IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON NEW TRENDS IN BRAIN SIGNAL PROCESSING AND ANALYSIS

Received January 23, 2019, accepted February 1, 2019, date of publication March 20, 2019, date of current version April 12, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2904624*

# Towards a Quantum-Inspired Binary Classifier

**PRAYAG TIWARI AND MASSIMO MELUCCI**
Department of Information Engineering, University of Padova, 35131 Padova, Italy
Corresponding author: Prayag Tiwari (prayag.tiwari@dei.unipd.it)

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant 721321.

**ABSTRACT** Machine Learning classification models learn the relation between input as features and output as a class in order to predict the class for the new given input. Several research works have demonstrated the effectiveness of machine learning algorithms but the state-of-the-art algorithms are based on the classical theories of probability and logic. Quantum Mechanics (QM) has already shown its effectiveness in many fields and researchers have proposed several interesting results which cannot be obtained through classical theory. In recent years, researchers have been trying to investigate whether the QM can help to improve the classical machine learning algorithms. It is believed that the theory of QM may also inspire an effective algorithm if it is implemented properly. From this inspiration, we propose the quantum-inspired binary classifier, which is based on quantum detection theory. We used text corpora and image corpora to explore the effect of our proposed model. Our proposed model outperforms the state-of-the-art models in terms of precision, recall, and F-measure for several topics (categories) in the 20 newsgroup text corpora. Our proposed model outperformed all the baselines in terms of recall when the MNIST handwritten image dataset was used; F-measure is also higher for most of the categories and precision is also higher for some categories. Our proposed model suggests that binary classification effectiveness can be achieved by using quantum detection theory. In particular, we found that our Quantum-Inspired Binary Classifier can increase the precision, recall, and F-measure of classification where the state-of-the-art methods cannot.

**INDEX TERMS** Binary classification, quantum mechanics, signal detection.

## I. INTRODUCTION

In the 16th century, Johannes Kepler used the data collected for analysis purposes by Brahe and Copernicus in order to explore unrevealed patterns: the rotation of planets around the Sun at one focus of an ellipse [1]. The unrevealed patterns from astronomical data gave rise to some mathematical models such as the Gradient Descent algorithm to find out optimal points [2], Laplace's Equation for least square fitting [3], Gauss-Newton's algorithm for solving linear equations [4] and Lagrange's method for the polynomial interpolations [5]. Late 19th and 20th centuries gave rise to a wider range of mathematical methods for mining the unrevealed patterns from the data and their root can be found to the beginning of artificial intelligence and artificial neural network research in 1950s [6], [7]. In the 21st century, Machine Learning (ML) is a subpart of Artificial Intelligence where automated algorithms learn from the data [8]. Generally, there are two kinds of ML tasks, i.e. the supervised tasks and the unsupervised tasks. Classification is a key supervised task in data mining, recommender systems, and Information Retrieval (IR); we will focus on classification.

Despite its effectiveness in several domains, classification algorithms are still inadequate in some domains due to the nature of the data, which may be very heterogeneous, and the number of features, which may be inadequate. It can be seen in Figure 5 and 6 that baseline performances are very low in terms of recall as well as F-measure and precision to some extent in order to support the inadequacy of existing classification models. Since the data is growing exponentially, the need for algorithms that are more advanced than the state of the art is also growing. As an alternative to finding other methods within the classical frameworks, the main objective of this paper is to replace the classical probability theory underlying the current state-of-the-art learning algorithms with the quantum probability theory underlying QM and to figure out the novel models stemming from QM that cannot be observed through the lens of classical frameworks. In this

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Hugo Albuquerque.

paper, we designed, implemented and experimented with a Quantum-Inspired Binary Classifier (QIBC), which is a step moving from the classical model to quantum inspired models of classification and then of ML. Firstly, QIBC outperforms all the baselines in terms of recall for almost all feature ranges that can be seen in Figure 5 as well as F-measure and precision for a certain range of features which can be seen in Figures 6, 7, 8, 9 and 10. Since the dimension of the feature space is also increasing monotonically, computation time also increases with the feature set, which is a big issue. Therefore, another focus of this paper is the computation time: the QIBC takes much less time (similar to Naïve Bayes (NB) and Decision Tree (DT)) to compute the function as compared to Support Vector Machine (SVM) and $k$ Nearest Neighbours (KNN) which take a very long time, especially for an image dataset as can be seen in Figure 11. Our model suggests that high levels of effectiveness and less computation time in classification can be achieved by using the theory of QM. The main motivation of using QM in this paper is to make use of the superposition state that provides higher degree of freedom in decision making. The rest of the paper is organized as follows: Section II presents the literature survey. A technical background is presented in Section III ; after that Sections IV and V present the proposed methodology and the experiments, respectively. Finally, Section VI concludes the paper by suggesting possible future work.

## II. LITERATURE SURVEY

ML is a well-established discipline and a wide catalog of books is available including, for example [9], [10]. ML classification is the way to categorize documents; therefore, it is little surprising that classification in IR can be found dating back to the seventies. Based on the given set of topic, there are a number of publications addressing classification with IR [11]–[16]. Apart from IR, classification is a topic of computer vision where classification is based on contextual information in the images. ‘‘Contextual’’ information here means there is a focus on the relationship with the neighborhood pixels [17].

A novel non-parametric estimation model has been proposed, inspired by QM [43]. Each data sample is associated with its function by kernel density estimation. The probability density function is computed by taking the sum of all the kernels in the classical kernel estimation theory. In the quantum estimation theory, every data sample is associated with a quantum particle which has a radial activation field throughout it. In QM, the Schroedinger differential equation is utilized to describe the location of the particle using their given detected energy level. The location of each data sample is determined and their associated density function is modeled utilizing the correlation with the quantum potential function. The kernel scale has been computed from the distribution of KNN statistics. The local Hessian for finding the modes in quantum hyperspace is used in order to implement the proposed model for the purpose of classification. Every mode is incorporated with a nonparametric class that is explained by using a region growing model. The proposed model has been implemented on the artificial dataset as well as for the topography segmentation from the radar image of terrain.

Since the obtained classification dataset very often consists of useless and redundant features, feature selections become important steps in pre-processing the dataset in order to resolve the classification problem [44]. This kind of issue is generally resolved by implementing an evolutionary algorithm to minimize the dimensions of the features. The primary aim is to eliminate the useless features and distinguish the relevant features properly in feature space, which can improve the classification accuracy. A novel quantum-inspired binary gravitational search algorithm with the K-nearest neighbor (QBGSAKNN) model with leave-one-out cross-validation (LOOCV) has been proposed. The main motivation of this model is to enhance the classification accuracy with a suitable feature set in case of binary problems. This model has been implemented in several UCI machine learning benchmarks in order to check the performance and obtain high classification accuracy.

The issue of binary classification has been addressed by utilizing the quantum-type version of the Nearest Mean Classifier (NMC) [45]. This proposed model is efficient enough to be consistently derived to an absolute number of features and demonstrate the optimal performance measures while respecting the classical version of NMC for various collections. Moreover, quantum inspired NMC is not uniform under re-scaling. The re-scaling factor as a free parameter comes into play with the re-scaling that might be effective in achieving additional enhancement in the classification performance measures.

A novel quantum-inspired evolutionary algorithm with binary-real representation (QIEABRR) is proposed for the advancement of the neural network [46]. This QIEABRR model is the extended version of Quantum-Inspired Evolutionary Algorithm for Numerical Optimization (QIEANO) [47]. This QIEABRR model is capable of configuring the feed forward neural network with regards to choosing the appropriate input variables, quantity of neurons in the hidden layer and overall extant weights. The QIEABRR model has been implemented in financial credit evaluations.

The challenges with the state-of-the-art-models such as the issues with relevance assessment and aboutness were reviewed and discussed in [21], thus implicating the requirement to utilize more than a single ‘‘gold standard’’ method when estimating retrieval and indexing, and suggesting a new evaluation model. The model is informed by a methodical review of the literature on several different evaluation methods: estimating the quality of indexing by using the gold standard or by using an evaluator, and estimating the quality of indexing directly through the retrieval performance.

In contrast to the classical computers that are built on the physical application of the two different states $'0'$ and $'1'$, quantum computers exploit the qubit's superposition of two different quantum states $|0\rangle$ and $|1\rangle$ [22], [23]. In recent years, researchers have revealed the strength of quantum computers for ML [24], [25]. Information preserved in a quantum system is somehow restricted by the laws of QM and it is very challenging to come up with quantum algorithms that outperform their classical algorithms [26]. It is necessary to note that the use of QM suggested in this paper is not about the use of quantum computers to perform ML tasks.

## III. TECHNICAL BACKGROUND

In this section, all the technical concepts are described in order to understand the proposed QIBC.

### A. CLASSIFICATION

ML methods can be supervised or unsupervised. In the case of supervised tasks (e.g. classification), the input is arranged as high dimensional feature vectors and are associated with labels which are given to an algorithm in order to learn how to relate data to labels and then predict the labels for the new data. In the case of unsupervised learning, there are no labels and the algorithm has to cluster the data into several groups. The main aim of unsupervised algorithms is to compute distances between the feature vectors in time proportional to the dimension of the given vectors in the classical models.

The goal of classification is to learn a mapping from the given input $X$ to the output or target $Y$, where $Y \in \{0, 1, \dots, C-1\}$ and $C$ is the number of classes. If $C = 2$ then this is known as binary classification where we often assume $Y \in \{0, 1\}$. If $C > 2$ then this is known as multi-class classification. If the labels are not mutually exclusive then it is known as multi-label classification.

### B. CHI-SQUARE

Feature selection is an essential issue in ML. There are several feature selections available but Chi-square is used in order to keep the dependency between features and classes. Chi-square, which is also denoted as $\chi^2$, is a statistical method that is applied to examine the independence of two possible events, where event $A$ and event $B$ are interpreted to be independent if $P(A|B) = P(A)$ and $P(B|A) = P(B)$ or, uniformly $P(AB) = P(A)P(B)$. $\chi^2$ can be calculated as,

$$\chi^2 = \sum_{i=1}^{n} (O_C - E_C)^2 \,/\, E_C \tag{1}$$

where $O_C$ is the number of observations in the class $C$ and $E_C$ is the number of expected observations in the class $C$. The supervised feature selection method, $\chi^2$ which helps to measure the deviation between observed count and expected count. If there is dependency between two events, the occurrence of a feature can be used to predict the occurrence of the class. The main aim is to opt for the top features whose occurrence is more dependent on the occurrence of the class.

If there is independence between two events, the expected count and observed counts will be closer, which simply means that there will be small $\chi^2$ scores. The hypotheses of independence is incorrect if the value of $\chi^2$ is high. The most essential features which do not fit the expected values well are opted for in classification; that is, Chi-square is estimated between every target and feature to opt for the features with the highest $\chi^2$ scores.

### C. NAIVE BAYES

Bayes' theorem presents a way to compute the posterior probability, $P(B|A)$, from the $P(B)$, $P(A)$ and $P(A|B)$ as follows:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)},$$

where $A, B$ are events of a probability space, the posterior probability of the target class for the given feature vector is represented by $P(B|A)$, the prior probability of the target class is represented by $P(B)$, the likelihood that the feature vector can be observed in the class is represented by $P(A|B)$ and the prior probability of the feature vector is represented by $P(A)$ [27].

The classifier called Naive Bayes assumes that there is independence among features for each class in order to reduce the computational burden; this assumption is often known as the conditional independence of class and can be stated as follows:

$$P(A|B) = \Pi_{i=1}^{d} P(A_i|B),$$

where $d$ is the dimension of the feature vector space and $A_i$ is the $i$-th feature of vector $A \in \mathbb{R}^d$.

### D. DECISION TREE

A decision tree (DT) is one of the most popular methods in ML models. A DT is utilized for both regression as well as classification. A DT is a tree where each node denotes an attribute or feature, each link denotes the rule or decision and each leaf denotes the output which may be continuous or have a categorical value. The main idea is to make a tree for the whole dataset and process a single output at each leaf. In order to create the tree, a root node (a feature which better classifies the training data) is necessary. Information gain is used to determine the best feature. To explain information gain accurately, entropy is used, which distinguishes the impurity of a random collection of examples. [28]

Entropy $H(S)$ is the rate of the amount of uncertainty in the given set $S$ (data):

$$H(S) = \sum_{c \in C} -P(c) \log_2 P(c), \tag{2}$$

where $S$ is the dataset for which entropy calculates, $C \in \{0, 1\}$ is the set of classes in $S$ and $P(c)$ is the ratio of the number of features in $c$ to the number of features in $S$.

Information gain $IG(S, F)$ explains to what extent uncertainty in $S$ was lessened after splitting $S$ for feature $F$:

$$IG(S, F) = H(S) - H(S, F) \tag{3}$$

Alternatively, It can be written as follows:

$$IG(S, F) = H(S) - \sum_{t \in T} P(t)H(t). \tag{4}$$

where $T$ is the subset from splitting the set, $P(t)$ is the ratio of the number of features in $t$ to the number of features in $S$ set and $H(t)$ is the entropy of $t$ subset.

### E. SUPPORT VECTOR MACHINE

SVMs takes the input samples as a point in the geometrical space. The main aim is to identify the best hyperplane that distinguishes points between negative and positive categories in the case of binary classification. SVM utilizes a set of feature functions in terms of flexibility instead of using event space in NB.

In case of linearly separable data, $x^+$ and $x^-$ are the closest positive and negative training points to the hyperplane. The margin is written as follows:

$$\text{margin} = |w.x^-| + |w.x^+| / ||w||. \tag{5}$$

A vector $w$ must be identified in order to solve the optimization problem of identifying the better hyperplane. The optimization problem can be expressed as identifying $||w||^2/2$ subject to $w.x_i \leq -1, \forall i$ such that class$(i)$ = negative and $w.x_i \geq 1, \forall i$, class$(i)$ = positive. In reality, most of the datasets are non-linearly separable and few datasets are linearly separable. [29]

### F. K-NEAREST NEIGHBOURS

The K-Nearest Neighbours method stores all the possible points and classifies new given points based on a similarity or distance function. KNN has been utilized in pattern recognition since the 1970s as a non-parametric method. A point is classified by the majority of votes of its neighbor point, with the point has been allocated to the class most common among their $k$ nearest neighbors. Some distance functions can be seen below:

- Euclidean distance function, $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$
- Manhattan distance function, $\sum_{i=1}^{k}|x_i - y_i|$
- Minkowski distance function, $(\sum_{i=1}^{k}(|x_i - y_i|)^q)^{\frac{1}{q}}$
- Hamming distance function, $D_H = \sum_{i=1}^{k}|x_i - y_i|$

The first three distance functions are valid for continuous variables. The hamming distance is generally used for categorical variables; so, if $x = y$ then $D_H = 0$ and if $x \neq y$ then $D_H = 1$.

The optimal value of $k$ can be chosen by examining the data. Generally, a higher value of $k$ is better as it helps to reduce the noise but this cannot be guaranteed. Another way is to use cross validation to find a better $k$ value by utilizing independent data to validate the $k$.

### G. CLASSICAL SIGNAL DETECTION THEORY

Signal Detection Theory (SDT) gives a general framework to illustrate and inspect decisions made in uncertain and ambiguous situations. SDT is based on three main concepts, i.e. signal, noise, and decision. Channel, receiver and source are the components of any signal detector (Figure 1).

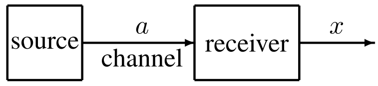


**FIGURE 1. Classical communication system.**

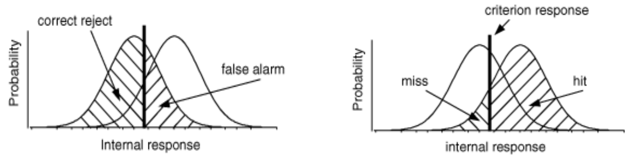


**FIGURE 2. Signal Detection Theory (SDT).**

We utilize SDT because classification is a task directly involving human users and SDT has been utilized widely in the field of psychophysics, that is, the field that examines the connection between stimulus and its psychological effect.

SDT needs some inference about how many types of decisions are considered under uncertainty. The three grounded questions in SDT are as follows: (a) Are we certain about the signal? (b) How can we know if the signal is correct? (c) How do we determine to react or not?

SDT function is used to identify whether the detected impulse is caused by the signal or noise [30]–[32]. Decision theory manages the possibilities among the hypotheses about the system at hand. Estimation Theory generally manages data $x_1, x_2, x_3, \ldots, x_n$ whose joint probability density function $P(x;\theta) = P(x_1, x_2, x_3, \ldots, x_n; \theta_1, \theta_2, \theta_3, \ldots, \theta_m)$ depends on some unknown parameter $\theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$ which is to be computed. For example, the given data can be samples $x_j = x(t_j)$ for the given input $x(t) = s(t;\theta) + n(t)$ to the receiver, consist of noise $n(t)$ with statistical characteristics and a signal $s(t;\theta)$ which depends on $\theta$, i.e. time of arrival, amplitude and carrier frequency. Estimation Theory sets up an estimation of seriousness or cost of errors in the estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m)$ of some parameters. The very common cost function, which is the weighted sum of the squared errors is as follows:

$$K(\hat{\theta}, \theta) = \sum_{k=1}^{m} W_k(\hat{\theta}_k - \theta_k)^2. \tag{6}$$

The main issue is to identify the $\hat{\theta}_k = \hat{\theta}_k(x_1, x_2, x_3, \ldots, x_n)$ as a function of that data in order to minimize the average cost.

In the case of binary decision, i.e. binary classification, there are two hypotheses, each hypothesis being represented by the presence or absence of the signal $s(t)$ in the input form $x(t)$ to the receiver during some observation interval $(0, T)$, where $n(t)$ is noise with statistical characteristics. The two hypotheses can be written as follows:

- Null hypothesis, $A_0 : x(t) = n(t)$
- Alternative hypothesis, $A_1 : x(t) = n(t) + s(t)$

**TABLE 1. Summary of the decision about the presence / absence of a signal. For each actual state and decision the corresponding outcome has costs.**

| | Presence of signal (1) | Absence of signal (0) |
|---|---|---|
| Decision as Yes (1) | Detection ($K_{11}$) | False Alarm ($K_{01}$) |
| Decision as No (0) | Miss ($K_{10}$) | Rejection ($K_{00}$) |

The best method of deciding between two hypotheses can be expressed in two ways. In terms of NB, an observer is aware of the following:

- the prior probabilities $\xi$ and $(1-\xi)$ of hypothesis $A_0$ and $A_1$
- the four costs $K_{ij}$ of choosing hypothesis $A_i$ when $A_j$ is correct $(i, j \in \{0, 1\})$.

The costs are required by the circumstances and the actions following the choices, in such a way that average cost would be minimal. Table 1 summarizes the decision costs.

Another way to describe the optimal binary decision is given by the theory of Neyman and Pearson [33], [34] which explains the two possible types of errors:

- opting for $A_1$ when $A_0$ is true, which is also known as false alarm or first type of error and its probability under the given method is represented by $Q_0$;
- opting for $A_0$ when $A_1$ is true, which is also known as miss or second type of error and its probability is represented by $Q_1$.

The complement $Q_d = 1 - Q_1$ is often known as the probability of detection. The strategy finding the best decision achieves the highest probability $Q_d$ for each first type of error probability $Q_0$.

### H. QUANTUM MECHANICS

Quantum Mechanics is one of the most surprising parts of modern physics and eminent theoretical achievements. It was originated to describe puzzling observations that were not possible to interpret by utilizing classical physics. The researchers obtained this by bringing a completely new set of principles into play. During the process of defining the new mechanics of the invisible world, a physicist also created a novel theory of a dynamic probabilistic system, which is more accepted than previous classical theory in particle physics [35], [36].

QM was first presented by Erwin Schroedinger and Werner Heisenberg and it reached maturity in the 1920s and 1930s [48]. In particular, Schroedinger formulated the fundamental equation of wave mechanics which is known as Schroedinger's equation: $H\psi = E\psi$, where $\psi$ is the eigenfunction which describes the state of the system, $H$ is a Hamiltonian operator and $E$ is the eigenvalue for the system energy. Heisenberg introduced the Uncertainty Principle stating that the momentum and position of any particle cannot be measured at the same time at arbitrary precision [37], thus establishing the principle that Nature is basically uncertain. While measurement of macroscopic world entities is deterministic according to the classical theory because of the fact that a system state can be absolutely measured several times, randomness is an intrinsic feature of the invisible world according to QM. This uncertainty does not happen because of the imprecision in measurement, it is rather due to the intrinsic randomness of the state of the system [38].

Quantum superposition is the core of QM, as said by Feynman. It can be simply represented by a mathematical equation:

$$\psi = \alpha_1|\phi_1\rangle + \alpha_2|\phi_2\rangle \quad |\alpha_1|^2 + |\alpha_2|^2 = 1. \tag{7}$$

The equation 7 can simply be understood as follows: a particle that is set at a certain state $\psi$ is simultaneously set at both states $\phi_1$ and $\phi_2$. But, when the particle is measured and a decision is taken as to whether the particle has been set at $\phi_1$ or $\phi_2$, either $\phi_1$ or $\phi_2$ will result and the particle will no longer be set at the superposed state $\phi$, which is called "collapse".

Collapse is a random event whose mechanics cannot be described in a deterministic way. An observer can only estimate the probability $|\alpha_i|^2$ that the particle collapses to $\phi_i$. A superposed state $\psi$ of one particle and the probability distribution thereof is radically different from a collection of individual particles where $\alpha_1 \times 100\%$ of the particles are at the state $\phi_1$ and $\alpha_2 \times 100\%$ of the particles are at state $\phi_2$. The latter distribution is called mixture and refers to a collection of individual particles whereas superposition refers to one individual particle.

Both superposition and mixture are described by density operators. A density operator is a mathematical object "universally" utilized to encode probability distributions and describe superpositions $\rho$'s or equivalently $\psi$'s such as

$$\rho = |\psi\rangle\langle\psi|$$

and mixtures of superpositions as can be seen as follows:

$$\sum p_i|\rho_i\rangle\langle\rho_i|,$$

where $p_i$ is the proportion of particles of the mixture at state $\psi_i$.

## IV. PROPOSED ARCHITECTURE AND METHODOLOGY

The overall proposed architecture of our model can be seen in Figure 3, showing how SDT, QM, Quantum SDT, Classical ML, DT are intersected in order to develop the Quantum Detection Model, which is used as a classifier. As in the general classification (supervised learning) task, input features with the given label are used to train the model and then the model is tested based on test features without any labels. Also in QIBC a label with input features is used to train our proposed model and tested on the test dataset.

The methodology of the proposed QIBC framework in the Section IV-B which is inspired by quantum SDT (Section IV-A).

### A. QUANTUM SDT

In quantum SDT, there is coder between the channel and source as shown in Figure 4. On the side of the source,

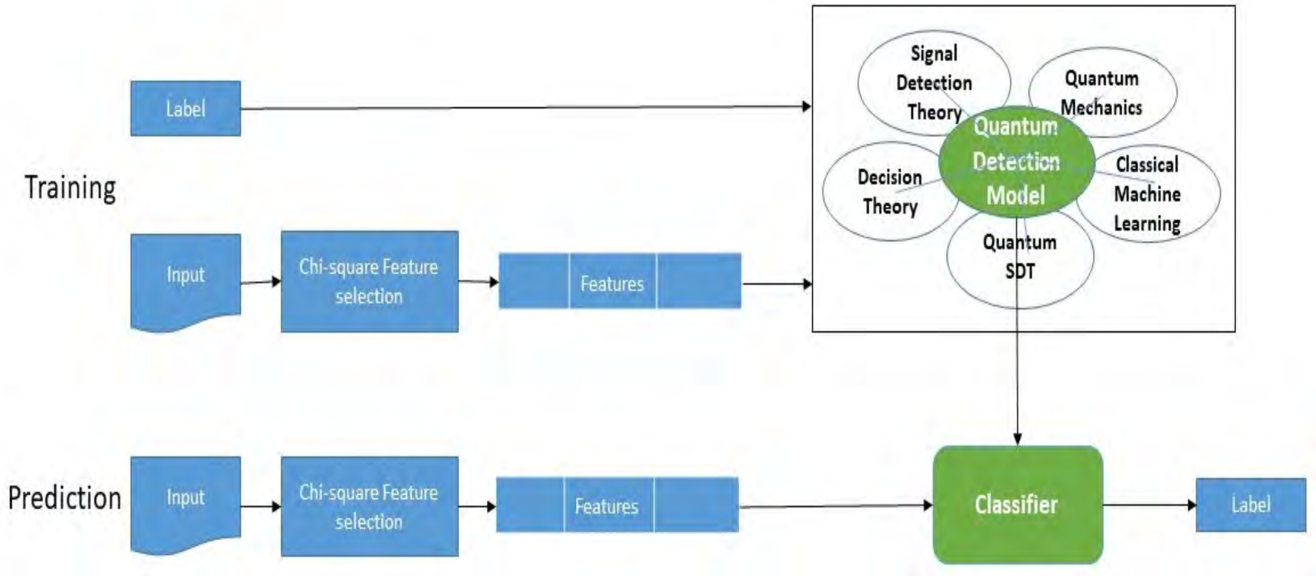


**FIGURE 3. Quantum-Inspired Binary Classifier (QIBC).**

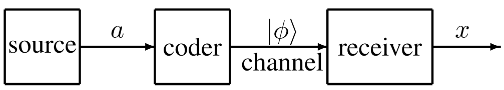


**FIGURE 4. Quantum communication system.**

the encoding part is carried out by the coder from the signal into some particle, i.e. a qubit whose pure state is $|\phi\rangle$. On the side of the receiver, a measurement is implemented, like the classical detection framework. The main difference between the classical and quantum model lies in what is encoded by an encoder and then what is decoded by the decoder. Mathematically speaking, this contrast means that the projector corresponding to the prime measurement can optimally be measured by utilizing classical theory in the case of the classical model. [38]–[41].

The two hypotheses subjected to decision are the presence or the absence of the signal. In the case of quantum framework, the decision about a system is to be made between these two density operators, i.e. either:

- $A_0$ with density operator $\rho_0$ and with the prior probability $\xi$, or
- $A_1$ with density operator $\rho_1$ and with the prior probability $(1 - \xi)$.

Consider some output $x_1, x_2, x_3, \ldots$ of some observables $X_1, X_2, X_3, \ldots$. The decision will be based upon a detection operator $\Delta$. The detection operator must be the one whose eigenvalues are 0 and 1, and this kind of operator is denoted as projection operator. The choice should be made between 0 and 1, and opt for $A_0$ when $\Delta = 0$ and $A_1$ when $\Delta = 1$. The detection operator $\Delta$ yields 1 under $A_0$ with the following probability:

$$Q_0 = P(\Delta = 1|A_0) = \mathrm{tr}(\rho_0 \Delta). \tag{8}$$

Similarly,

$$Q_1 = P(\Delta = 1|A_1) = \mathrm{tr}(\rho_1 \Delta). \tag{9}$$

The average cost can be written as

$$\bar{K} = \xi K_{00} + (1-\xi)K_{01} - (1-\xi)(K_{01} - K_{11})\mathrm{tr}(\rho_1 - \lambda\rho_0)\Delta, \tag{10}$$

where

$$\lambda = \frac{\xi(K_{10} - K_{00})}{(1-\xi)(K_{01} - K_{11})}. \tag{11}$$

As long as $K_{01} > K_{11}$, $\bar{K}$ will be minimum if $\mathrm{tr}\left((\rho_1 - \lambda\rho_0)\,\Delta\right)$ can be maximized.

The best detection operator is provided by the eigenstates $|e_l\rangle$ of the operator $\rho_1 - \lambda\rho_0$ corresponding to the positive eigenvalues, where the eigensystem is provided by

$$(\rho_1 - \lambda\rho_0)|e_l\rangle = e_l|e_l\rangle \quad l = 1, \ldots, (\text{rank of } \rho_1 - \lambda\rho_0). \tag{12}$$

So it is essential to maximize

$$\mathrm{tr}(\rho_1 - \lambda\rho_0)\Delta = \sum_l e_l \langle e_l|\Delta|e_l\rangle \tag{13}$$

and this can be obtained if

$$e_k\langle e_l|\Delta|e_l\rangle = 1, e_l \geq 0 \quad \text{and } e_l\langle e_l|\Delta|e_l\rangle = 0, e_l < 0.$$

Hence, the estimation of the optimal projection operator between $A_0$ and $A_1$ can be written as

$$\Delta = \sum_{l:e_l \geq 0} |e_l\rangle\langle e_l|. \tag{14}$$

So the probabilities of error can be written as:

$$Q_0 = \sum_{l:e_l \geq 0} \langle e_l|\rho_0|e_l\rangle \qquad \text{and } Q_1 = 1 - \sum_{l:e_l \geq 0} \langle e_l|\rho_1|e_l\rangle. \tag{15}$$

and the minimum average cost can be written as:

$$\bar{K}_{min} = \xi K_{00} + (1-\xi)K_{01} - (1-\xi)(K_{01} - K_{11}) \sum_{l:e_l > 0} e_l \tag{16}$$

Quantum SDT in IR was introduced in [42] to re-weight the query terms and re-rank the retrieved documents. Consider the vector $|y\rangle$, which is the input query of an IR system providing a ranked list of documents, each document being represented by a vector $|x\rangle$. The relevance assessments allow to estimate the density operators $\rho_0$ by using nonrelevant documents and $\rho_1$ by using relevant documents. The use of quantum SDT in IR consists of a relevance feedback algorithm which projects both the query vector $|y\rangle$ and the document vectors $|x\rangle$ by means of the optimal detection operator; therefore, re-ranking is computed by:

$$\langle x|\Delta|y\rangle. \tag{17}$$

### B. QUANTUM-INSPIRED BINARY CLASSIFIER

A novel QIBC inspired by quantum detection theory is described in this section. For each topic (category) we suppose that each training sample is about the category or not. For a given category and the set of training samples, we use the projector $\Delta$ for each category to identify whether the test sample is about the category or not. To determine whether the test sample is about the category, $\Delta$ is examined against a vectorial representation of the test sample.

Consider a set of distinct features calculated from the whole sample collection. Each sample can be represented as a vector of features; each element in the feature vector is a non-negative number such as frequency. Each sample in the training set has a binary label in $\{0, 1\}$. The main goal of QIBC is to obtain one binary label for each sample in the test set.

The QIBC estimates the density operators $\rho_0$ and $\rho_1$ by using the training sample; in particular, for each class, the negative training samples were utilized to estimate $\rho_0$ and the positive training samples were utilized to estimate $\rho_1$.

In order to achieve these density operators $\rho_0$ and $\rho_1$, we firstly calculated the total number of samples with non-zero values for each particular feature. In this way, one vector $|v\rangle$ was obtained for each class. Since we were considering the binary case, two vectors $|v_0\rangle$ and $|v_1\rangle$ were obtained; the former refers to the negative training sample and the latter refers to the positive training sample. These vectors can be considered as statistics of the features in a class. We normalized the acquired vectors to obtain $|\langle v|v\rangle|^2 = 1$. Then, we calculated the outer product in order to obtain the density operators $\rho_0$ and $\rho_1$ as follows:

$$\rho_0 = \frac{|v_0\rangle\langle v_0}{tr(|v_0\rangle\langle v_0)} \qquad \rho_1 = \frac{|v_1\rangle\langle v_1}{tr(|v_1\rangle\langle v_1)} \tag{18}$$

We set $\lambda = 1$, meaning that the prior probability is the same for both classes ($\xi = 0.5$); moreover, there is no cost for correct detection $K_{00} = K_{11} = 0$; finally, the costs of false alarm and miss are constant ($K_{01} = K_{10}$). Eventually, we determined the binary label for the given test sample $x$ represented by $|x\rangle$ by inspecting the value of $\langle x|\Delta|x\rangle$. If $\langle x|\Delta|x\rangle \geq 0.5$ then $x$ was assigned to the class, otherwise it was not assigned.

## V. EXPERIMENTS

### A. DATASET DESCRIPTION

To evaluate the effectiveness of our proposed model, the 20 Newsgroups[1] text corpora have been used. The 20 Newsgroups text corpora is a set of 18,846 newsgroup documents, categorized evenly across 20 different newsgroups and each newsgroup corresponds to some different topics

"(alt.atheism, comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, comp.sys.mac.hardware, comp.windows.x, misc.forsale, rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey, sci.crypt, sci.electronics, sci.med, sci.space, soc.religion.christian, talk.politics.guns, talk.politics.mideast, talk.politics.misc, talk.religion.misc)". There are 18,846 documents in which 11,314 documents are used for training purposes and 7,532 for testing[2].

To check the effect of our proposed model on an image dataset as well, the MNIST handwritten image dataset [3] was used. There are 60,000 training set examples, and 10,000 test set examples in the dataset. There were 10 categories from 0 to 9 but the last category was not taken into consideration because there were much fewer samples. This dataset is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.

### B. PROCEDURE

Feature selection is always an essential step. The evaluation measures vary by changing the number of selected features, so selecting the top 400 or the top 10 features greatly affects the evaluation measures. We used the chi-square feature selection method to select those top features. In order to see the changes we decided to do analysis on the range of features, like starting from the top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400. We used these ranges of numbers randomly as the basis of the obtained experiment.

We used NB, SVM, KNN and DT as baselines in order to compare with QIBC. These evaluation parameters were used i.e. precision, recall, and F-measure, to measure the effectiveness of QIBC and of the baselines.

[1] http://qwone.com/~jason/20Newsgroups/
[2] http://www.cad.zju.edu.cn/home/dengcai/Data/TextData.html
[3] http://yann.lecun.com/exdb/mnist/

There are multiple categories available in both datasets; therefore, we used one-vs-all strategy for each category, that is, the sample in the training set labeled as pertinent to the category was considered a positive example while the rest were negative examples. In some cases, there were very few samples or no samples for a given category so some categories were excluded. Five-fold cross-validation was performed while training the model.

### C. MEASURES

The measures that are often used in ML to check the performance of the proposed model are Precision, Recall and F-measure.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

where

- TP is the number of true positives, i.e. the samples that are correctly classified in a certain class;
- FP is the number of false positives, i.e. the samples that are not classified in the class;
- TN is the number of true negatives, i.e. the samples that are not correctly classified in the class;
- FN is the number of false negatives, i.e. the samples that are incorrectly classified in the class.

### D. RESULTS

#### 1) EXPERIMENTAL RESULTS ON MNIST HANDWRITTEN IMAGE DATASET

The three evaluation parameters, i.e., Precision, Recall and F-measure, were used to evaluate the performance of the models. We used the MNIST handwritten image dataset to check the performance of our model on an image dataset. The evaluation parameters change when altering the range of features. We selected a range of features, i.e. the top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400, respectively. Analysis was performed for each category. QIBC outperforms all the baselines for several categories in terms of precision, recall and F-measures which can be seen in Table 2, 3 and 4.

Our proposed QIBC model outperformed all baselines in terms of recall for almost the entire range of features, as can be seen from Figure 5. In terms of recall, QIBC works very well on an image dataset and can be used safely where high recall is required as it performs better overall baselines. Computation time is another issue as some baselines take a long time to compute, for example SVM and KNN, while QIBC takes much less time to compute which is another advantage.

In terms of F-measure, QIBC outperformed all the existing models in the starting range of features, i.e. the top 5, 10, 15, 20 and 30, for all categories except category 0 as can be seen in Figures 6(b), 6(c), 6(d), 6(e), 6(f), 6(g), 6(h) and 6(i). QIBC outperformed existing models in all ranges up to the top 50 features as can be seen in Figures 6(b) 6(c) 6(f) 6(g) for categories 1, 2, 5, 6 and 7. QIBC outperformed existing models in all ranges up to the top 70 features for category 6 as can be seen in Figure 6(g). Subsequently, QIBC performance started decreasing with an increasing number of features, i.e. 100, 150, 200 and 400. Therefore, QIBC can provide some good results where other baselines cannot perform, for instance when the top 5 features are selected.

In terms of precision, QIBC outperformed all the baselines for several categories, i.e. categories 0, 1, 5, 6 and 7 when the top 10 features were selected as can be seen in Figures 7(a), 7(b), 7(f), 7(g) and 7(h). QIBC had results similar to the baselines for categories 1, 2, 5 and 6 when the top 15 features were selected as can be seen in Figures 7(b), 7(c), 7(f) and 7(g). The QIBC performance was similar to the baselines in the cases of categories 5 and 6 when the top 20 features were selected as can be seen in Figure 7(f) and 7(g). QIBC still outperformed some baselines, for instance SVM, when the top 50 features were selected for category 1 as can be seen in 7(b). It also outperformed SVM when the top 70 features were selected for category 3 as can be seen in 7(d). QIBC outperformed KNN when the top 20 features were selected for category 4 as can be seen in 7(e) and it outperformed SVM when the top 40 features were selected for category 5 as can be seen in 7(f). QIBC outperformed SVM when the top 70 features were selected for category 6 as can be seen in 7(g) and it outperformed SVM when the top 50 features were selected for category 7 as can be seen in 7(h). Lastly, QIBC outperformed SVM when the top 30 and 40 features were selected for category 8 as can be seen in 7(d).

#### 2) EXPERIMENTAL RESULTS ON 20 NEWSGROUP TEXT CORPORA

QIBC outperforms all the baselines for several topics in terms of precision, recall and F-measures which can be seen in Table 5, 6 and 7.

In terms of Precision, QIBC outperformed most of the baselines, i.e. topics 2, 7, 19, 20, 1, 3, 4, 6, 8, 9, 10, 14, 16 and 18 for the top 5 features as can be seen in Figure 8(a). DT failed to even perform for 15 out of 20 topics and DT performed better for only 5 topics. It can be seen in Figure 8 and Figure 11 that the performance of QIBC in terms of precision and computation time was better than the other baselines if we take DT into less consideration because it failed to even perform for most of the topics. A similar situation happened when the top 10 features were selected as can be seen in Figure 8(c) where QIBC outperformed most of the baselines, i.e. topics 5, 4, 20, 9, 1, 3, 6, 8, 10, 15, 16 and 18, in terms of precision and computation time. QIBC outperformed all the baselines for most of the topics, i.e. 8, 1, 3, 4, 7, 9, 10, 14, 16, 18 and 20, when the top 15 features were selected as again DT failed to perform for most of the topics. QIBC performed better again for the topics, i.e. 3, 4, 7, 9, 10, 14, 15, 18 and 20 as can be seen in Figure 8(b). By increasing the number of selected features, the performance of QIBC decreased slightly with the increasing number of features but again QIBC takes much less time to perform and always

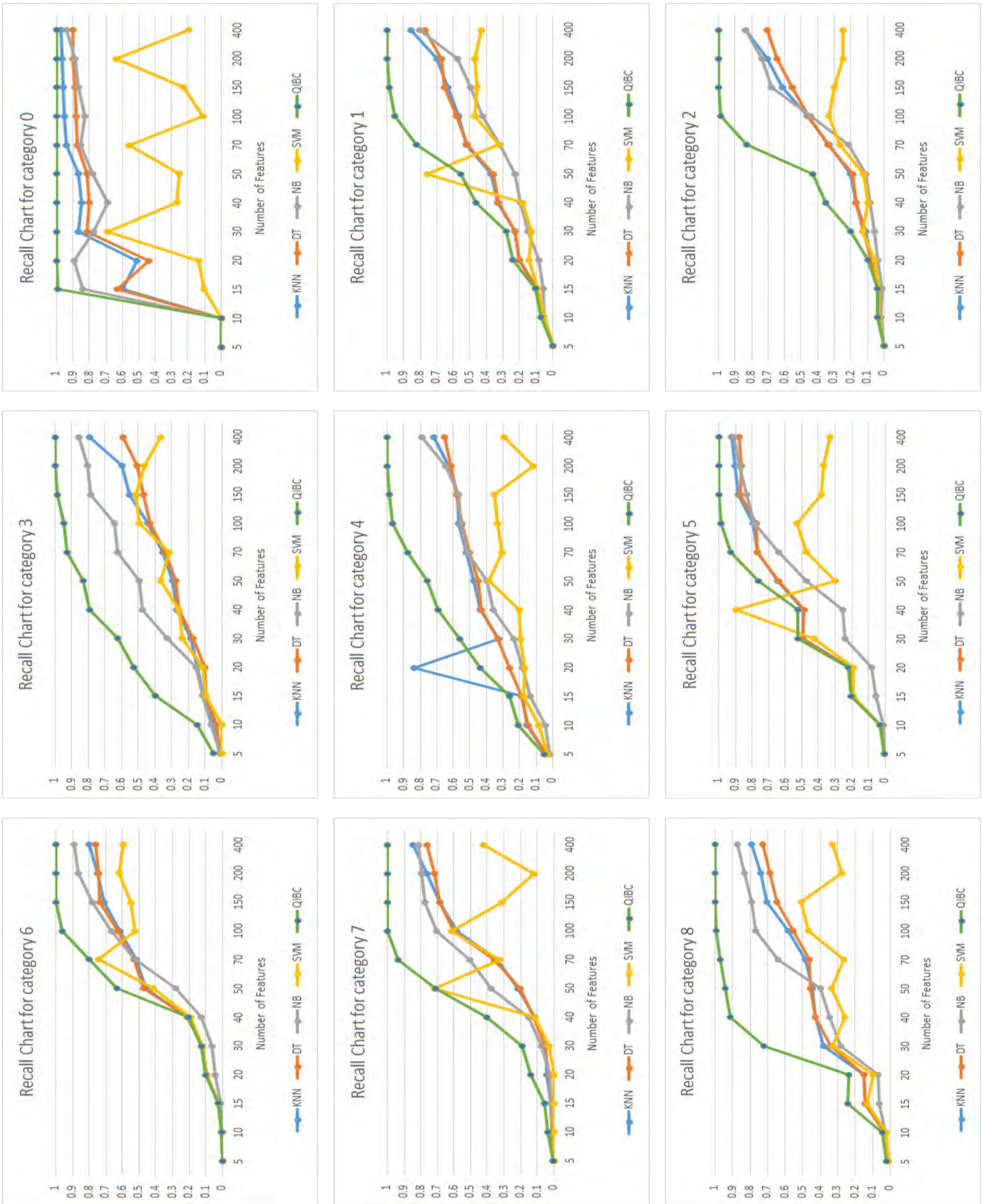


**FIGURE 5.** Recall chart for each category by changing number of features 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400 among KNN, DT, NB, SVM and QIBC on MNIST handwritten image dataset.

**TABLE 2.** Precision Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on 20 newsgroup dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | $QIBC_2 = 0.693$<br>$QIBC_9 = 1.00$ | $QIBC_{19} = 0.818$<br>$QIBC_{10} = 1.00$ | $QIBC_{20} = 0.967$<br>$QIBC_{14} = 1.00$ | $QIBC_1 = 0.972$<br>$QIBC_{16} = 1.00$ | $QIBC_1 = 1.00$<br>$QIBC_{18} = 1.00$ | $QIBC_4 = 1.00$ | $QIBC_6 = 1.00$ | $QIBC_8 = 1.00$ |
| 10 | $QIBC_5 = 0.923$<br>$QIBC_{10} = 1.00$ | $QIBC_4 = 0.923$<br>$QIBC_{15} = 1.00$ | $QIBC_{20} = 0.947$<br>$QIBC_{16} = 1.00$ | $QIBC_9 = 0.95$<br>$QIBC_{18} = 1.00$ | $QIBC_1 = 0.971$ | $QIBC_3 = 1.00$ | $QIBC_6 = 1.00$ | $QIBC_8 = 1.00$ |
| 15 | $QIBC_8 = 0.918$<br>$QIBC_{14} = 1.00$ | $QIBC_1 = 0.952$<br>$QIBC_{16} = 1.00$ | $QIBC_3 = 1.00$<br>$QIBC_{18} = 1.00$ | $QIBC_4 = 1.00$<br>$QIBC_{20} = 1.00$ | $QIBC_7 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{10} = 1.00$ |
| 20 | $QIBC_3 = 1.00$<br>$QIBC_{18} = 1.00$ | $QIBC_4 = 1.00$<br>$QIBC_{20} = 1.00$ | $QIBC_7 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{14} = 1.00$ | $QIBC_{15} = 1.00$ |
| 30 | $QIBC_{12} = 0.933$ | $QIBC_3 = 1.00$ | $QIBC_6 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{13} = 1.00$ | $QIBC_{16} = 1.00$ | $QIBC_{18} = 1.00$ |
| 40 | $QIBC_{12} = 0.933$ | $QIBC_2 = 1.00$ | $QIBC_6 = 1.00$ | $QIBC_8 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{13} = 1.00$ | $QIBC_{16} = 1.00$ | |
| 50 | $QIBC_4 = 0.745$ | $QIBC_2 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{13} = 1.00$ | $QIBC_{16} = 1.00$ | | |
| 70 | $QIBC_1 = 0.961$ | $QIBC_2 = 1.00$ | $QIBC_3 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{13} = 1.00$ | $QIBC_{16} = 1.00$ | | |
| 100 | $QIBC_1 = 0.961$<br>$QIBC_{16} = 1.00$ | $QIBC_5 = 0.975$ | $QIBC_2 = 1.00$ | $QIBC_3 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{13} = 1.00$ | $QIBC_{14} = 1.00$ |
| 150 | $QIBC_{17} = 0.916$<br>$QIBC_{16} = 1.00$ | $QIBC_2 = 1.00$<br>$QIBC_{18} = 1.00$ | $QIBC_3 = 1.00$<br>$QIBC_{20} = 1.00$ | $QIBC_7 = 1.00$ | $QIBC_9 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{11} = 1.00$ | $QIBC_{14} = 1.00$ |
| 200 | $QIBC_2 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{11} = 1.00$ | $QIBC_{14} = 1.00$ | $QIBC_{20} = 1.00$ | | | |
| 400 | $QIBC_2 = 1.00$ | $QIBC_{10} = 1.00$ | $QIBC_{14} = 1.00$ | $QIBC_{17} = 1.00$ | | | | |

**TABLE 3.** Recall Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on 20 newsgroup dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | $QIBC_2 = 0.023$ | $QIBC_{20} = 0.119$ | $QIBC_{18} = 0.154$ | | | | | |
| 10 | $QIBC_4 = 0.061$ | $QIBC_7 = 0.065$ | $QIBC_{17} = 0.145$ | $QIBC_{20} = 0.215$ | | | | |
| 15 | $QIBC_2 = 0.069$ | | | | | | | |
| 20 | $QIBC_8 = 0.111$ | $QIBC_{19} = 0.177$ | $QIBC_{16} = 0.560$ | | | | | |
| 30 | $QIBC_4 = 0.071$<br>$QIBC_{15} = 0.359$ | $QIBC_7 = 0.078$<br>$QIBC_{12} = 0.463$ | $QIBC_2 = 0.120$<br>$QIBC_{15} = 0.624$ | $QIBC_8 = 0.136$ | $QIBC_{19} = 0.206$ | $QIBC_{14} = 0.231$ | $QIBC_{20} = 0.235$ | |
| 40 | $QIBC_4 = 0.089$<br>$QIBC_{18} = 0.486$ | $QIBC_3 = 0.097$<br>$QIBC_{12} = 0.531$ | $QIBC_7 = 0.102$<br>$QIBC_{10} = 0.586$ | $QIBC_{19} = 0.235$<br>$QIBC_{11} = 0.651$ | $QIBC_{14} = 0.254$ | $QIBC_{20} = 0.254$ | $QIBC_5 = 0.326$ | $QIBC_{15} = 0.397$ |
| 50 | $QIBC_4 = 0.096$<br>$QIBC_{15} = 0.426$ | $QIBC_7 = 0.104$<br>$QIBC_{18} = 0.494$ | $QIBC_3 = 0.104$<br>$QIBC_{12} = 0.546$ | $QIBC_{19} = 0.258$<br>$QIBC_{11} = 0.654$ | $QIBC_{20} = 0.270$ | $QIBC_6 = 0.276$ | $QIBC_{14} = 0.292$ | $QIBC_5 = 0.334$ |
| 70 | $QIBC_7 = 0.115$<br>$QIBC_{12} = 0.579$ | $QIBC_4 = 0.127$<br>$QIBC_{10} = 0.649$ | $QIBC_8 = 0.174$<br>$QIBC_{11} = 0.681$ | $QIBC_{19} = 0.270$ | $QIBC_{20} = 0.302$ | $QIBC_6 = 0.315$ | $QIBC_{14} = 0.389$ | $QIBC_{18} = 0.513$ |
| 100 | $QIBC_4 = 0.140$<br>$QIBC_{12} = 0.602$ | $QIBC_7 = 0.143$<br>$QIBC_{11} = 0.714$ | $QIBC_8 = 0.278$ | $QIBC_{19} = 0.287$ | $QIBC_{20} = 0.334$ | $QIBC_6 = 0.351$ | $QIBC_{15} = 0.515$ | $QIBC_{18} = 0.542$ |
| 150 | $QIBC_4 = 0.150$ | $QIBC_{13} = 0.254$ | $QIBC_{19} = 0.30$ | $QIBC_8 = 0.344$ | $QIBC_6 = 0.387$ | $QIBC_{15} = 0.553$ | | |
| 200 | $QIBC_7 = 0.178$<br>$QIBC_{18} = 0.603$ | $QIBC_3 = 0.179$<br>$QIBC_9 = 0.705$ | $QIBC_{13} = 0.287$<br>$QIBC_{16} = 0.713$ | $QIBC_{19} = 0.309$ | $QIBC_6 = 0.397$ | $QIBC_8 = 0.397$ | $QIBC_5 = 0.407$ | $QIBC_{15} = 0.589$ |
| 400 | $QIBC_3 = 0.237$<br>$QIBC_{18} = 0.659$ | $QIBC_7 = 0.264$<br>$QIBC_{16} = 0.768$ | $QIBC_{13} = 0.343$<br>$QIBC_9 = 0.770$ | $QIBC_{20} = 0.406$<br>$QIBC_{11} = 0.802$ | $QIBC_6 = 0.484$ | $QIBC_8 = 0.513$ | $QIBC_{15} = 0.635$ | |

**TABLE 4.** F-measure Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on 20 newsgroup dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | $QIBC_2 = 0.044$ | $QIBC_1 = 0.202$ | $QIBC_{20} = 0.212$ | $QIBC_{18} = 0.267$ | | | | |
| 10 | $QIBC_4 = 0.1141$ | $QIBC_7 = 0.121$ | $QIBC_{17} = 0.244$ | $QIBC_{20} = 0.350$ | | | | |
| 15 | $QIBC_1 = 0.398$ | | | | | | | |
| 20 | $QIBC_8 = 0.198$ | $QIBC_{19} = 0.291$ | | | | | | |
| 30 | $QIBC_4 = 0.130$ | $QIBC_7 = 0.139$ | $QIBC_{19} = 0.329$ | $QIBC_{14} = 0.369$ | $QIBC_{15} = 0.517$ | $QIBC_{12} = 0.619$ | $QIBC_{11} = 0.756$ | |
| 40 | $QIBC_4 = 0.160$<br>$QIBC_{10} = 0.731$ | $QIBC_3 = 0.175$<br>$QIBC_{11} = 0.776$ | $QIBC_7 = 0.176$ | $QIBC_{19} = 0.362$ | $QIBC_{14} = 0.399$ | $QIBC_5 = 0.475$ | $QIBC_{15} = 0.555$ | $QIBC_{18} = 0.646$ |
| 50 | $QIBC_4 = 0.171$<br>$QIBC_{15} = 0.582$ | $QIBC_7 = 0.180$<br>$QIBC_{18} = 0.645$ | $QIBC_3 = 0.188$<br>$QIBC_{12} = 0.685$ | $QIBC_{19} = 0.390$<br>$QIBC_{11} = 0.777$ | $QIBC_6 = 0.419$ | $QIBC_{14} = 0.442$ | $QIBC_5 = 0.483$ | $QIBC_1 = 0.534$ |
| 70 | $QIBC_7 = 0.190$<br>$QIBC_{12} = 0.707$ | $QIBC_4 = 0.215$<br>$QIBC_{10} = 0.774$ | $QIBC_8 = 0.281$<br>$QIBC_{11} = 0.794$ | $QIBC_{19} = 0.398$ | $QIBC_{20} = 0.443$ | $QIBC_6 = 0.459$ | $QIBC_{14} = 0.536$ | $QIBC_{18} = 0.658$ |
| 100 | $QIBC_7 = 0.226$<br>$QIBC_{11} = 0.813$ | $QIBC_{16} = 0.302$ | $QIBC_{19} = 0.399$ | $QIBC_8 = 0.409$ | $QIBC_{20} = 0.470$ | $QIBC_6 = 0.493$ | $QIBC_{15} = 0.649$ | $QIBC_{18} = 0.678$ |
| 150 | $QIBC_4 = 0.232$ | $QIBC_{13} = 0.357$ | $QIBC_{19} = 0.400$ | $QIBC_8 = 0.475$ | $QIBC_6 = 0.527$ | $QIBC_1 = 0.590$ | $QIBC_{15} = 0.663$ | |
| 200 | $QIBC_3 = 0.275$ | $QIBC_{13} = 0.386$ | $QIBC_{19} = 0.399$ | $QIBC_8 = 0.518$ | $QIBC_5 = 0.527$ | $QIBC_6 = 0.533$ | $QIBC_{15} = 0.665$ | |
| 400 | $QIBC_7 = 0.304$ | $QIBC_3 = 0.324$ | $QIBC_{13} = 0.401$ | $QIBC_6 = 0.578$ | $QIBC_8 = 0.580$ | $QIBC_{15} = 0.651$ | $QIBC_{11} = 0.797$ | |

performs for the hard topics, where others cannot perform, for example DT. Interestingly, QIBC performed slightly better compared to DT for several topics, i.e. topics 17, 5, 2, 3, 7, 9, 10, 11, 14, 16, 18 and 20, when the top 150 features were selected as can be seen in Figure 8(j). The performances of other baselines such as KNN, NB and SVM were low for most

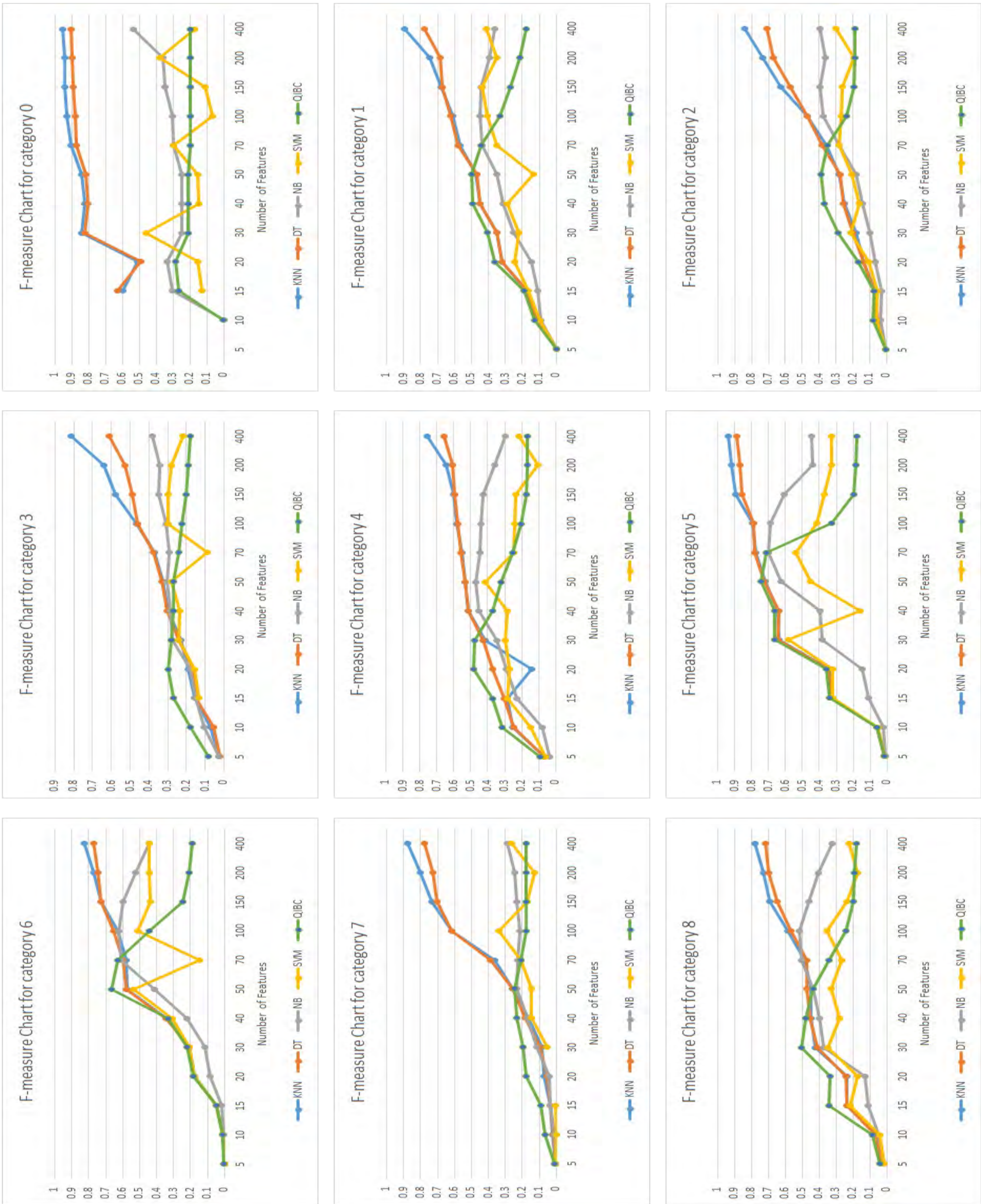


**FIGURE 6.** F-measure chart for each category by changing number of features 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400 among KNN, DT, NB, SVM and QIBC on MNIST handwritten image dataset.

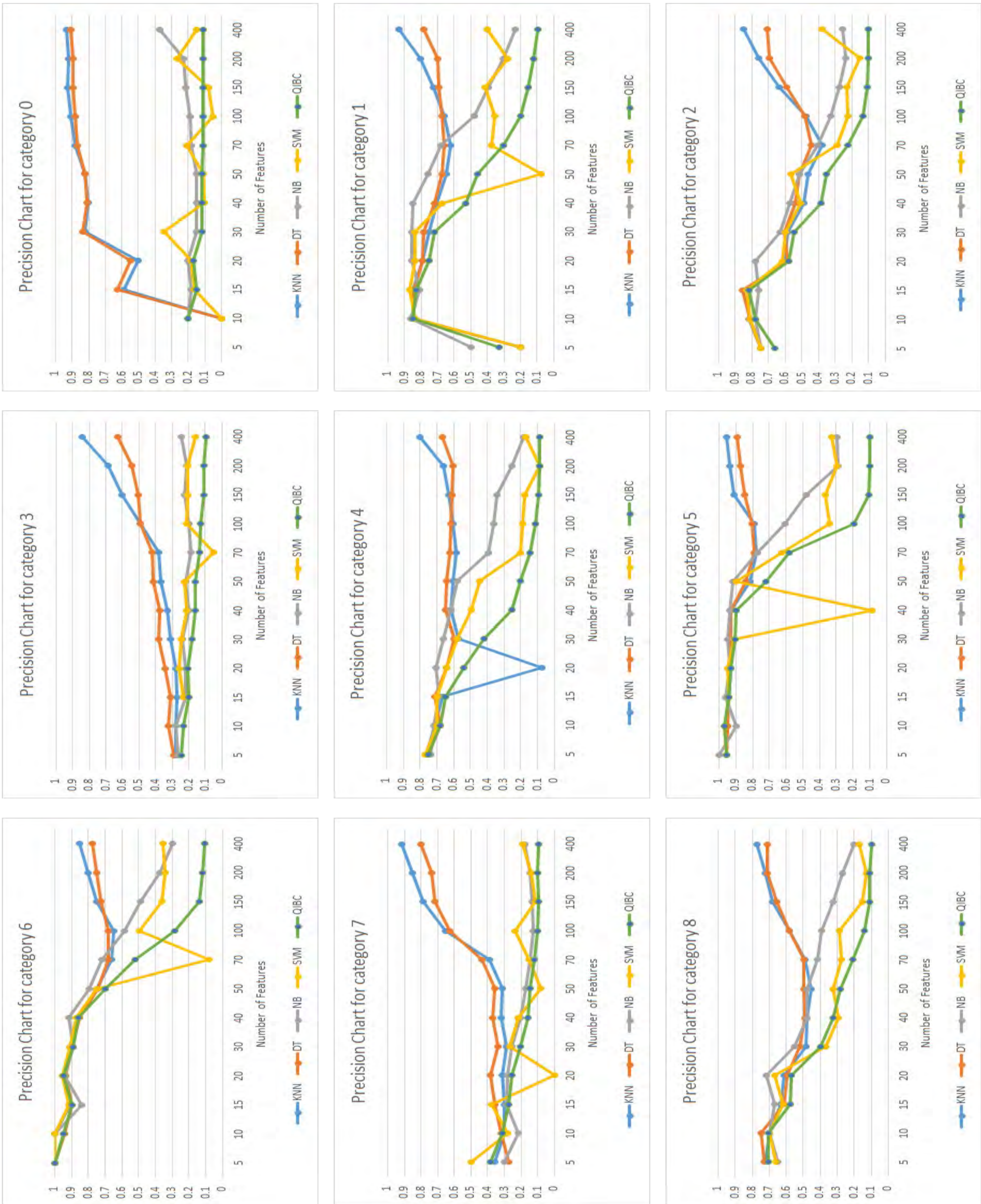


**FIGURE 7.** **Precision chart for each category by changing number of features 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400 among KNN, DT, NB, SVM and QIBC on MNIST handwritten image dataset.**

**TABLE 5.** Recall Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on MNIST handwritten image dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | $QIBC_0 = 0.00$ $QIBC_8 = 0.022$ | $QIBC_1 = 0.0007$ | $QIBC_2 = 0.0014$ | $QIBC_3 = 0.051$ | $QIBC_4 = 0.049$ | $QIBC_5 = 0.008$ | $QIBC_6 = 0.001$ | $QIBC_7 = 0.0066$ |
| 10 | $QIBC_0 = 0.001$ $QIBC_8 = 0.047$ | $QIBC_1 = 0.07$ | $QIBC_2 = 0.041$ | $QIBC_3 = 0.147$ | $QIBC_4 = 0.208$ | $QIBC_5 = 0.032$ | $QIBC_6 = 0.006$ | $QIBC_7 = 0.038$ |
| 15 | $QIBC_0 = 0.998$ | $QIBC_1 = 0.240$ | $QIBC_2 = 0.097$ | $QIBC_3 = 0.531$ | $QIBC_5 = 0.221$ | $QIBC_6 = 0.104$ | $QIBC_7 = 0.140$ | $QIBC_8 = 0.237$ |
| 20 | $QIBC_0 = 0.00$ $QIBC_8 = 0.022$ | $QIBC_1 = 0.0007$ | $QIBC_2 = 0.0014$ | $QIBC_3 = 0.051$ | $QIBC_4 = 0.049$ | $QIBC_5 = 0.008$ | $QIBC_6 = 0.001$ | $QIBC_7 = 0.0066$ |
| 30 | $QIBC_0 = 1.00$ $QIBC_8 = 0.721$ | $QIBC_1 = 0.283$ | $QIBC_2 = 0.2$ | $QIBC_3 = 0.626$ | $QIBC_4 = 0.56$ | $QIBC_5 = 0.522$ | $QIBC_6 = 0.128$ | $QIBC_7 = 0.189$ |
| 40 | $QIBC_0 = 0.99$ | $QIBC_1 = 0.466$ | $QIBC_2 = 0.354$ | $QIBC_3 = 0.792$ | $QIBC_4 = 0.691$ | $QIBC_7 = 0.402$ | $QIBC_8 = 0.911$ | |
| 50 | $QIBC_0 = 0.99$ $QIBC_8 = 0.943$ | $QIBC_1 = 0.554$ | $QIBC_2 = 0.429$ | $QIBC_3 = 0.829$ | $QIBC_4 = 0.756$ | $QIBC_5 = 0.761$ | $QIBC_6 = 0.634$ | $QIBC_7 = 0.71$ |
| 70 | $QIBC_0 = 1.00$ $QIBC_8 = 0.968$ | $QIBC_1 = 0.823$ | $QIBC_2 = 0.828$ | $QIBC_3 = 0.929$ | $QIBC_4 = 0.878$ | $QIBC_5 = 0.928$ | $QIBC_6 = 0.801$ | $QIBC_7 = 0.937$ |
| 100 | $QIBC_0 = 1.00$ $QIBC_8 = 0.994$ | $QIBC_1 = 0.953$ | $QIBC_2 = 0.989$ | $QIBC_3 = 0.95$ | $QIBC_4 = 0.97$ | $QIBC_5 = 0.987$ | $QIBC_6 = 0.964$ | $QIBC_7 = 0.999$ |
| 150 | $QIBC_0 = 1.00$ $QIBC_8 = 1.00$ | $QIBC_1 = 0.989$ | $QIBC_2 = 0.999$ | $QIBC_3 = 0.987$ | $QIBC_4 = 0.989$ | $QIBC_5 = 1.00$ | $QIBC_6 = 0.998$ | $QIBC_7 = 1.00$ |
| 200 | $QIBC_0 = 1.00$ $QIBC_8 = 1.00$ | $QIBC_1 = 0.998$ | $QIBC_2 = 1.00$ | $QIBC_3 = 0.997$ | $QIBC_4 = 1.00$ | $QIBC_5 = 1.00$ | $QIBC_6 = 0.999$ | $QIBC_7 = 1.00$ |
| 400 | $QIBC_0 = 1.00$ $QIBC_8 = 1.00$ | $QIBC_1 = 1.00$ | $QIBC_2 = 1.00$ | $QIBC_3 = 1.00$ | $QIBC_4 = 1.00$ | $QIBC_5 = 1.00$ | $QIBC_6 = 1.00$ | $QIBC_7 = 1.00$ |

**TABLE 6.** F-measure Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on MNIST handwritten image dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | $QIBC_1 = 0.0014$ $QIBC_8 = 0.042$ | $QIBC_2 = 0.0028$ | $QIBC_3 = 0.084$ | $QIBC_4 = 0.092$ | $QIBC_5 = 0.016$ | $QIBC_6 = 0.002$ | $QIBC_7 = 0.012$ | |
| 10 | $QIBC_0 = 0.002$ $QIBC_8 = 0.088$ | $QIBC_1 = 0.129$ | $QIBC_2 = 0.078$ | $QIBC_3 = 0.179$ | $QIBC_4 = 0.318$ | $QIBC_5 = 0.061$ | $QIBC_6 = 0.012$ | $QIBC_7 = 0.068$ |
| 15 | $QIBC_1 = 0.189$ | $QIBC_2 = 0.074$ | $QIBC_3 = 0.268$ | $QIBC_4 = 0.372$ | $QIBC_5 = 0.339$ | $QIBC_6 = 0.05$ | $QIBC_7 = 0.094$ | $QIBC_8 = 0.34$ |
| 20 | $QIBC_1 = 0.364$ $QIBC_8 = 0.335$ | $QIBC_2 = 0.166$ | $QIBC_3 = 0.299$ | $QIBC_4 = 0.483$ | $QIBC_5 = 0.357$ | $QIBC_6 = 0.187$ | $QIBC_7 = 0.18$ | |
| 30 | $QIBC_1 = 0.406$ $QIBC_8 = 0.507$ | $QIBC_2 = 0.29$ | $QIBC_3 = 0.279$ | $QIBC_4 = 0.478$ | $QIBC_5 = 0.662$ | $QIBC_6 = 0.224$ | $QIBC_7 = 0.197$ | |
| 40 | $QIBC_1 = 0.496$ | $QIBC_2 = 0.368$ | $QIBC_7 = 0.231$ | $QIBC_8 = 0.48$ | | | | |
| 50 | $QIBC_1 = 0.502$ | $QIBC_2 = 0.387$ | $QIBC_5 = 0.742$ | $QIBC_6 = 0.663$ | | | | |
| 70 | $QIBC_6 = 0.63$ | | | | | | | |

of the topics in addition to the fact that KNN and SVM take a long time for computation as well.

In terms of recall, the performance of QIBC improved by increasing the number of top selected features as can be seen in Figure 9. For example, if we consider the top 50 features then QIBC outperformed the baselines for most topics, i.e. topics 4, 7, 3, 19, 20, 6, 14, 5, 15, 18, 12 and 11 as can be seen in Figure 9(g). DT again failed to perform for most of the topics. The QIBC performance increased with an increasing number of features as can be seen in Figure 9, although it performed better for few topics when the top 150 features were selected as can be seen in Figure 9(k). KNN always performed better for topic 1.

In terms of F-measure, QIBC performance started improving by increasing the number of top selected features as can be seen in Figure 10. QIBC performance started improving for top selected features, i.e. the top 30, 40, 50, 70 and so on. For example, if we consider the top 50 features then QIBC outperformed the baselines for most topics, i.e. topics 4, 7, 3, 19, 20, 6, 14, 5, 1, 15, 18 and 11 as can be seen in Figure 10(g). QIBC outperformed more than half of the topics when the top 40, 50, 70 and 100 features were selected. QIBC performed better for only 6 topics when the top 150 features were selected. When the top 200 and 400 features were selected then QIBC performed better for 7 and 8 topics out of 20. As with the increasing number of selected features, computation time also increased for all baselines as well as for QIBC but QIBC is more effective because it requires less computation time with the better measures.

**TABLE 7.** Precision Table where QIBC outperform all the baselines (i.e. $QIBC_t$ stands for the performance of QIBC for the topic or category $t$) on MNIST handwritten image dataset for all range of features, i.e. top 5, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200 and 400.

| | | |
|---|---|---|
| 5 | $QIBC_6 = 1.00$ | |
| 10 | $QIBC_0 = 0.2$ | $QIBC_5 = 0.97$ |
| 20 | $QIBC_6 = 0.95$ | |


**FIGURE 8.** Precision chart based on top selected features from top 5 to top 400 among KNN, DT, NB, SVM and QIBC on 20 Newsgroup Text Corpora.

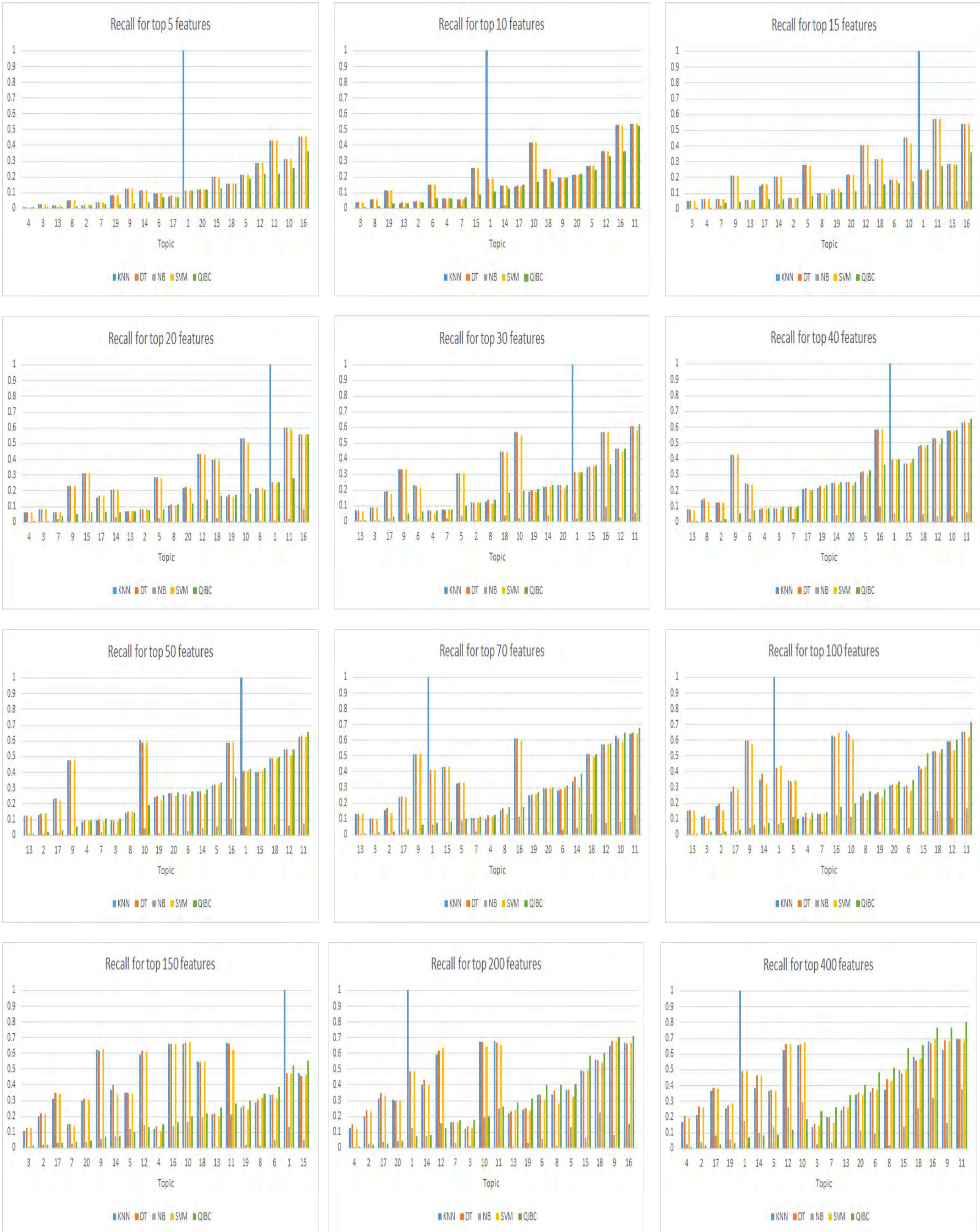


**FIGURE 9.** Recall chart based on top selected features from top 5 to top 400 among KNN, DT, NB, SVM and QIBC on 20 Newsgroup Text Corpora.

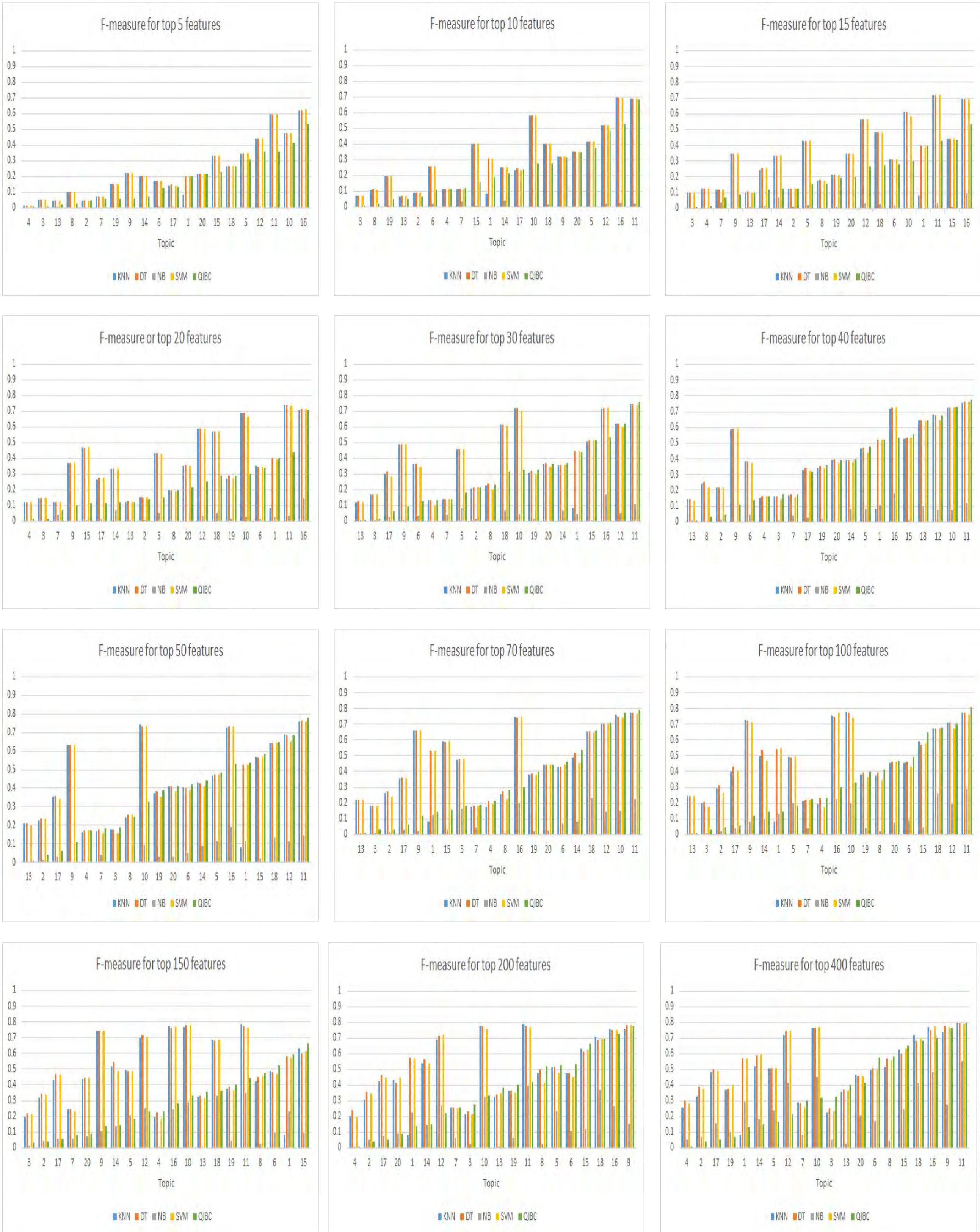


**FIGURE 10.** **F-measure chart based on top selected features from top 5 to top 400 among KNN, DT, NB, SVM and QIBC on 20 Newsgroup Text Corpora.**

| Features | KNN | DT | NB | SVM | QIBC |
|---|---|---|---|---|---|
| 5 | 31.7634 | 0.4756 | 1.0224 | 1841.3 | **2.0952** |
| 10 | 48.5518 | 0.6199 | 1.0682 | 2702.7 | **2.1392** |
| 15 | 45.8418 | 0.7878 | 1.1197 | 3659.4 | **2.2417** |
| 20 | 57.7713 | 1.1635 | 1.2810 | 5348.7 | **2.6347** |
| 30 | 81.7262 | 1.8554 | 1.4868 | 5924.3 | **3.2154** |
| 40 | 121.1638 | 2.8600 | 1.8492 | 6984.8 | **3.1704** |
| 50 | 147.7935 | 3.6617 | 2.0850 | 8107.1 | **3.7973** |
| 70 | 215.7735 | 6.0609 | 2.5716 | 11022 | **5.6506** |
| 100 | 321.4240 | 9.5287 | 3.3811 | 17017 | **10.0570** |
| 150 | 486.7824 | 13.7031 | 4.7636 | 25687 | **19.2916** |
| 200 | 639.0312 | 17.3438 | 6.0727 | 32468 | **29.1939** |
| 400 | 1199.4 | 29.1856 | 11.1044 | 54941 | **166.8406** |

(a)

| Features | KNN | DT | NB | SVM | QIBC |
|---|---|---|---|---|---|
| 5 | 5.2256 | 0.2455 | 0.2304 | 10.466 | **0.8577** |
| 10 | 8.0862 | 0.2795 | 0.2823 | 10.1312 | **0.8983** |
| 15 | 7.3662 | 0.3689 | 0.3456 | 10.4679 | **0.9458** |
| 20 | 8.6009 | 0.4472 | 0.4567 | 10.5934 | **1.136** |
| 30 | 12.3551 | 0.6619 | 0.5872 | 10.8999 | **1.397** |
| 40 | 15.6101 | 0.9297 | 0.7308 | 10.9707 | **1.7699** |
| 50 | 20.1709 | 1.3405 | 0.8987 | 11.778 | **2.1004** |
| 70 | 26.823 | 2.0612 | 1.1737 | 12.4184 | **2.6618** |
| 100 | 39.6668 | 3.4751 | 1.7378 | 14.6174 | **3.6232** |
| 150 | 62.1047 | 6.2484 | 2.3273 | 17.0815 | **6.5877** |
| 200 | 86.231 | 9.7856 | 3.0617 | 20.07 | **13.237** |
| 400 | 172.9039 | 25.7639 | 6.0116 | 30.1179 | **101.2075** |

(b)

**FIGURE 11. Computation Time in second of each classifier for MNIST Handwritten Image Dataset and 20 Newsgroup Text Corpora among KNN, DT, NB, SVM, QIBC. (a) Computation Time in second among KNN, DT, NB, SVM and QIBC on MNIST Handwritten Image Dataset. (b) Computation Time in second among KNN, DT, NB, SVM and QIBC on 20 Newsgroup Text Corpora.**

### *E. CASE STUDY*

QIBC output varies by changing the threshold value of $\langle x|\Delta|x\rangle$ from 0.5 to something else. Some case studies have been done on the MNIST handwritten image dataset as can be seen in Figure 12. QIBC precision is optimal for categories 0 and 8 when $\langle x|\Delta|x\rangle > 0.5$. QIBC precision is higher for categories 1 and 4 when $\langle x|\Delta|x\rangle > 0.4$. Precision is higher for category 2 when $\langle x|\Delta|x\rangle > 0.6$ and has a similar precision for categories 3 and 7. Precision is higher for categories 5 and 6 when $\langle x|\Delta|x\rangle > 0.7$. In terms of recall, recall is similar in all cases for category 0. Recall rate is higher for categories 1 and 4 when $\langle x|\Delta|x\rangle > 0.6$. Recall is higher for categories 2 and 3 when $\langle x|\Delta|x\rangle > 0.7$. Again, there is higher recall for categories 5 and 8 when $\langle x|\Delta|x\rangle > 0.4$. Recall for category 6 is also higher when $\langle x|\Delta|x\rangle > 0.5$. They have similar F-measure for categories 0 and 5 in all four cases. F-measure is higher for categories 1 and 3 when $\langle x|\Delta|x\rangle > 0.5$. F-measure rate is higher for categories 2 and 4 when $\langle x|\Delta|x\rangle > 0.7$ and so on.

We also experimented on the 20 Newsgroups dataset as can be seen in Figure 13. In terms of precision, QIBC precision is higher for categories 1 and 5 when $\langle x|\Delta|x\rangle > 0.4$. QIBC has similar precision for categories 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19 and 20. Precision is higher for category 6 when $\langle x|\Delta|x\rangle\rangle > 0.6$ and 0.7, and higher for

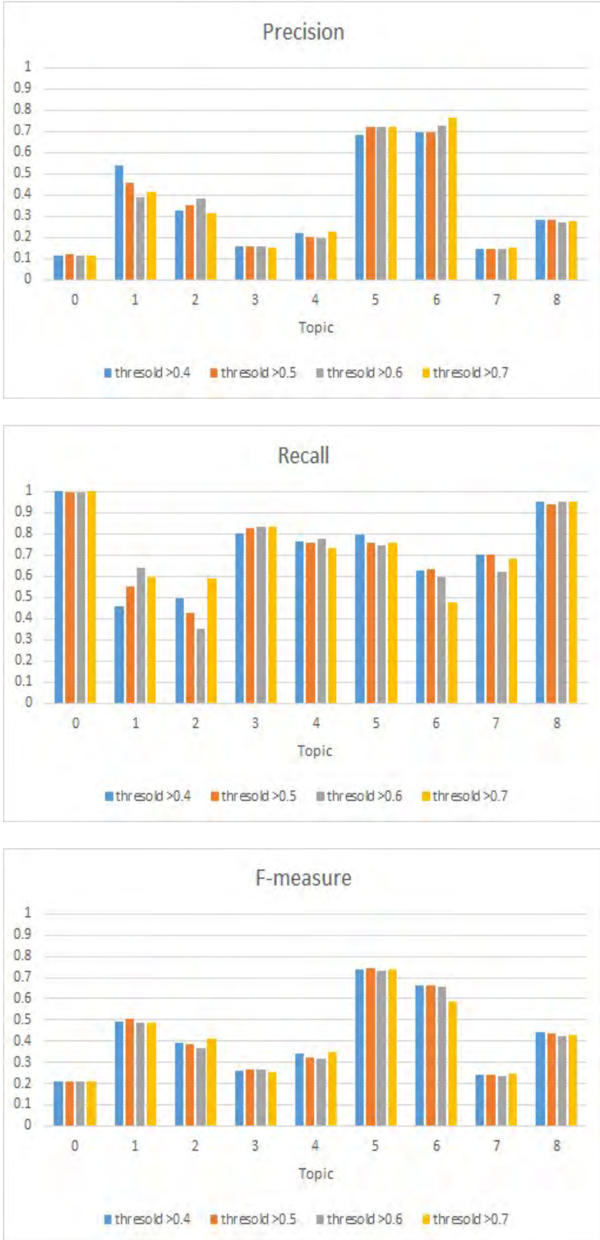


**FIGURE 12. Threshold chart of QIBC by changing the $\langle x|\Delta|x\rangle > 0.4$, 0.5, 0.6 and 0.7 on MNIST handwritten image dataset.**

category 17 when $\langle x|\Delta|x\rangle > 0.6$. Recall is higher for categories 1, 2, 3, 5, 6, 9, 10, 14, 15 and 16 when $\langle x|\Delta|x\rangle > 0.4$. Recall is similar for categories 4, 7, 8, 11, 12, 13, 17, 18, 19 and 20 in all cases. F-measure is higher for categories 1, 2, 3, 5, 6, 9, 10, 14, 15, 16, and 17 for each case. F-measure is similar for categories 4, 7, 8, 11, 12, 13, 18, 19 and 20 for each case. So the main reason for this case study was to check the deviation in the measures by changing threshold values. It is possible to get high measures also in certain circumstances as can be seen throughout the case study.

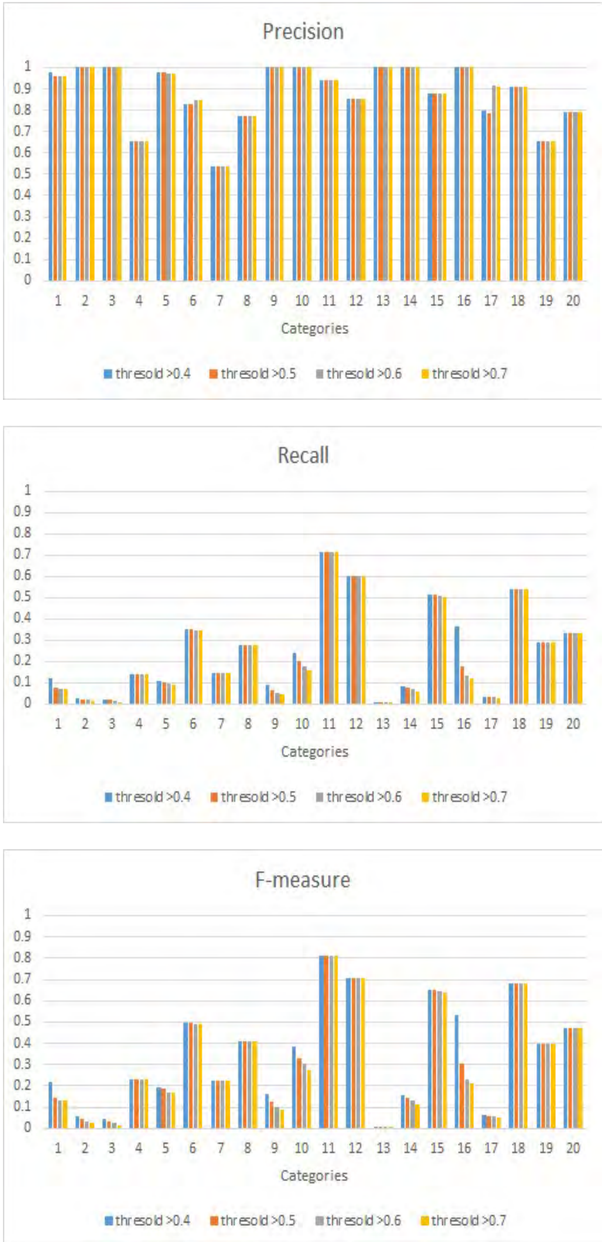


**FIGURE 13. Threshold chart of QIBC by changing the $\langle x|\Delta|x\rangle > 0.4$, 0.5, 0.6 and 0.7 on 20 Newsgroup Dataset.**

## VI. CONCLUSIONS AND FUTURE WORKS

QIBC outperformed the baselines for several categories. We investigated how QIBC performs better in the range of features for example, recall increases with an increasing number of features and F-measure starts decreasing with an increasing number of features for the image dataset. This outcome encourages us to further investigate the world of quantum-inspired ML frameworks. However, some results are still unsatisfactory for some categories (topics).

In order to understand the reason as to why our proposed model fails for some topics and performs better for other topics, we need to do a micro-analysis and find the main reason for failure and success. This analysis might help us to improve our model. Our proposed framework is just the gateway to the quantum-inspired ML world and we need to put more effort into understanding the effectiveness of QM in order to develop more efficient algorithms.

We implemented our model on text corpora as well as on the image corpora to check the performance. We achieved high precision for several topics (categories) when text corpora was used as well as high recall with F-measure for most of the topics (categories). We achieved high recall for every category as well as high F-measure for most categories when image corpora was used. The result obtained from the image corpora can be beneficial in several domains, i.e. patent search and biomedical image classification, where recall is crucial for tasks aiming to find all the pertinent items of a class.

We addressed binary classification. Although it is not a limitation, we are trying to implement a multi-class classifier inspired by quantum detection theory for multi-class classification problems. Besides, we are also working on multi-label classification tasks in order to come up with multi-label classifiers.

## REFERENCES

[1] J. Kepler, *The Secret of the Universe*, vol. 9. Norwalk, CT, USA: Abaris Books, 1981.
[2] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.
[3] P. Jarratt and C. Mack, "A least squares method for Laplace's equation with Dirichlet boundary conditions," *Comput. J.*, vol. 11, no. 1, pp. 83–87, 1968.
[4] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[5] P. G. Ciarlet and P. A. Raviart, "General Lagrange and Hermite interpolation in $R^n$ with applications to finite element methods," *Arch. Rational Mech. Anal.*, vol. 46, no. 3, pp. 177–199, 1972.
[6] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, p. 386, 1958.
[7] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards, *Artificial Intelligence: A Modern Approach*, vol. 3. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.
[8] C. Robert, "Machine learning, a probabilistic perspective," *CHANCE*, vol. 27, no. 2, pp. 62–63, 2014. doi: 10.1080/09332480.2014.914768.
[9] E. Alpaydin, *Introduction To Machine Learning*. Cambridge, MA, USA: MIT Press, 2009.
[10] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.
[11] G. Salton, and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, 1986.
[12] K. S. Jones, *Automatic Keyword Classification for Information Retrieval*. London, U.K.: Butterworths, 1971.
[13] R. R. Larson, "Introduction to information retrieval," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 4, pp. 852–853, 2010.
[14] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, vol. 283. Boston, MA, USA: Addison-Wesley, 2010.
[15] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[16] N. Craswell, W. B. Croft, M. de Rijke, J. Guo, and B. Mitra, "Neural information retrieval: Introduction to the special issue," *Inf. Retr. J.*, vol. 21, nos. 2–3, pp. 107–110, 2018.
[17] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. Berlin, Germany: Springer-Verlag, 2010.
[18] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432.

[19] X. Yang, C. Macdonald, and I. Ounis, "Using word embeddings in twitter election classification," *Inf. Retr. J.*, vol. 21, nos. 2–3, pp. 183–207, 2018.
[20] I. Vulić and M. F. Moens, "Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* New York, NY, USA: ACM, Aug. 2015, pp. 363–372.
[21] K. Golub, D. Soergel, G. Buchanan, D. Tudhope, M. Lykke, and D. Hiom, "A framework for evaluating automatic indexing or classification in the context of retrieval," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 1, pp. 3–16, 2016.
[22] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[23] N. S. Yanofsky and M. A. Mannucci, *Quantum Computing for Computer Scientists*, vol. 20. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[24] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nature Phys.*, vol. 10, no. 9, pp. 631–633, 2014.
[25] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. New York, NY, USA: Academic, 2014.
[26] S. Aaronson, "Read the fine print," *Nature Phys.*, vol. 11, no. 4, pp. 291–293, 2015.
[27] K. P. Murphy, *Naive Bayes Classifiers*, vol. 18. Vancouver, BC, Canada: Univ. British Columbia, 2006.
[28] D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991.
[29] I. Steinwart and A. Christmann, *Support Vector Machines*, 1st ed. Berlin, Germany: Springer-Verlag, 2008.
[30] H. Abdi, "Signal detection theory (SDT)," *Encyclopedia of Measurement and Statistics*, N. Salkind, Ed., 2007.
[31] J. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics*. New York, NY, USA: Psychology Press, 1996. doi: 10.4324/9781315806167.
[32] T. D. Wickens, *Elementary Signal Detection Theory*. New York, NY, USA: Oxford University Press, 2002.
[33] J. Neyman and E. S. Pearson, "The testing of statistical hypotheses in relation to probabilities a priori," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 29, no. 4. Cambridge, U.K.: Cambridge Univ. Press, pp. 492–510, 1933.
[34] J. Neyman and E. S. Pearson, "IX. On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London. Ser. A*, vol. 231, nos. 694–706, pp. 289–337, 1933.
[35] J. Busemeyer and P. D. Bruza, *Quantum Models of Cognition and Decision*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[36] I. Pitowsky, *Quantum Probability Quantum Logic*, 1989.
[37] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," *Original Scientific Papers Wissenschaftliche Originalarbeiten*. Berlin, Germany: Springer, 1985, pp. 478–504.
[38] G. Cariolaro, *Quantum Communications*. Berlin, Germany: Springer, 2015.
[39] C. W. Helstrom, "Quantum detection and estimation theory," *J. Statist. Phys.*, vol. 1, no. 2, pp. 231–252, 1969.
[40] C. W. Helstrom, *Quantum Detection and Estimation Theory*. California, MA, USA: California Univ., 1971.
[41] C. W. Helstrom, *Quantum Detection and Estimation Theory*. New York, NY, USA: Academic, 1976, pp. 74–83.
[42] M. Melucci, "Relevance feedback algorithms inspired by quantum detection," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 1022–1034, Apr. 2016.
[43] N. Nasios and A. G. Bors, "Kernel-based classification using quantum mechanics," *Pattern Recognit.*, vol. 40, no. 3, pp. 875–889, 2007.
[44] X. Han, L. Quan, X. Xiong, and B. Wu, "Facing the classification of binary problems with a hybrid system based on quantum-inspired binary gravitational search algorithm and K-NN method," *Eng. Appl. Artif. Intell.*, vol. 26, no. 10, pp. 2424–2430, 2013.
[45] G. Sergioli, G. M. Bosyk, E. Santucci, and R. Giuntini, "A quantum-inspired version of the classification problem," *Int. J. Theor. Phys.*, vol. 56, no. 12, pp. 3880–3888, 2017.
[46] A. G. de Pinho, M. Vellasco, and A. V. A. da Cruz, "A new model for credit approval problems: A quantum-inspired neuro-evolutionary algorithm with binary-real representation," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 445–450.
[47] A. V. Abs da Cruz, M. M. B. R. Vellasco, and M. A. C. Pacheco, *Quantum-Inspired Evolutionary Algorithm for Numerical Optimization* (Computational Intelligence Book Quantum Inspired Intelligent Systems), vol. 121. Berlin, Germany: Springer, 2008, pp. 115–132.
[48] P. Kaye, R. Laflamme, and M. Mosca, *An Introduction to Quantum Computing*. New York, NY, USA: Oxford Univ. Press, 2007.

**PRAYAG TIWARI** received the M.S. degree from NUST MISIS, Moscow. He is currently pursuing the Ph.D. degree with the University of Padova, Italy. He was a Research Assistant with NUST MISIS and he has had teaching and industrial work experience. He is currently a Marie Sklodowska-Curie Researcher with the University of Padova, Italy. He has several publications in journal, book series and conferences of IEEE, ACM, Springer, Elsevier, MDPI, Taylor Francis, and IGI-Global. His research interests include machine learning, deep learning, quantum inspired machine learning, and information retrieval.

**MASSIMO MELUCCI** received the degree in statistical science and the Ph.D. degree in computer engineering from the University of Padova, Italy, in 1996, where he is currently an Associate Professor with the Department of Information Engineering.

Since 1990, he has been carrying out research in information retrieval (IR), that is, the automated retrieval of information relevant to user needs. In particular, he is interested in modeling, implementing and experimenting advanced methods for indexing, retrieving, and ranking documents, recently inspired by physics and engineering disciplines. He has authored more than about 170 publications and in particular two research books. He has been researching the correspondence between IR and quantum mechanics (QM) both from a theoretical and an experimental perspective, because effective contextual IR systems can be designed within the theoretical framework of QM. Indeed, QM provides a theoretical framework which well describes how contextual factors influence the user information needs and the retrieval of relevant documents. At the intersection between QM and detection theory, he found that the maximum effectiveness stated by the probability ranking principle for IR can be further improved without additional evidence for estimating the parameters of a probabilistic model.

Dr. Melucci co-ordinated the EU VII FP Marie Curie Action IRSES project QONTEXT, from 2010 to 2013, and he is currently coordinating the H2020 Marie Sklodowska Cure Action ITN project quantum information access and retrieval theory (QUARTZ), running from 2017 until 2020. He teaches databases, information retrieval, and computing methods for data science at the University of Padova.