



A quantum procedure for estimating information gain in Boolean classification task

Michal Koren¹ · Or Peretz¹

Received: 28 September 2023 / Accepted: 17 February 2024
© The Author(s) 2024

Abstract

A substantial portion of global quantum computing research has been conducted using quantum mechanics, which recently has been applied to quantum computers. However, the design of a quantum algorithm requires a comprehensive understanding of quantum mechanics and physical procedures. This work presents a quantum procedure for estimating information gain. It is aimed at making quantum computing accessible to those without preliminary knowledge of quantum mechanics. The procedure can be a basis for building data mining processes according to measures from information theory using quantum computers. The main advantage of this procedure is the use of amplitude encoding and the inner product of two quantum states to calculate the conditional entropy between two vectors. The method was implemented using the IBM simulator and tested over a dataset of six features and a Boolean target variable. The results showed a correlation of 0.942 between the ranks achieved by the classical and quantum computations with a significance of $p < 0.005$.

Keywords Information gain · Entropy · Decision trees · Information theory

1 Introduction and related work

Quantum computing (QC) stands out as a highly promising domain within computation, commanding a significant presence in global research efforts (Ying 2010). At its core, QC relies on the foundational principle of quantum physics, positing that electrons can simultaneously exhibit wave-like and particle-like properties (Robertson 1943). However, the development and maintenance of quantum computers include formidable challenges due to their vulnerability to noise and anomalies (Bennett et al. 1997; De Wolf 2019). Many technological firms have been involved in quantum computing and invested heavily in developing this industry (Zeng et al. 2017). Over the years, a broader view of the pros and cons of QC has emerged, and this discussion remains relevant in modern times (Boyer et al. 1998; De Wolf 2019).

Compared to classical computing, QC demonstrates the potential to reduce computational complexity, enabling extensive simultaneous processing, a concept well supported by research (Biamonte et al. 2017; Wiebe 2020). However, algorithms are not limited to classical or quantum computers; they can work together to achieve better results (Buffoni and Caruso 2021). Consequently, combining QC and classical computing has given rise to a growing discipline called Quantum Machine Learning (QML). The transformation of classical machine learning algorithms into quantum equivalents requires translating classical algorithmic logic into circuits composed of quantum gates (Benedetti et al. 2019; Alchieri et al. 2021). Recent studies have published various quantum algorithm applications, including learning the behavior of random variables (González et al. 2022; Pirhooshyaran and Terlaky 2021), the development of quantum convolutional networks for image learning (Hur et al. 2022; Tüysüz et al. 2021), the creation of generative adversarial networks (GANs) and transfer learning (Assouel et al. 2022; Azevedo et al. 2022; Zoufal et al. 2021), as well as the implementation of reinforcement learning (Dalla Pozza et al. 2022).

In physics, entropy plays a vital role in characterizing the uncertainty in the state of matter (Bein 2006). Due to the rapid development of information technology, entropy has gained importance in information theory in recent years. Consequently, considering the amount of

✉ Michal Koren
michal.koren@shenkar.ac.il

Or Peretz
or.peretz@shenkar.ac.il

¹ School of Industrial Engineering and Management, Shenkar—Engineering, Design, Art, Anne Frank 12, 5252317 Ramat Gan, Israel

data in events, random variables and probability, considerable information can be identified through distributions of data (Wehr 1978). These information measures have significant applications in artificial intelligence and machine learning, such as constructing decision trees and optimizing classifier models (Kapur and Kesavan 1992). Additionally, entropy is an essential metric in data mining and machine learning that indicates model inconclusiveness or inaccuracy. Generally, low entropy indicates that valuable information can be easily extracted from the data, while high entropy indicates a more significant challenge to generate meaningful insights (Kaufmann et al. 2020; Kaufmann and Vecchio 2020; Liu et al. 2022).

When dealing with a wide variety of data types, calculating entropy using a quantum computer can be challenging. First, it is necessary to ensure that the input is encoded in such a way that it fits the constraints of quantum states. Second, a series of quantum gates must be constructed to approximate the input data's entropy. As quantum simulators are now becoming more widely available, methods have been developed to calculate the entropy of a random variable using quantum circuits. An example of such a method is the entropy “black box” that uses variable distribution to determine its amplitude encoding and estimate its entropy (Koren et al. 2023).

Decision trees are fundamental elements in machine learning, offering a flexible and easily comprehensible approach to decision-making and predictive tasks (Navada et al. 2011). Their hierarchical structure breaks down complex decision-making processes into simple, often binary, questions accessible to non-specialized individuals. Their ability to understand complex data relationships makes decision trees valuable and reliable for building predictive models (Ahmed and Kim 2017). Furthermore, understanding that a decision tree is not homogeneous is essential, as a series of internal nodes can generate different decision trees. The maximum number of trees that can be generated from given data is exponential (Charbuty and Abdulazeez 2021).

The Iterative Dichotomiser 3 (ID3) algorithm is a traditional method for constructing decision trees (Hssina et al. 2014; Jin et al. 2009). This method follows a divide-and-conquer strategy and uses information gain (IG) to assess the accuracy of the classification (Kent 1983). The value of the IG indicates the rate of entropy reduction and is calculated by the entropy of the distribution of the subtraction in the primary data structure. Consequently, higher IG values indicate a higher percentage of removed entropy (Batra and Agrawal 2018; Guleria et al. 2014). Thus, for each recursive iteration, the algorithm selects the highest IG feature and uses it to build the next step of the tree.

This work presents a quantum procedure for estimating IG in datasets with Boolean target variables. It is a generic procedure that can be applied to data mining processes.

Section 2 describes the procedure's correctness and general implementation using quantum logic circuits. The motivation to carry out this research is expressed in two main aspects. The first refers to the accessibility of quantum computation for the data mining process without the need for prior physical knowledge. The second is the calculation of IG in QC as a basis for building decision tree models and other measures according to information theory. The proposed method's main advantage is the amplitude of its encoding and the inner product of two quantum states used to calculate the conditional entropy between two vectors. Section 3 presents a case study of a simple dataset that compares the proposed method with the classical computer method's results. The results of the proposed method will then be compared using a dataset with six features, which will be presented in Section 4. Last, Section 5 will discuss the main conclusions and suggestions for future directions.

2 Quantum information gain

This section presents a new method for calculating quantum information gain for discrete value features and a Boolean target variable. The proposed method uses the Quantum Entropy “Black Box” (QEBB), which inputs a vector of occurrences and calculates the entropy using amplitude encoding and parameterized vectors (Koren et al. 2023). First, the procedure and its quantum logic will be described. Then, the implementation and correctness of the method will be presented. Table 1 presents the notations used in this study.

2.1 Quantum logic and gates

Let D be a dataset consisting of m features and n records, denoted as $F = \{F_1, \dots, F_m\}$, and a target variable, denoted as Y_D . The method inputs D and a feature denoted as A , such that $A \in F$. At the beginning, the method uses the QEBB to calculate the entropy of Y_D , the initial entropy of the dataset, denoted as $QH(D)$. The method consists of three sub-procedures, as follows:

Table 1 The notation used in this study

Symbol	Remarks
D	Dataset
m	Number of features
$F = \{F_1, \dots, F_m\}$	Set of features in D
Y_D	The Boolean target variable of D
$D_E(A) = \{a_1, \dots, a_p\}$	Set of distinct elements in feature A
$QH(D)$	The quantum entropy of Y_D
$QH_A(D)$	The quantum conditional entropy of feature A
$IG_A(D)$	Information gain of dataset D by feature A

- Classical computer preprocessing** – Given $A \in F$, the method iterates over distinct values in A , denoted as $D_E(A) = \{a_1, \dots, a_p\}$, for $p \leq n$. For each $a_i \in D_E(A)$, the method stores Y_i , the target variable of records such that $A = a_i$. Let $W = \{w_1, \dots, w_p\}$ be a set of proportional parameters, such that each w_i equals $|Y_i|$ and represents the weight of a_i in the dataset.
- Initialization of amplitude vectors** – At this point, the method creates Q_h , a set of all QEBB(Y_i). The algorithm transforms both W, Q_h to an amplitude encoding by concatenating all items into a single amplitude vector. Let \tilde{W}, \tilde{Q} be the amplitude vectors, such that $|\tilde{W}|^2 = |\tilde{Q}|^2 = 1$, satisfies:

$$\tilde{W} = \frac{1}{\sqrt{\sum_{i=1}^p w_i}} \cdot \sum_{i=1}^p \sqrt{w_i} \tag{1a}$$

$$\tilde{Q} = \frac{1}{\sqrt{\sum_{i=1}^p q_i}} \cdot \sum_{i=1}^p \sqrt{q_i} \tag{2a}$$

Thus, each $w_i \in W$ is converted to $\frac{\sqrt{w_i}}{\sqrt{\sum_{i=1}^p w_i}}$ and each $q_i \in Q_h$ to $\frac{\sqrt{q_i}}{\sqrt{\sum_{i=1}^p q_i}}$.

- Quantum operations and output** – The proposed method creates two quantum circuits: first for W with $\lceil \log_2 |W| \rceil + 1$ qubit and second for Q_h with $\lceil \log_2 |Q_h| \rceil + 1$ qubit. (1a), (2a) are set as the initial states, denoted as, $|\psi_w\rangle$ and $|\psi_q\rangle$, respectively. Next, the method applies the U gate with the parameters $\theta = \frac{\pi}{2}, \phi = 0, \lambda = \pi$ for both circuits to move the states into superposition, which is equivalent to applying the Hadamard gate. The inner product of the states, denoted as $\langle \psi_w | \psi_q \rangle$, can be used to measure the overlap between the state vectors ψ_w, ψ_q . The probability of observing the system in the state ψ_w , given that it is in state ψ_q , can be calculated by $|\langle \psi_w | \psi_q \rangle|^2$. Thus, the inner product of both states describes the amount of conditional entropy achieved in the sub-dataset of feature A . Subtracting $|\langle \psi_w | \psi_q \rangle|^2$ from the original entropy of the dataset, $QH(D)$ yields the information gain achieved in dataset D by feature A .

ClassicalComputerPreprocessing(D,A)

- $W, Y \leftarrow \emptyset$
- $D_E(A) \leftarrow \{a_1, \dots, a_p\}$ set of distinct elements
- For each $a_i \in D_E(A)$:
 - $Y_i \leftarrow$ target values of records, such that $A = f_i$
 - $W \leftarrow W \cup \{|Y_i|\}$
 - $Y \leftarrow Y \cup \{Y_i\}$
- **Return** W, Y

QuantumInformationGain(D,A)

- $QH(D) \leftarrow$ QEBB(Y_D)
- $W, Y \leftarrow$ ClassicalComputerPreprocessing(D,A)
- $Q_h \leftarrow \{ \text{QEBB}(Y_i) \mid Y_i \in Y \}$
- $|\psi_w\rangle, |\psi_q\rangle \leftarrow \left\{ \frac{\sqrt{w_i}}{\sqrt{\sum_{w_i \in W} w_i}} \mid w_i \in W \right\}, \left\{ \frac{\sqrt{q_i}}{\sqrt{\sum_{q_i \in Q_h} q_i}} \mid q_i \in Q_h \right\}$
- Apply $U \left(\theta = \frac{\pi}{2}, \phi = 0, \lambda = \pi \right)$ in both states $|\psi_w\rangle, |\psi_q\rangle$
- Return $QH(D) - |\langle \psi_w | \psi_q \rangle|^2$

Table 2 Legend and description for Fig. 1

Step	Input/output	Description
1	Input dataset and an Attribute (A)	
2	Input step (1) and yield the quantum entropy (Q_h) and weight vector (W)	Execution of the classic computer preprocessing and QEBB (the quantum entropy “black box”)
3	Input step (2) and initialize qubits for initial states	The method normalizes Q_h, W values by amplitude encoding and allocates the total of $2(\log_2 p + 1)$ qubits, where p is the number of distinct elements in feature A
4	Input step (3) and returns the inner product of the states	Apply $U\left(\frac{\pi}{2}, 0, \pi\right)$ gate on both state vectors, calculate the inner product and returns $ \langle \psi_w \psi_q \rangle ^2$, which is the probability of overlap between the states

Table 2 provides the legend and description of the flow presented in Fig. 1

2.2 Correctness

Let D be a dataset consisting of m features and n records, denoted as $F = \{F_1, \dots, F_m\}$, and a target variable, denoted as Y_D . Given $A \in F$, let $D_E(A) = \{a_1, \dots, a_p\}$ be a set of distinct elements in feature A , for $p \leq n$. For each value $a_i \in D_E(A)$, the method calculates w_i . The proportional parameter of a_i equals $|Y_i|$. Let W be a set of all w_i . Clearly,

$$\sum_{w_i \in W} w_i = \sum_{i=1}^p |Y_i| = |D| = n$$

The algorithm transforms W to an amplitude encoding by concatenating all items into a single amplitude vector. Let $(1a)$ be the amplitude vector, such that $|\widetilde{W}|^2 = 1$. The normalization constant, denoted as \widetilde{W}_c , satisfies:

$$\widetilde{W}_c = \frac{1}{\sqrt{\sum_{i=1}^p w_i}}$$

Each $w_i \in W$ is converted to $\frac{\sqrt{w_i}}{\sqrt{\sum_{i=1}^p w_i}}$. Therefore, it satisfies the following:

$$\begin{aligned} \widetilde{W} &= \widetilde{W}_c \cdot \sum_{i=1}^p \sqrt{w_i} \\ |W|^2 &= \sum_{i=1}^p \left(\frac{\sqrt{w_i}}{\sqrt{\sum_{i=1}^p w_i}} \right)^2 = \sum_{i=1}^p \frac{w_i}{\sum_{i=1}^p w_i} = \frac{\sum_{i=1}^p w_i}{\sum_{i=1}^p w_i} = 1 \end{aligned} \tag{1b}$$

Note that the same correctness holds for (2a).

$$\widetilde{Q} = \widetilde{Q}_c \cdot \sum_{i=1}^p \sqrt{q_i} \tag{2b}$$

The Eq. (1a) is equivalent to (1b) as well as (2a) to (2b). The input vectors can be represented in the computational basis as $(1b)|i\rangle$ and $(2b)|i\rangle$. Since a quantum system of n qubits provide 2^n amplitudes, encoding (1b), (2b) requires the use of $\lceil \log_2 |W| \rceil + 1$ qubit each, i.e., $\lceil \log_2 p \rceil + 1$. It is important to note that in cases where the length of (1b) or (2b) is not to the power of two, zeros are added as their values do not change the IG calculation.

At this point, the method set the (1b), (2b) amplitude vectors as initial states, denoted as, $|\psi_w\rangle$ and $|\psi_q\rangle$, respectively. Since the normalization constant sums up to one, the coefficients

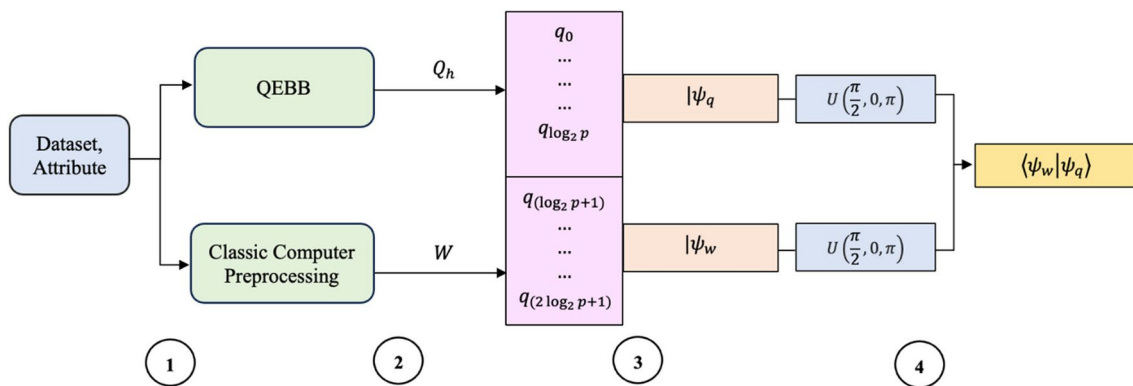


Fig. 1 The quantum procedure flow of conditional entropy calculation

of $|\psi_w\rangle$ can describe the probability of each state, which is equivalent to the proportional parameter. Similarly, the coefficients of $|\psi_q\rangle$ describe the relative entropy of each proportional parameter.

Next, the method creates two quantum circuits with $2(\lfloor \log_2 p \rfloor + 1)$ as the total number of qubits. By analyzing the worst case, all values in feature A are distinct, i.e., $p = n$, and the total number of qubits equals twice the number needed for the complete dataset. Once the method allocates all qubits, it applies the U gate with the parameters $\theta = \frac{\pi}{2}, \phi = 0, \lambda = \pi$ on both circuits independently, which moves the states into superposition. The following equations describe the quantum circuit over a single qubit, although the generalization to a higher dimension can be done with tensor products:

$$|\psi_w\rangle = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \psi_q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

$$H|\psi_w\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} w_1 + w_2 \\ w_1 - w_2 \end{bmatrix}$$

$$H|\psi_q\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} q_1 + q_2 \\ q_1 - q_2 \end{bmatrix}$$

Let $\langle \psi_w | \psi_q \rangle$ be the inner product of the states and let $|\langle \psi_w | \psi_q \rangle|^2$ be the probability of observing the system in the state ψ_w , given that it is in state ψ_q . The inner product can be understood as measuring the overlap between the state vectors ψ_w, ψ_q . Since the value of $|\langle \psi_w | \psi_q \rangle|^2$ is a probability, it satisfies:

$$0 \leq |\langle \psi_w | \psi_q \rangle|^2 \leq 1$$

Thus, the inner product of these states:

$$\begin{aligned} \langle \psi_w | \psi_q \rangle &= \psi_w^\dagger \psi_q = \\ &= \frac{1}{2}(w_1 + w_2)(q_1 + q_2) + \frac{1}{2}(w_1 - w_2)(q_1 - q_2) = \\ &= \frac{1}{2}(w_1q_1 + w_2q_2 + w_1q_1 + w_2q_2) = \\ &= w_1q_1 + w_2q_2 \end{aligned}$$

Last, subtracting $|\langle \psi_w | \psi_q \rangle|^2$ from the original entropy of the dataset, $QH(D)$, yields the IG achieved in dataset D by feature A .

3 Case study

This section demonstrates a quantum IG calculation compared to classical computer computation. Table 3 presents a mockup dataset (D) consisting of feature A and a Boolean target variable, Y . First, the classic computer computation will be described, followed by the quantum procedure.

3.1 Classical computer computation

The IG is defined as the difference between the dataset entropy, $H(Y)$, and the conditional entropy achieved by feature A :

$$IG_A(D) = H(Y) - H(Y|A) = H(Y) - \sum_{a_i \in A} P(A = a_i) \cdot H(Y|A = a_i)$$

The initial entropy of the dataset can be calculated by:

$$H(Y) = - \sum_{y_i \in Y} P(y_i) \cdot \log_2(p(y_i)) = -\left[\frac{1}{3} \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \log_2\left(\frac{2}{3}\right)\right] = 0.918$$

The calculated conditional entropy is based on the probability distribution function of feature A . For the demonstration, Table 4 presents the probability function, achieved by a simple preprocessing procedure.

The conditional entropy, $H(Y|A)$, is defined as:

$$\begin{aligned} H(Y|A = 1) &= \frac{1}{6} \cdot (-1 \cdot \log_2(1)) = 0 \\ H(Y|A = 2) &= \frac{1}{3} \cdot \left(-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right)\right) = 0.270 \\ H(Y|A = 3) &= \frac{1}{6} \cdot \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) = 0.167 \\ H(Y|A = 4) &= \frac{1}{3} \cdot (-1 \cdot \log_2(1)) = 0 \\ H(Y|A) &= \sum_{a_i \in A} P(A = a_i) \cdot H(Y|A = a_i) = 0.437 \end{aligned}$$

Last, the IG of dataset D , achieved by feature A , is:

$$IG_A(D) = H(Y) - H(Y|A) = 0.918 - 0.437 = 0.481$$

3.2 Quantum computer computation

This section presents the case study and experiments of the IG calculation using the proposed method. The experiment was simulated using Qiskit (Cross 2018) and an IBM simulator

Table 3 Mockup dataset for demonstration

A	1	1	2	2	2	2	3	3	4	4	4	4
Y	1	1	0	0	0	1	0	1	1	1	1	1

Table 4 The probability distribution function of feature A

a_i	1	2	3	4
$P[A = a_i]$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$

with 1024 shots. To begin, the method used the QEBB (Koren et al. 2023) to obtain the initial entropy of the target variable.

$$QH(Y) = 0.911$$

Next, the method created and initialized W , the proportional parameter of each distinct value in A , and Q_h , the vector of quantum entropy achieved by QEBB:

$$W = [6, 3, 6, 3]$$

$$Q_h = [0, 0.811, 1, 0]$$

The quantum circuit converted W, Q_h into amplitude vectors. Applying the U gate and pushing the vectors into the superposition yielded state vectors of:

$$|\psi_w\rangle = \frac{\sqrt{6}}{6}|00\rangle + \frac{\sqrt{3}}{3}|01\rangle + \frac{\sqrt{6}}{6}|10\rangle + \frac{\sqrt{3}}{3}|11\rangle$$

$$|\psi_q\rangle = 0|00\rangle + 0.669|01\rangle + 0.743|10\rangle + 0|11\rangle$$

The inner product of both quantum states was defined as:

$$\langle \psi_w | \psi_q \rangle = \psi_w^\dagger \psi_q = 0.689$$

$$\left| \langle \psi_w | \psi_q \rangle \right|^2 = 0.475$$

Last, the IG of dataset D achieved by feature A was:

$$IG_A(D) = 0.911 - 0.475 = 0.436$$

The difference between the result obtained in classical computing and the proposed method is 0.045. For further analysis, see Section 4.

4 Results

This section compares and analyzes the proposed and classical computer computation methods for calculating IG. For the comparison, the diabetes dataset was used (Kahn 1994), which is a dataset of 768 diabetic and non-diabetic women. It consists of eight features and a Boolean target variable. Since the proposed method was designed for discrete values, the ‘‘BMI’’ and ‘‘DiabetesPedigreeFunction’’ features were removed.

Table 5 presents the six features of the dataset and compares the IG obtained in both classical computer calculation and quantum computation. The initial entropy was 0.933 for the classic calculation and 0.932 for the quantum computation. The minimal error obtained in the dataset appears in the glucose feature, with an error of 0.001, while the highest error was accrued in the skin thickness feature, with an error of 0.075.

Building a decision tree using the ID3 algorithm includes searching for the feature with the maximum IG value to be the root of the tree. In both methods, the glucose feature was chosen as the tree’s root. Since the error of this feature was the minimal error achieved, it can be concluded that there was correspondence between the methods.

This study uses two inner quantum states to estimate the conditional entropy. Thus, Fig. 2 compares the conditional entropy achieved by the quantum and classical methods. High

Table 5 A comparison of classic and quantum computation of information gain

Attribute (A)	Classic Computation		Quantum Computation		Error
	$H(A Y)$	$IG_A(D)$	$H(A Y)$	$IG_A(D)$	
Pregnancies	0.871	0.061	0.838	0.094	0.033
Glucose	0.304	0.629	0.302	0.630	0.001
Blood Pressure	0.873	0.059	0.808	0.124	0.065
Skin Thickness	0.851	0.081	0.776	0.156	0.075
Insulin	0.656	0.277	0.711	0.221	0.056
Age	0.792	0.140	0.740	0.192	0.052

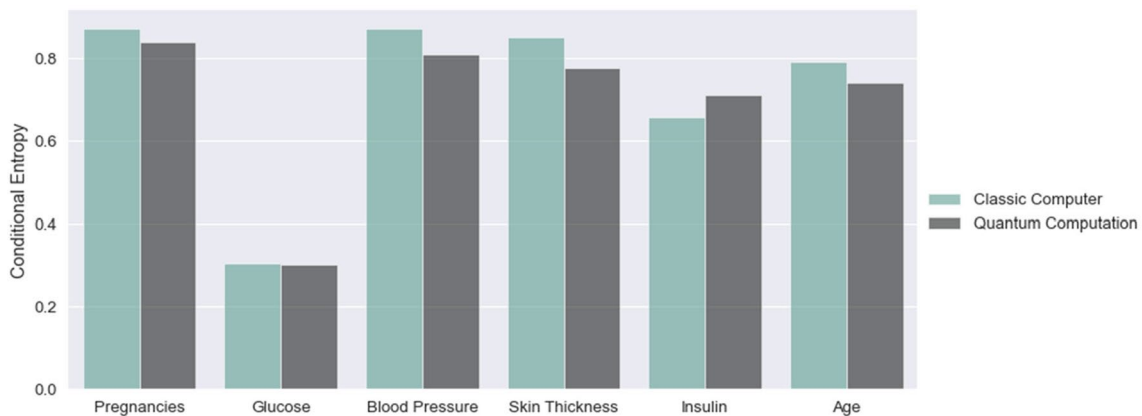


Fig. 2 A comparison of the conditional entropy of each feature

Table 6 Ranks of information gain and its differences for the Spearman correlation coefficient

Attribute (A)	Classic Ranking		Quantum Ranking		$d_i = R_c - R_q$
	$IG_A(D)$	R_c	$IG_A(D)$	R_q	
Pregnancies	0.061	5	0.094	6	1
Glucose	0.629	1	0.630	1	0
Blood Pressure	0.059	6	0.124	5	1
Skin Thickness	0.081	4	0.156	4	0
Insulin	0.277	2	0.221	2	0
Age	0.140	3	0.192	3	0

agreement was obtained among the features when, in most cases (five out of six), the quantum result scored slightly lower than the actual value. This conclusion also helps refine and understand the relationship between the conditional entropy and the inner product of quantum states.

The IG measure can be interpreted as ranking the features in the dataset according to the level of mutual information with the target feature. Thus, for further analysis, the Spearman correlation coefficient, denoted as r_s , was used to examine the ranking correlation between the classic and quantum computation results (Myers and Sirosis 2004).

Table 6 presents the ranks and their differences between the classical and quantum computation methods. Since all ranks were distinct integers, the r_s was computed as follows:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot (1^2 + 1^2)}{6 \cdot (6^2 - 1)} = 0.942$$

Given that $r_s = 0.942$ and $p < 0.005$, this indicated that there was a very strong positive correlation between the ranks with a probability of 0.995 (Ramsey 1989).

5 Conclusions and discussion

This study proposes a quantum procedure for information gain calculation. The presented procedure can be applied in data mining, information analysis, and machine learning algorithms. The proposed method involves amplitude encoding and uses the inner product of the quantum states to estimate the conditional entropy. Its main innovation is the use of quantum computers to calculate the IG without having to transform the problem from classical to quantum computation. Furthermore, it is accessible to those without a previous understanding of QC. The following are the current study’s main conclusions:

1. The procedure is based on the inner product of the quantum states. The squared value of the inner product is the

probability of observing the system in one state given the other state. By using amplitude encoding for the input vectors, the probability represents the conditional entropy of the target variable given a feature.

2. The minimum error achieved between the value found using classical calculation and the proposed quantum procedure was 0.001, while the maximum error was 0.075. It can be concluded that, in the case of a Boolean target variable, the conditional entropy can be estimated by the inner product of quantum states.
3. To compare the rating of the features according to the IG, the Spearman correlation coefficient (r_s) was calculated for the ratings obtained by the classical and quantum calculations. The correlation coefficient value was 0.942 with a p -value < 0.005 , indicating a strong level of agreement between the ranks with a probability of 0.995.

This study's limitation relates to the use of the inner product as the conditional entropy. Due to QC constraints, the conditional entropy was bounded between zero and one, which holds only for Binary target features. Future work should examine two main issues. First, the generalization of the proposed method supports multiclass classification (i.e., a target value with at least three distinct values) and continuous features. Second, the evaluation and analysis of additional datasets consisting of mixed types of features should be further studied.

Author contributions Michal Koren: Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation, Project administration. Or Peretz: Writing - Review & Editing ,Software, Data Curation, Visualization.

Funding Open access funding provided by Shenkar College of Engineering and Design. This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The data used in this study is available and accessible in UCI Machine Learning Repository – Datasets (<https://archive.ics.uci.edu/ml/datasets.php>).

Declarations

Competing interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmed F, Kim KY (2017) Data-driven weld nugget width prediction with decision tree algorithm. *Procedia Manuf* 10:1009–1019. <https://doi.org/10.1016/j.promfg.2017.07.092>
- Alchieri L, Badalotti D, Bonardi P, Bianco S (2021) An introduction to quantum machine learning: from quantum logic to quantum deep learning. *Quantum Mach Intell* 3:28. <https://doi.org/10.1007/s42484-021-00056-8>
- Assouel A, Jacquier A, Kondratyev A (2022) A quantum generative adversarial network for distributions. *Quantum Mach Intell* 4:28. <https://doi.org/10.1007/s42484-022-00083-z>
- Azevedo V, Silva C, Dutra I (2022) Quantum transfer learning for breast cancer detection. *Quantum Mach Intell* 4:5. <https://doi.org/10.1007/s42484-022-00062-4>
- Batra M, Agrawal R (2018) Comparative analysis of decision tree algorithms. In: *Nature Inspired Computing: Proceedings of CSI 2015*. Springer, Singapore, pp 31–36
- Bein B (2006) Entropy. *Best Practice & Research. Clinical Anaesthesiology* 20:101–109. <https://doi.org/10.1016/j.bpa.2005.07.009>
- Benedetti M, Lloyd E, Sack S, Fiorentini M (2019) Parameterized quantum circuits as machine learning models. *Quantum Sci Technol* 4:043001. <https://doi.org/10.1088/2058-9565/ab4eb5>
- Bennett CH, Bernstein E, Brassard G, Vazirani U (1997) Strengths and weaknesses of quantum computing. *SIAM Journal on Computing* 26:1510–1523. <https://doi.org/10.1137/S0097539796300933>
- Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. *Nature* 549:195–202. <https://doi.org/10.1038/nature23474>
- Boyer M, Brassard G, Høyer P, Tapp A (1998) Tight bounds on quantum searching. *Fortschritte Der Phys* 46:493–505. [https://doi.org/10.1002/\(SICI\)1521-3978\(199806\)46:4/5<493::AID-PROP493>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1521-3978(199806)46:4/5<493::AID-PROP493>3.0.CO;2-P)
- Buffoni L, Caruso F (2021) New trends in quantum machine learning (a). *Europhysics Letters* 132:60004. <https://doi.org/10.1209/0295-5075/132/60004>
- Charbuty B, Abdulazeez A (2021) Classification based on decision tree algorithm for machine learning. *J Appl Sci Technol Trends* 2:20–28. <https://doi.org/10.38094/jastt20165>
- Cross A (2018) The IBM Q experience and QISKit open-source quantum computing software. *APS March Meet Abstr* 2018(L58):003
- Dalla Pozza N, Buffoni L, Martina S, Caruso F (2022) Quantum reinforcement learning: The maze problem. *Quantum Mach Intell* 4:11. <https://doi.org/10.1007/s42484-022-00068-y>
- De Wolf R (2019) Quantum computing: Lecture notes. arXiv:1907.09415. <https://doi.org/10.48550/arXiv.1907.09415>
- González FA, Gallego A, Toledo-Cortés S, Vargas-Calderón V (2022) Learning with density matrices and random features. *Quantum Mach Intell* 4:23. <https://doi.org/10.1007/s42484-022-00079-9>
- Guleria P, Thakur N, Sood M (2014) Predicting student performance using decision tree classifiers and information gain. In: *2014 International conference on parallel, distributed and grid computing* (pp. 126–129). IEEE, Solan, India, pp 126–129. <https://doi.org/10.1109/PDGC.2014.7030728>
- Hssina B, Merbouha A, Ezzikouri H, Erritali M (2014) A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications* 4(2):13–19
- Hur T, Kim L, Park DK (2022) Quantum convolutional neural network for classical data classification. *Quantum Mach Intell* 4:3. <https://doi.org/10.1007/s42484-021-00061-x>
- Jin C, De-Lin L, Fen-Xiang M (2009) An improved ID3 decision tree algorithm. In: *2009 4th international conference on computer science & education*. IEEE, Nanning, China, pp 127–130. <https://doi.org/10.1109/ICCSE.2009.5228509>
- Kahn M (1994) Diabetes. UCI Machine Learning Repository. <https://doi.org/10.24432/C5T59G>
- Kapur JN, Kesavan HK (1992) Entropy optimization principles and their applications. In: Singh VP, Fiorentino M (eds) *Entropy and energy dissipation in water resources*. Springer, Dordrecht, pp 3–20
- Kaufmann K, Vecchio KS (2020) Searching for high entropy alloys: A machine learning approach. *Acta Materialia* 198:178–222. <https://doi.org/10.1016/j.actamat.2020.07.065>
- Kaufmann K, Maryanovsky D, Mellor WM, Zhu C, Rosengarten AS, Harrington TJ, Oses C, Toher C, Curtarolo S, Vecchio KS (2020) Discovery of high-entropy ceramics via machine learning. *NPJ Computational Materials* 6:42. <https://doi.org/10.1038/s41524-020-0317-6>
- Kent JT (1983) Information gain and a general measure of correlation. *Biometrika* 70(1):163–173. <https://doi.org/10.1093/biomet/70.1.163>
- Koren M, Koren O, Peretz O (2023) A quantum “black box” for entropy calculation. *Quantum Mach Intell* 5(2):37. <https://doi.org/10.1007/s42484-023-00127-y>
- Liu X, Zhang J, Pei Z (2022) Machine learning for high-entropy alloys: Progress, challenges and opportunities. *Progress in Materials Science* 131:101018. <https://doi.org/10.1016/j.pmatsci.2022.101018>
- Myers L, Sirois MJ (2004) Spearman correlation coefficients, differences between. *Encycl Stat Sci* 12. <https://doi.org/10.1002/0471667196.ess5050.pub2>
- Navada A, Ansari AN, Patil S, Sonkamble BA (2011) Overview of use of decision tree algorithms in machine learning. In: *2011 IEEE control and system graduate research colloquium*. IEEE, Shah Alam, Malaysia, pp. 37–42. <https://doi.org/10.1109/ICSGRC.2011.5991826>
- Pirhooshyaran M, Terlaky T (2021) Quantum circuit design search. *Quantum Mach Intell* 3:25. <https://doi.org/10.1007/s42484-021-00051-z>
- Ramsey PH (1989) Critical values for Spearman’s rank order correlation. *Journal of Educational Statistics* 14(3):245–253. <https://doi.org/10.3102/10769986014003245>
- Robertson JK (1943) The role of physical optics in research. *American Journal of Physics* 11:264–271. <https://doi.org/10.1119/1.1990496>
- Tüysüz C, Rieger C, Novotny K, Demirköz B, Dobos D, Potamianos K, Vallecorsa S, Vilmant JR, Forster R (2021) Hybrid quantum classical graph neural networks for particle track reconstruction. *Quantum Mach Intell* 3:29. <https://doi.org/10.1007/s42484-021-00055-9>
- Wehrl A (1978) General properties of entropy. *Reviews of Modern Physics* 50(2):221–260. <https://doi.org/10.1103/RevModPhys.50.221>
- Wiebe N (2020) Key questions for the quantum machine learner to ask themselves. *New Journal of Physics* 22:091001. <https://doi.org/10.1088/1367-2630/abac39>
- Ying M (2010) Quantum computation, quantum theory and AI. *Artificial Intelligence* 174:162–176. <https://doi.org/10.1016/j.artint.2009.11.009>
- Zeng W, Johnson B, Smith R, Rubin N, Reagor M, Ryan C, Rigetti C (2017) First quantum computers need smart software. *Nature* 549:149–151. <https://doi.org/10.1038/549149a>
- Zoufal C, Lucchi A, Woerner S (2021) Variational quantum Boltzmann machines. *Quantum Mach Intell* 3:7. <https://doi.org/10.1007/s42484-020-00033-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.