *pure* state, the closer those two metrics become, and the smaller the difference between their BAs. Table 3 makes this relationship explicit.

Low-purity centroids in the SA encoding make the JTC-derived fidelity a worse estimate for the trace distance, hence a large performance gap. With IS, centroids are pure ($P = 1$), so the estimation of the fidelity is better and the gap almost vanishes. INF sits in between on both counts.

In short, SA wins when the true trace distance is used, hinting that direction dominates class separation, while high-purity encodings (IS, INF) make the JTC fidelity metric a better replacement of the trace distance, giving more accurate results. This results highlights the importance of the encoding choice in quantum-inspired classifiers, especially when using optical implementations like the JTC.

The choice of encoding is not the only improvement we can do. Another way to improve the BA of the QNM-c is to supply $m > 1$ identical copies of the test state. Taking the tensor product $\rho^{\otimes m}$ is the quantum analogue of a polynomial feature expansion: it embeds the data in a larger space where classes separate more easily. Even a modest setting such as $m = 2$ or 3 can improve the BA by up to 10 percentage points, but that is left for future work.

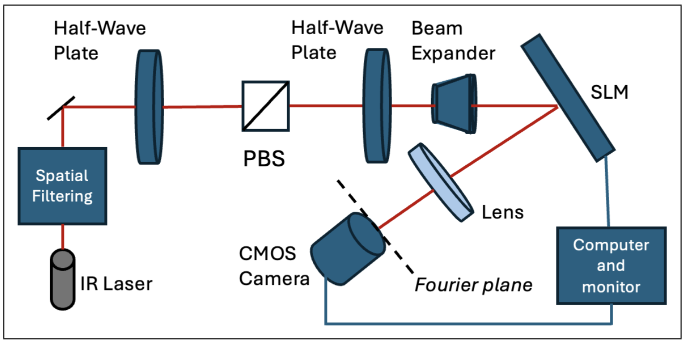


Figure 5: Schematic of the experimental setup for the JTC-based optical correlator. The laser beam is expanded and collimated, modulated by the SLM, and Fourier-transformed onto the CMOS camera.

V. Experimental results

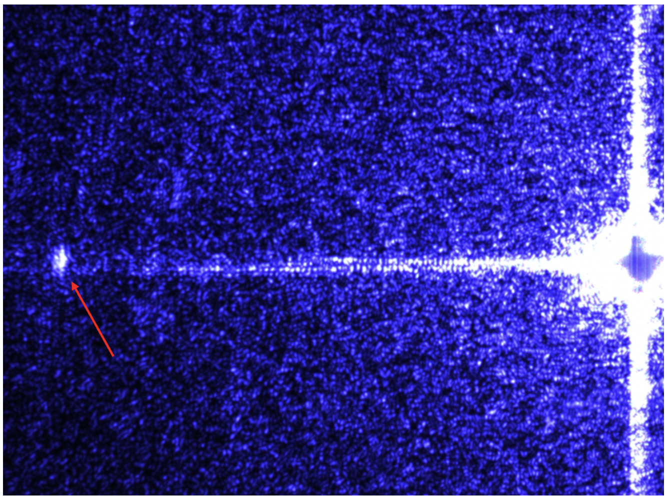
The optical setup used to implement the binary JTC (Fig. 5) consisted of a reflective, phase-only SLM (Thorlabs Exulus-HD3), illuminated by a IR laser at a wavelength of $\lambda = 845$ nm. The laser beam was collimated and expanded to achieve a uniform illumination across the SLM. A Fourier-transform lens with a focal length of $f = 150$ mm was used to perform the optical Fourier transform. The CMOS camera used was a Thorlabs model DCC1645C, placed at the fourier plane with a resolution of 1280×1024 pixels, capturing the joint power spectrum and the correlation plane. We calibrated the SLM to our specific wavelength to ensure that the grayscale values in the images accurately corresponded to the intended phase shifts. Errors in the phase range (0 to π) can significantly degrade the correlation results.

Initialy, the camera was saturating due to the dynamic range of the spectrum, so we applied a checkerboard pattern to the SLM to replicate the fourier transforms at the corners ([12]) and therefore reduce the dynamic range.

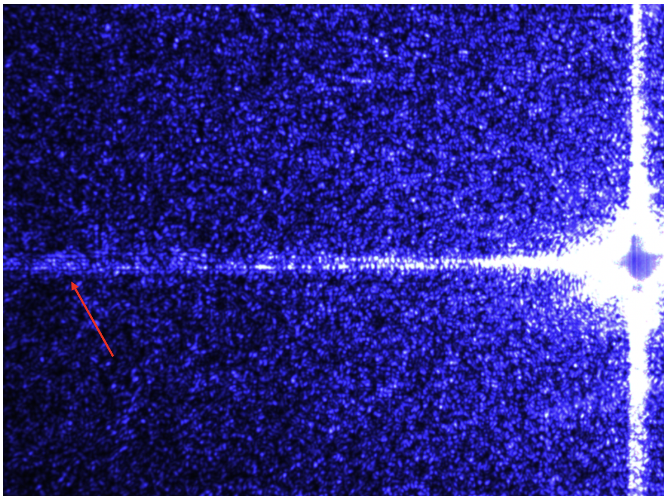
Although an off-axis cross-correlation peak is visible in Fig. 6a, its contrast is reduced by a strong constant term around the center of the Fourier plane. For Figurre 6b we see no cross-correlation peak, which is expected because the two digits are not similar. With a perfectly aligned ± 1 checkerboard the central term should vanish, because the same number of $+1$ and -1 pixels cancel out the zero-order light [12]. Its presence here suggests either a small mis-registration between the checkerboard and the SLM pixel grid, or an extra background signal that reaches the camera.

According to equation 11, the cross-correlation peaks should be located at $(\pm 2x_0, 0)$, where x_0 is the distance from the axis of the SLM to the image center. By varying the scale (i.e., the size of the images displayed on the SLM), we change the distance x_0 and therefore the position of the cross-correlation peaks. That is what we observe in Figures 7a and 7b, where the cross-correlation peaks get closer to the center of the correlation plane as we reduce the scale of the images.

To evaluate the performance of the JTC we recorded the correlation plane for twenty identical-digit pairs and, after a Gaussian pre-filter ($\sigma = 3$), extracted two key

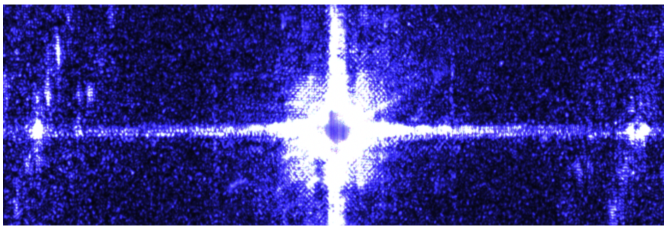


(a) Digits 1 vs. 1. After reducing the range to only capture the left part of the correlation plane, a off-axis peak (indicated by a red arrow) reveals the high similarity of the two inputs.

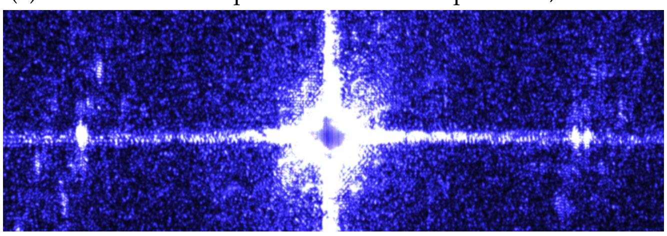


(b) Digits 1 vs. 2. Under the same scaling no noticeable off-axis peak is visible, indicating low similarity.

Figure 6: Experimental correlation planes recorded with the JTC setup. A match (a) shows a cross-correlation peak, whereas a mismatch (b) does not.



(a) Cross-correlation peaks at the off-axis positions, scale 0.6.



(b) Cross-correlation peaks at the off-axis positions, scale 0.5. Notice that the peaks move toward the centre.

Figure 7: Experimental correlation planes recorded with the JTC setup for the digit pair 1 vs. 1. Reducing the image scale shifts the cross-correlation peaks toward the centre of the plane.

metrics for each peak: the full-width at half-maximum (FWHM) along both axes and the signal-to-noise ratio (SNR) defined in [5]:

$$\mathrm{SNR} = \frac{R_{\max}}{\sigma_n},$$

where $R_{\max}$ is the correlation-peak amplitude and σ_n is the rms noise measured outside the 50 % response region.

Table 4: FWHM and SNR for identical-digit correlation peaks.

FWHM (px)	SNR (linear)	SNR (dB)
$(29, 27) \pm (5, 5)$	8.16 ± 0.06	18.23 ± 0.05

Table 4 shows that the mean peak size almost matches the 28×28 input digits, confirming that the JTC preserves all spatial detail required for a parallel implementation. The slight x–y asymmetry (2 px) and the 18 % spread are consistent with minor alignment drift rather than optical diffraction. An average SNR of 8.16 ± 0.06 places the peak about eight times above the background, well above typical detection thresholds, yet still limited by the 8-bit dynamic range of the camera-therefore several peaks already clip at full scale. Lowering the laser power or reducing the exposure time of the camera would increase headroom and boost the SNR.

The experimental figures coincide with those predicted by our numerical simulations, validating the optical model. At present, the principal bottleneck is not optical but electronic: the data link between the computer, SLM, and camera limits the achievable rate of the JTC to 2 correlations per second. Upgrading this interface, by using C++ instead of Python, or moving, for example, to on-board FPGA processing would allow the demonstrated optical resolution and SNR to translate directly into higher number of correlations per second.

VI. Conclusion

We have shown that a single-lens Joint Transform Correlator can act as a compact, power-efficient setup for both classical and quantum-inspired nearest-mean classifiers. Full-wave LightPipes simulations predict that, using the informative encoding, an optical JTC achieves a balanced accuracy of 82.39 % on MNIST when its correlation peak is interpreted as fidelity and mapped to trace distance. This already surpasses the 72.42 % reached by a classical CNM-C implemented on the same JTC simulation. Across all three encodings, quantum-inspired metrics consistently yield higher accuracy—improving performance by up to 10 percentage points over the classical baseline—and do so without relying on electronic post-processing and minimal pre-processing.

Our prototype confirms that these performance gains are physically realizable, achieving optical classification at 2 correlations per second using relatively simple components, with results consistent with simulation predictions. This rate is primarily constrained by the interface between the PC, SLM, and camera, with additional limitations from the camera's frame rate and the SLM's refresh rate. Throughput could be significantly improved by not only upgrading to faster components—such as a ferroelectric liquid crystal SLM (FLC-SLM) [12] and a high-speed camera—but also by optimizing the interface between them.

We also explored multiple encodings from real vectors to quantum states, and observed that higher-purity embeddings tend to yield trace distance estimates closer to their software counterparts, even though this does not always translate to better classification accuracy. This suggests that state purity plays a key role in the optical calculation of fidelity, but is not the sole determinant of classification performance. Together, these results validate the JTC as a possible fast and energy-efficient platform for analog quantum-inspired machine learning.

VII. Future Work

There are several promising directions to extend this work. **Parallel JTC implementation** represents a natural next step. While our current setup performs correlations sequentially, the JTC's inherent spatial parallelism [6] allows for multiple reference patterns to be displayed simultaneously on the SLM. With all classes encoded side by side, a single optical shot could evaluate all similarities at once, reducing inference time.

Encoding strategies play a central role in the performance of quantum-inspired classifiers. Our results suggest that centroid purity significantly influences the effectiveness of optical trace distance approximations. Exploring other encodings—especially ones that preserve purity while retaining directional and norm information—could yield further improvements.

Using **multiple copies** of the input state is another promising direction. By taking the tensor product of m identical copies, we can embed the data in a higher-dimensional space where classes may separate more easily. This approach has been shown to improve classification performance in quantum-inspired classifiers [2], and could be adapted to work with the JTC.

We also see a big potential in extending this framework to **other quantum-inspired classifiers**, such as the pretty good measurement (PGM) classifier [2], and this work serves as a simpler starting point for implementation of more complex quantum-inspired classifiers.

Another important direction is the use of **Uhlmann's theorem** to estimate trace distances between mixed states. If we can find a practical way to compute the purifications of mixed states, we could compute the true trace distance between an input and a quantum centroid directly, thereby eliminating the fidelity approximation, and not being limited to high purity encodings.

Alternative optical setups are also very relevant. Beyond the JTC, there are other optical systems such as VanderLugt correlators ([11], [3]) or optic-disk based systems [8] that may offer different trade-offs between complexity, robustness, and sensitivity. Investigating these could lead to architectures better suited for specific applications.

Acknowledgments

I would like to thank Emmanuel Cruzeiro for all the guidance; Preeti, Carolina, and José for hands-on help in the experiments; and the QuLab group in general for steady support and helpful discussions.

References

[1] A. Alfalou and C. Brosseau. Understanding correlation techniques for face recognition: From basics to applications. *ISEN Brest, L@bISEN*, 2024.

[2] E.Z. Cruzeiro, C. De Mol, S. Massar, et al. Quantum-inspired classification based on quantum state discrimination. *Quantum Machine Intelligence*, 6:79, 2024.

[3] Andrei Drăgulinescu. Optical correlators for cryptosystems and image recognition: A review. *Sensors*, 23(2):907, 2023.

[4] Wesley E. Foor and Mark A. Neifeld. Adaptive, optical, radial basis function neural network for handwritten digit recognition. *Applied Optics*, 34(32):7545–7555, 1995.

[5] Bahram Javidi and Joseph L. Horner. Single spatial light modulator joint transform correlator. *Applied Optics*, 28(5):1027–1032, 1989.

[6] Mikhail Kozoil, Mikhail Frolow, and Marina Frolow. LightPipes for Python 2.1.5, 2021. Python library for simulating propagation of coherent light beams.

[7] Peter L. McMahon. The physics of optical computing. *Nature Photonics*, 17(4):226–238, 2023. School of Applied and Engineering Physics, Cornell University.

[8] M. A. Neifeld and D. Psaltis. Optical implementations of radial basis classifiers. *Applied Optics*, 32:1370–1379, 1993.

[9] D. Psaltis, E. G. Paek, and S. S. Venkatesh. Optical image correlation with a binary spatial light modulator. *Optical Engineering*, 23:698–704, 1984.

[10] G. Sergioli, G. Russo, E. Santucci, A. Stefano, S.E. Torrisi, S. Palmucci, C. Vancheri, and R. Giuntini. Quantum-inspired minimum distance classification in biomedical context. *quant-ph*, 2018.

[11] A. Vander Lugt. Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory*, 10(2):139–145, Apr 1964.

[12] T. D. Wilkinson and W. A. Crossland. The binary phase-only 1/f joint transform correlator. In *Proceedings of SPIE — Optoelectronic Processing (Optics & Photonics)*, volume 3386, pages 65–69, 1998.