Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

Optimizing Quantum Neural Networks for Predicting the Effectiveness of Drug Compounds as Corrosion Inhibitors

Lubna Mawaddah[1], Muhammad Reesa Rosyid[1], Akbar Priyo Santosa[1], Muhamad Akrom[2,*]

[1] Faculty of Computer Science, Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia
[2] Research Center for Materials Informatics, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202013132@mhs.dinus.ac.id, [2]111202012665@mhs.dinus.ac.id, [3]111202012656@mhs.dinus.ac.id, [4,*]m.akrom@dsn.dinus.ac.id
Correspondence Author Email: m.akrom@dsn.dinus.ac.id
Submitted: **11/06/2024**; Accepted: **25/06/2024**; Published**: 26/06/2024**

Abstract−Corrosion, caused by electrochemical reactions in corrosive environments, can degrade the quality and lifespan of materials, potentially leading to significant losses in various industrial sectors. One common strategy to reduce corrosion rates is by using corrosion inhibitors. A significant challenge in this field is the time-consuming and costly process of testing new corrosion inhibitors in the laboratory. Consequently, there is a need for more efficient and cost-effective methods to predict the effectiveness of potential corrosion inhibitors using machine learning techniques. This research addresses this problem by applying a quantum machine learning (QML) approach with quantum neural network (QNN) algorithms to evaluate the effectiveness of drug compounds as corrosion inhibitors. The study aims to optimize QNN models by investigating three different quantum circuit configurations to identify the most effective design. The results showed that Model-01, consisting of three layers, demonstrated the best performance with an MSE of 38.81, an RMSE of 6.23, and an MAE of 6.19, along with the shortest training time of 32 seconds, indicating an optimal balance between complexity and generalizability. Overall, this QML approach provides new insights into the predictive ability of QNN models in assessing the effectiveness of drug compounds as corrosion inhibitors, demonstrating the potential of quantum computing to enhance predictive accuracy and efficiency in investigating anti-corrosion materials.

Keywords: Corrosion Inhibitor; Drug Compounds; Quantum Circuit; Quantum Machine Learning; Quantum Neural Network

1. INTRODUCTION

Corrosion is the process of decreasing the quality and service life of materials, especially metals, caused by electrochemical reactions in corrosive environments. This phenomenon is of significant concern across various industrial sectors, particularly in the oil and gas industry, due to its potential for causing substantial losses [1]. The economic impact of corrosion amounts to US$ 2.5 trillion annually worldwide, which has been steadily increasing over the past decade [2]. The impact of corrosion is not only limited to the economy, but also affects environmental damage, security, and safety.

One effective strategy to mitigate corrosion is the use of anti-corrosion materials, known as corrosion inhibitors [3]. These are chemicals added to environments that can damage metals (aggressive environments) to slow down the process of metal deterioration. The most effective organic compounds as corrosion inhibitors are those containing specific atoms such as nitrogen (N), phosphorus (P), sulfur (S), arsenic (As), and oxygen (O), or specific bonds (π) in their structures [4]. In recent years, researchers have been developing research on corrosion inhibitors using synthetic compounds to create organic corrosion inhibitors with excellent corrosion inhibition properties [5]. However, the primary challenge lies in the time-consuming and costly process of synthesizing and testing new corrosion inhibitors in the laboratory. Therefore, there is a critical need for efficient and cost-effective methods to predict the effectiveness of potential corrosion inhibitors by utilizing machine learning capabilities.

Studies on the application of quantitative structure activity relationship (QSAR) and classical machine learning (CML) to predict the effectiveness of corrosion inhibitors have been extensively discussed in the literature, as these methods replace costly and resource-intensive laboratory experiments [6]-[9]. For instance, Perez et al. [10] developed a QSAR-ARX model to predict the corrosion inhibition efficiency of expired drugs on steel surfaces using quantum mechanical descriptors, achieving a root mean square error (RMSE) of 7.03. Ser et al. [11] employed artificial neural network (ANN) models to predict the corrosion inhibition efficiency of pyridine and quinoline compounds, achieving an RMSE of 8.8. Alamri et al. [12] investigated pyrimidine compounds as corrosion inhibitors using partial least squares regression (PLS) and random forest (RF), resulting in mean squared error (MSE) values of 64.641 and 32.602, respectively. Similarly, focusing on pyrimidine compounds, Quadri et al. [13] employed an ANN model and achieving an RMSE of 10.56. Meanwhile, Pham et al. [14] used the gradient boosting decision tree (GB) algorithm along with the permutation feature importance (PFI) technique to predict the corrosion inhibition efficiency of organic compounds on carbon steel, obtaining an RMSE of 6.40.

While these CML methods have provided valuable insights, they are limited by computational constraints, as classical computers require more time to solve problems, especially when the dataset is large. Therefore, a quantum machine learning (QML) approach is necessary. By leveraging the power of quantum computing, QML is expected to enhance the accuracy and efficiency in predicting corrosion inhibition efficiency. Quantum computing (QC) has reached a historic milestone with its ability to solve certain problems faster than classical computing. This achievement is attracting growing interest and supporting future research [15], [16]. QML is a branch of quantum information processing that focuses on developing quantum algorithms for learning from data, leveraging the processing power of quantum computers, as well as the scalability and capacity of machine learning algorithms [17].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

Variational quantum algorithm (VQA) is a quantum computing algorithm that uses a hybrid quantum-classical scheme. In VQA, quantum processors are used to prepare the target quantum state, while measurements are performed to extract useful information for the classical computer in the exploration and optimization process. VQA has been widely used in quantum optimization problem solving, quantum simulation, and quantum machine learning [15], [18]. Quantum neural network (QNN) is a part of VQA that operates by designing quantum circuits consisting of qubits and parameterized rotation gate operations. Qubits can be likened to neurons in classical neural networks, while rotation gates can be likened to weights connecting neurons. Thus, QNN combines quantum theory features with neural network properties [19]-[21].

This study aims to investigate the ability of the QNN algorithm to estimating the corrosion inhibition efficiency of drug compounds. This approach explores several important aspects, including the model's ability to learn complex patterns in the training data. Model evaluation is performed using the metrics mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Training time and expressibility is also measured to assess the prediction quality of the developed QNN model. By exploring the potential of QML, this research aims to offer a novel approach that can significantly enhance the prediction and optimization of corrosion inhibitors, ultimately contributing to the reduction of the economic and environmental impacts of corrosion.

2. RESEARCH METHODOLOGY

2.1 Research Stages

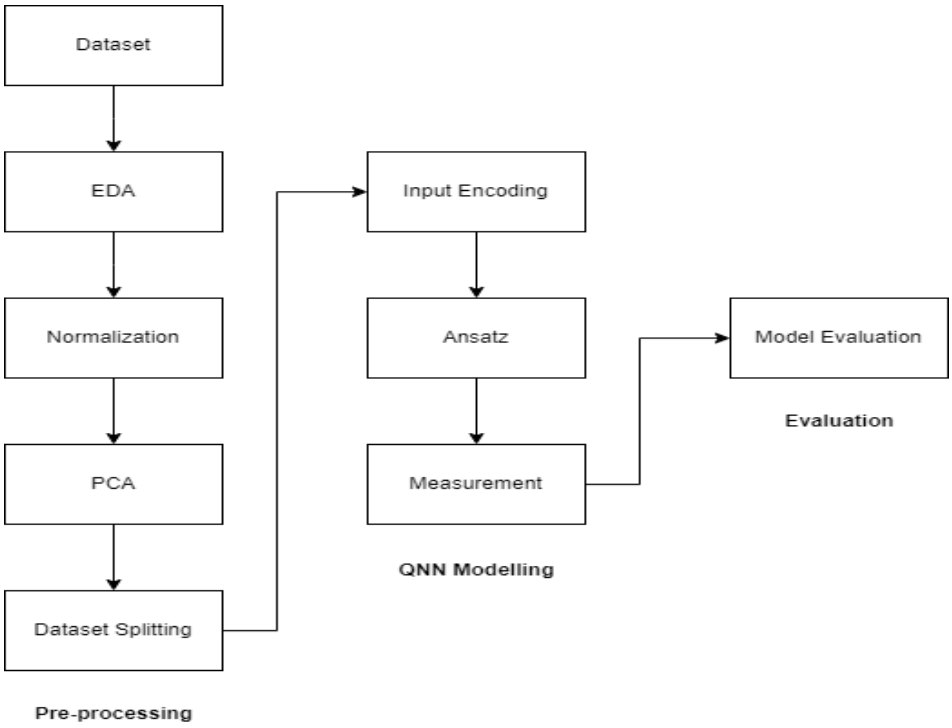

Figure 1. QML model framework

The research stages illustrated in Figure 1 depict the process undertaken in this study, starting with data collection, exploratory data analysis (EDA), normalization, principal component analysis (PCA), and data splitting. The next step involves building the QNN model with an architecture consisting of input encoding, the ansatz, and measurement. The final step is model evaluation. Detailed explanations can be found in sections 2.2-2.5.

2.2 Dataset Description

This study uses a published dataset containing 260 drug compounds with 14 features and 1 target [10]. The features used are quantum chemical descriptors representing molecular inhibitor properties, including molecular weight (MW), acid dissociation constant (pKa), octanol-water partition coefficient (log P), water solubility (log S), polar surface area (PSA), polarizability (α), the energy of highest occupied molecular orbital (E-HOMO), the energy of lowest unoccupied molecular orbital (E-LUMO), ionization energy (I), electron affinity (A), electronegativity (χ), hardness (η), electrophilicity (ω), and the fraction of shared electrons (ΔN). The target variable is corrosion inhibition efficiency (IE (%)), which is the corrosion inhibition performance of the inhibitor.

Each drug compound has a size parameter that depends on its molecular weight. The acidity strength in a solution is measured by pKa = -logKa, which is calculated by the negative base-10 logarithm of the acid dissociation constant of a solution. Log P indicates the ability of organic compounds to dissolve in polar (water) and non-polar

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

(octanol) phases. Additionally, the log S and PSA descriptors are two important parameters used to predict the water solubility of a substance. Log S provides direct information about water solubility, while PSA provides information about molecular polarity that can affect water solubility [10], [22]. Molecular polarizability is affected by the electron density distribution and the ability of the molecule to change the electron density distortion [1].

E-HOMO refers to the molecule's ability to transfer electrons, where molecules with high E-HOMO values tend to transfer electrons to acceptors with low-energy and empty molecular orbitals. While E-LUMO refers to the molecule's ability to accept electrons [23]. Ionization energy is the amount of energy required to remove an outermost electron from an atom, used to measure the reactivity level of atoms or molecules. Meanwhile, electron affinity refers to the amount of energy released to capture one mole of electrons [1]. Electronegativity measures the relative ability of an atom to attract electrons in a compound. Pyrimidines with more electronegative centers are predicted to exhibit better inhibition efficiency. Global hardness measures a molecule's resistance to charge transfer [11], [12]. Electrophilicity index measures the energy loss due to maximum electron flow between donors and acceptors [24]. Lastly, the fraction of shared electrons (ΔN) indicates the number of electrons transferred from the metal surface to the inhibitor molecule. A higher ΔN value indicates higher corrosion inhibition efficiency [25].

2.3 Data Pre-processing

Exploratory data analysis (EDA) is a crucial phase in data analysis that focuses on examining the dataset, understanding the nature and characteristics of the data, and generate an initial understanding of the data [26]. From the EDA results, it was found that most of the data had missing values. Therefore, data cleaning was performed by removing rows containing missing values, resulting in 78 data points remaining out of the initial 260 data points.

Next, data normalization was performed using the min-max scaling method. This method adjusts each feature in the dataset so that its values are within a specified range, typically between 0 and 1 or -1 and 1 [27]. The equation can be seen in equation (1).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x represents the original value of the feature, x_{min} is the minimum value of the feature, x_{max} is the maximum value of the feature, and x_{new} is the new normalized value of the feature [28].

Subsequently, the principal component analysis (PCA) technique is applied to the features in the dataset. PCA is a technique used to reduce the dimensionality of a dataset, while retaining as much relevant statistical information as possible. The PCA process involves creating new variables that are linear combinations of the original variables, thereby maximizing the variance of uncorrelated data [29]. PCA is implemented by retaining 95% of the total components, resulting in 4 most significant features out of the initial 14 features.

After the data has gone through the PCA process, the next step is to split the dataset into two parts; the training data and the testing data, with an 80:20 ratio. This division of data is important to prevent overfitting, which is a condition where the model fits the training data too well and cannot generalize well to new data [30].

2.3 Quantum Neural Network Model

The QNN architecture consists of three main components, as depicted in Figure 1. The first component is the input encoding or quantum feature mapping, which is responsible for transforming classical data into quantum representations. Each classical input vector is converted into a quantum state vector by applying an encoding transformation to the $|0\rangle$ ground state on each qubit. The second component is the ansatz or variational circuit, which is a quantum circuit consisting of a series of gates that can be iteratively adjusted using classical optimization. Its function is to approach a desired state in quantum problem solving. Variations in ansatz design can result in significant performance differences. In QNN, designing an efficient ansatz with high accuracy is still a challenge. The last component is measurement, which measures the quantum circuit and produces output values that have been converted from quantum states [31]-[35].

The objective function is constructed to compute the loss value based on the difference between the actual target values with the model's predicted values. The function is designed to assess how accurately the predictive model matches the target values for each sample. The next step is to minimize the loss value by applying an optimizer. Quantum parameters are optimized using the constrained optimization by linear approximations (COBYLA) optimizer, which is an optimization technique used to find the minimum value of the objective function without requiring derivatives of the function. This technique allows constraints to be placed on the optimized variables, enabling the solution of constrained optimization problems [36].

After obtaining the optimal parameters, the quantum circuit is applied to the training and testing data. Then, the performance of the model is evaluated using evaluation metrics such as MSE, RMSE, and MAE, along with the training time required in the model training process. This evaluation is conducted to assess the predictive quality of the developed QNN model.

2.3 Metrics Evaluation

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

Model evaluation is conducted to compare the performance of each circuit. This process aims to assess the quality and effectiveness of each model in predicting the corrosion inhibitor values in the drug compounds. Model performance evaluation is measured using evaluation metrics such as MSE, RMSE, and MAE.

MSE and MAE both assess the disparity between actual values and predicted values. The distinction lies in their calculation methods; MAE calculates the absolute difference, whereas MSE calculates the squared difference. Meanwhile, RMSE is the square root of MSE. RMSE standardizes the unit of measure of MSE [37], [38]. The formulas for each evaluation metric can be seen in equations 2-4.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (3)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \quad (4)$$

Where X_i is the actual value, Y_i is the predicted value, $\bar{Y}$ is the mean of the actual values, and m is the number of samples.

3. RESULT AND DISCUSSION

The primary components that PCA produces effectively function as novel features by encapsulating a significant amount of the relevant information about the connection between dataset characteristics and the target variable. As such, it is interesting to observe that important information about the feature-target link may persist after PCA integration and in many cases, it can even improve. This preservation encourages more effective analysis, especially when it comes to the interactions between characteristics and targets in scenarios involving regression analysis.

Table 1. Significance values of features between before and after PCA analysis

Feature	MW	pKa	log P	log S	PSA	α	HOMO	LUMO	I	A	χ	η	σ	Ω	ΔN
Significance (0.05)	0.03	0.04	0.02	0.02	0.03	0.04	0.02	0.02	0.04	0.03	0.03	0.03	0.03	0.04	0.03
Feature	PCA 1				PCA 2			PCA 3				PCA 4			
Significance (0.05)	0.03				0.04			0.02				0.03			

Table 1 shows the relationship between features and targets both before and after PCA integration. The features include chemical descriptors. The significance values (p-values) below 0.05 indicate statistically significant relationships between these features and the target variable. Post-PCA, the primary components (PCA 1, PCA 2, PCA 3, PCA 4) also show significant values, indicating that PCA has effectively retained the important information from the original dataset. The Pearson correlation coefficient and corresponding p-values are used to describe the results of the EDA. It is crucial to understand that even if PCA reduces the dimensionality of the dataset, the inherent information contained in the features is mostly unchanged. PCA is specifically designed to reduce the size of datasets while preserving the highest level of variability included in the original data.

After PCA is used, significant changes are observed in the relationship between the principal components and the objective. A marked increase in the correlation coefficient suggests a stronger connection between the principal components and the objective. Furthermore, there is a decrease in the p-value, which suggests increased statistical significance, when the connection between the principal component (PC) and the target post-PCA transformation to its pre-transformation condition is compared. As a result, PCA is helpful in maintaining important data from the original characteristics. Moreover, it reveals stronger correlations with higher statistical significance between the target variable and the attributes.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

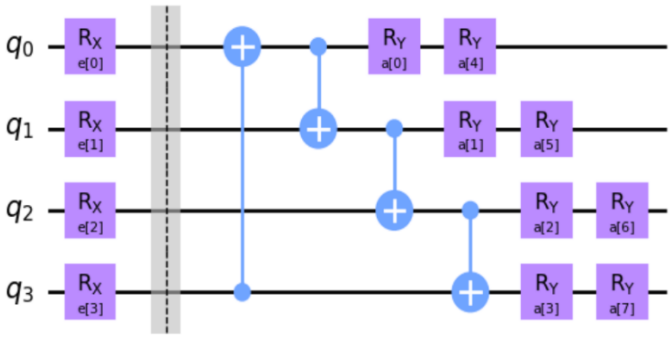

Figure 2. Quantum circuit 1 (Model-01)

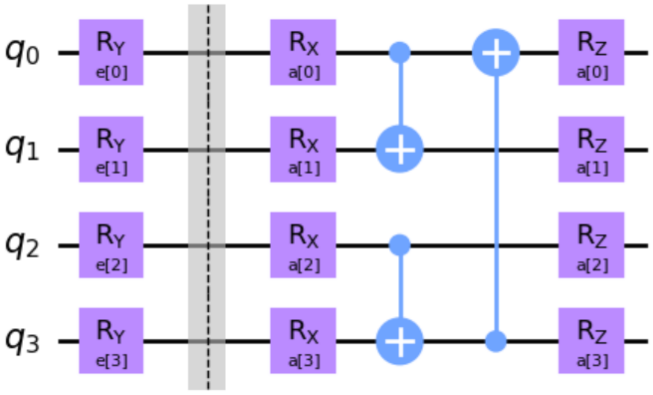

Figure 3. Quantum circuit 2 (Model-02)

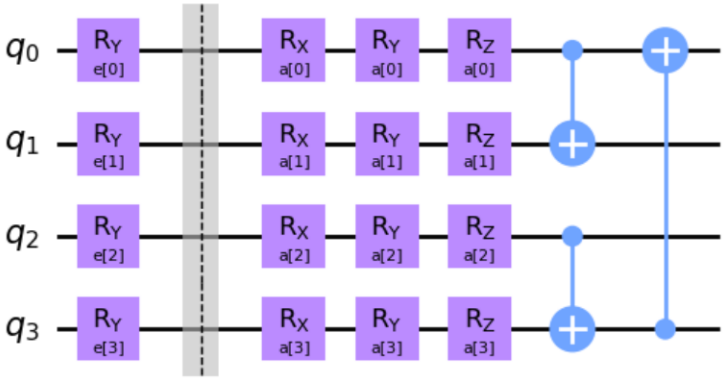

Figure 4. Quantum circuit 3 (Model-03)

Figures 2-4 depict the proposed QNN architecture in this study. The quantum circuits are constructed using 4 qubits (corresponding to the number of features retained after the PCA process). In this study, 3 quantum circuits are built to compare their effectiveness. The quantum circuits were designed with design variations to evaluate their effect on the model performance.

In this experiment, several quantum gates were used. The rotation Rx and Ry gates play roles in changing the qubit values, which creates a quantum superposition, transforming the qubit's ground state from $|0\rangle$ to a balanced superposition state of $|0\rangle$ and $|1\rangle$. There are Rx, Ry, and Rz gates combination, which are rotation gates allowing qubit rotation around the x, y, and z axes respectively, with adjustable parameters. To create entanglement between two qubits, the controlled-not (CX) gate is used, where one of the qubit acts as a control and the other as a target.

1. **Model-01:**
   The circuit in Figure 2 is built using Rx gates on each qubit as the input encoding, followed by variations of CX and Ry gates as the ansatz.
2. **Model-02:**
   The circuit in Figure 3 is built using Ry gates on each qubit as the input encoding, while the ansatz consists of Rx, CX, and Rz gates are repeatedly applied.
3. **Model-03:**
   The circuit in Figure 4 is built using Ry gates on each qubit as the input encoding, while the ansatz consists of Rx, Ry, Rz, and CX gates sequentially on each qubit.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

The number of layers in QNN models is a crucial hyperparameter that significantly influences the model's performance. Through systematic experimentation, it was determined that different models achieved optimal performance with varying numbers of layers: Model-01 with 3 layers, Model-02 with 5 layers, and Model-03 with 1 layer, as indicated by the highest validation R^2 scores. The primary challenge in determining the optimal number of layers is balancing overfitting and underfitting. Overfitting occurs when the model has too many layers, making it overly complex. The model captures noise and minute details in the training data, which do not generalize well to new, unseen data. Overfitted models exhibit high performance on training data but poor performance on validation or test data. This is evident from the need to avoid excess layers in model design. Underfitting occurs when the model has too few layers, making it too simplistic. The model fails to capture the underlying patterns and structures in the data. Underfitted models perform poorly on both training and validation data because they cannot effectively learn from the data. This was avoided in the study by ensuring even the simplest model (Model-03) had a sufficient, albeit minimal, number of layers.

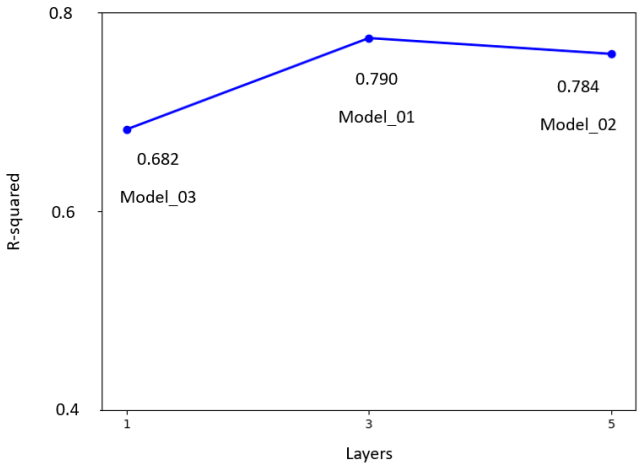

Figure 5. Best performance of QNN models based-on layers number

Figure 5 shows the performance of QNN models based-on layers number. Model-01 (3 layers) achieved the best validation R^2 score with three layers, indicating an optimal balance between complexity and generalizability for its specific dataset and task. Three layers allowed the model to learn adequately complex patterns without overfitting. Model-02 (5 layers) required more layers to reach optimal performance, suggesting that its data or the complexity of the task necessitated a deeper network to capture more intricate relationships within the data. Model-03 (1 layer) performed best with just one layer, which might indicate that the underlying patterns in this dataset are relatively simple and can be captured without additional complexity.

The systematic trials involved gradually adjusting the number of layers and observing the impact on validation R^2 scores. This method ensured a thorough exploration of the model's capacity to generalize, helped identify the point at which additional layers ceased to contribute to, or even detracted from, model performance, and provided a clear understanding of how different tasks and datasets might require different network depths. The results underscore the necessity of carefully selecting the number of layers in QNN models. Choosing the optimal number of layers is critical for the effectiveness of QNN models. The study highlights that a methodical approach to layer selection, based on empirical validation, is essential for achieving a balance between model complexity and the ability to generalize. This careful tuning results in models that can make reliable predictions, which is the ultimate goal in practical applications.

Table 2. Experiment results

Model	MSE	RMSE	MAE	Training Time
Model-01	38.81	6.23	6.19	32 second
Model-02	66.42	8.15	8.42	49 second
Model-03	99.40	9.97	9.43	58 second

Table 2 shows the performance of the three QNN models based on evaluation metrics (MSE, RMSE, and MAE) and training time. Model-01 outperforms the other two models across all evaluation metrics, achieving the lowest MSE, RMSE, and MAE. This indicates that it has the best predictive accuracy and generalization capability. It also has the shortest training time (32 seconds), suggesting that it is not only the most accurate but also the most efficient in terms of computational resources. The balance of three layers seems to provide the right level of complexity to capture the data patterns without overfitting or underfitting.

Model-02 shows intermediate performance with higher MSE, RMSE, and MAE compared to Model-01, but better than Model-03. This suggests that increasing the layers beyond three may introduce some degree of overfitting,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

thus reducing its effectiveness. The training time is significantly longer (49 seconds), indicating that the additional layers require more computational effort and time without a corresponding improvement in performance.

Model-03 has the highest MSE, RMSE, and MAE, indicating it struggles to capture the necessary patterns in the data, likely due to underfitting. Surprisingly, it also has the longest training time (58 seconds). This could be due to inefficiencies in the training process for simpler models or possibly the need for more iterations to compensate for the lack of complexity.

The results clearly demonstrate the trade-offs involved in model design concerning the number of layers. Model-01, with three layers, achieves the best balance of predictive performance and training efficiency. Model-02, while more complex, does not provide sufficient gains in accuracy to justify the additional computational cost and training time. This suggests that the model begins to overfit, capturing noise rather than meaningful patterns. Model-03, with its single layer, underfits the data, unable to learn the underlying relationships effectively, leading to poor performance on all metrics despite the longest training time. In practical applications, selecting the appropriate number of layers is crucial. Too many layers can lead to diminishing returns in accuracy and increased computational demands, while too few layers result in inadequate learning. Model-01's performance indicates that for this specific dataset and task, three layers are sufficient to capture the necessary complexity without incurring the drawbacks of overfitting or excessive computational cost.

The evaluation of the three QNN models underscores the importance of a balanced approach to model complexity. Model-01, with three layers, strikes the best balance, delivering superior predictive performance and training efficiency. This highlights the necessity of carefully selecting the number of layers to ensure models are both effective and computationally feasible.

Table 3. Expressibility values of models

Model	Expressibility
Model-01	0.15
Model-02	0.29
Model-03	0.38

Table 3 shows the expressibility values of each models. Expressibility in the context of QNN models is a measure of the model's ability to represent a wide variety of functions or data patterns. High Expressibility; the model can represent complex functions and capture intricate patterns in the data. Moderate Expressibility; the model captures essential patterns without overfitting. Achieved through a balanced network depth. Low Expressibility; the model captures only basic patterns, often leading to underfitting. Associated with very shallow networks. Model-01 is the best overall performance (lowest MSE, RMSE, MAE) and efficient training time, with expressibility value of 0.15 capable of capturing necessary data patterns without overfitting or underfitting. Model-02 with higher error metrics than Model-01, indicating potential overfitting, has expressibility value of 0.29, high expressibility due to greater depth, but suffers from overfitting. Model-03 with highest error metrics, indicating underfitting, reach expressibility value for 0.38, the lowest expressibility due to insufficient depth to capture complex patterns. Model-01 demonstrates an optimal balance between capturing complex patterns and maintaining generalizability. The three-layer structure provides adequate expressibility to learn important data features without overfitting, as evidenced by the lowest error metrics. Model-02 has a higher expressibility due to its additional layers, allowing it to capture more intricate patterns. However, this comes at the cost of increased risk of overfitting, reflected in the higher error metrics. Despite its theoretical higher expressibility, practical performance indicates a trade-off with generalizability. Model-03 has the lowest expressibility, constrained by its single-layer architecture. It struggles to capture sufficient patterns from the data, leading to underfitting. This is reflected in its high error metrics, despite being theoretically claimed to have the best expressibility in terms of avoiding overfitting. Expressibility is crucial for model performance, but it must be balanced with the model's ability to generalize. Model-01 achieves the best practical expressibility, reflected in its balanced performance metrics. Model-02, while having higher expressibility, overfits, reducing its effectiveness. Model-03, despite claims of best expressibility in avoiding overfitting, underperforms due to insufficient expressibility. Thus, expressibility values provide a clear understanding of each model's capacity to represent data patterns effectively.

4. CONCLUSION

This study implemented QNN models to predict the corrosion inhibition efficiency (IE) of drug compounds using three different quantum circuit configurations. The investigation highlighted the critical role of quantum circuit design in the effectiveness of QNN models. Model-01 emerged as the most effective configuration, balancing optimal expressibility and performance, with the lowest error metrics of MSE 38.81, RMSE 6.23, MAE 6.19, and shortest training time of 49 seconds. Model-02 showed higher expressibility but suffered from overfitting, resulting in higher error metrics and longer training time. It achieved an MSE of 66.42, RMSE of 8.15, and MAE of 8.42, with a training time of 49 seconds. Model-03 demonstrated the lowest expressibility and highest error metrics, indicating underfitting and inefficiency despite its simplicity. It had an MSE of 99.40, RMSE of 9.97, and MAE of 9.43, with the longest training time of 58 seconds. The study underscores the importance of balancing expressibility with generalizability

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

in QNN models. Demonstrates that variations in circuit design and parameterization can significantly improve model prediction quality. Future research directions explore a wider variety of quantum circuit designs, test alternative optimization methods and parameterizations, investigate the impact of different quantum gate types and configurations on model performance, develop advanced techniques to mitigate overfitting in deeper models like Model-02. This research advances QNN algorithm development, highlighting the potential for improved prediction quality through careful circuit design and parameterization. The findings suggest that with continued refinement, QNN models could become a powerful tool in materials science and other fields requiring accurate predictive modeling.

ACKNOWLEDGMENT

This research was supported by the Research Center for Materials Informatics, Universitas Dian Nuswantoro.

REFERENCES

[1] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, and H. K. Dipojono, "A machine learning approach for corrosion small datasets", *Npj Mater Degrad*, vol. 7, no. 1, Dec. 2023, doi: 10.1038/s41529-023-00336-7.
[2] M. Akrom, "Investigation Of Natural Extracts As Green Corrosion Inhibitors In Steel Using Density Functional Theory", J. Teori Apl. Fis, vol. 10, no. 1, pp. 89–102, Jan. 2022, doi: 10.23960/jtaf.v10i1.2927.
[3] M. Akrom, "DFT Investigation of Syzygium Aromaticum and Nicotiana Tabacum Extracts as Corrosion Inhibitor", Science Tech, vol. 8, no. 1, pp. 42–48, Feb. 2022, doi: 10.30738/st.vol8.no1.a11775.
[4] S. Budi, M. Akrom, G.A. Trisnapradika, T. Sutojo, W.A.E. Prabowo, "Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets", *Scientific Journal of Informatics*, vol. 10, no. 2, May. 2023, doi: 10.15294/sji.v10i2.43929.
[5] M. Akrom, S. Rustad, A. G. Saputro, and H. K. Dipojono, "Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors", *Comput Theor Chem*, vol. 1229, p. 114307, Nov. 2023, doi: 10.1016/J.COMPTC.2023.114307.
[6] M. Akrom and T. Sutojo, "Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors", Eksergi, vol. 20, no. 2, Art. no. 2, Jul. 2023, doi: 10.31315/e.v20i2.9864.
[7] M. Akrom, S. Rustad, A. G. Saputro, A. Ramelan, F. Fathurrahman, and H. K. Dipojono, "A combination of machine learning model and density functional theory method to predict corrosion inhibition performance of new diazine derivative compounds", Mater Today Commun, vol. 35, Jun. 2023, doi: 10.1016/j.mtcomm.2023.106402.
[8] T. W. Quadri et al., "Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies", Arabian Journal of Chemistry, vol. 15, no. 7, Jul. 2022, doi: 10.1016/j.arabjc.2022.103870.
[9] B. El Ibrahimi et al., "Theoretical evaluation of some α-amino acids for corrosion inhibition of copper in acidic medium: DFT calculations, Monte Carlo simulations and QSPR studies", J King Saud Univ Sci, vol. 32, no. 1, pp. 163–171, Jan. 2020, doi: 10.1016/j.jksus.2018.04.004.
[10] C. Beltran-Perez *et al.*, "A General Use QSAR-ARX Model to Predict the Corrosion Inhibition Efficiency of Drugs in Terms of Quantum Mechanical Descriptors and Experimental Comparison for Lidocaine", *Int J Mol Sci*, vol. 23, no. 9, May. 2022, doi: 10.3390/ijms23095086.
[11] C. T. Ser, P. Žuvela, and M. W. Wong, "Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships", *Appl Surf Sci*, vol. 512, May. 2020, doi: 10.1016/j.apsusc.2020.145612.
[12] A. H. Alamri and N. Alhazmi, "Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors", *Journal of Saudi Chemical Society*, vol. 26, no. 6, Nov. 2022, doi: 10.1016/j.jscs.2022.101536.
[13] T. W. Quadri Et Al., "Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors", Materials Today Communications, vol. 30, Mar. 2022, doi: 10.1016/j.mtcomm.2022.103163.
[14] T. H. Pham, P. K. Le, and D. N. Son, "A data-driven QSPR model for screening organic corrosion inhibitors for carbon steel using machine learning techniques", RSC Adv, vol. 14, no. 16, pp. 11157–11168, Apr. 2024, doi: 10.1039/d4ra02159b.
[15] A. Macaluso, L. Clissa, S. Lodi, and C. Sartori, "A variational algorithm for quantum neural networks", in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, vol. 12142, pp. 591–604, 2020, doi: 10.1007/978-3-030-50433-5_45.
[16] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, "Variational quantum algorithms for nonlinear problems", *Phys Rev A (Coll Park)*, vol. 101, no. 1, Jan. 2020, doi: 10.1103/PhysRevA.101.010301.
[17] A. Zeguendry, Z. Jarir, and M. Quafafou, "Quantum Machine Learning: A Review and Case Studies", *Entropy*, vol. 25, no. 2. MDPI, Feb. 01, 2023. doi: 10.3390/e25020287.
[18] M. Cerezo *et al.*, "Variational quantum algorithms", *Nature Reviews Physics*, vol. 3, no. 9. Springer Nature, pp. 625–644, Sep. 01, 2021. doi: 10.1038/s42254-021-00348-9.
[19] N. Mishra *et al.*, "Quantum Machine Learning: A Review and Current Status", in *Advances in Intelligent Systems and Computing*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 101–145, 2021, doi: 10.1007/978-981-15-5619-7_8.
[20] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks", *Nat Comput Sci*, vol. 1, no. 6, pp. 403–409, Jun. 2021, doi: 10.1038/s43588-021-00084-1.
[21] M. Schuld, I. Sinayskiy, and F. Petruccione, "The quest for a Quantum Neural Network", *Quantum Information Processing*, vol. 13, no. 11. Springer New York LLC, pp. 2567–2586, Oct. 21, 2014. doi: 10.1007/s11128-014-0809-8.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 207−215
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5318

[22] I. B. Obot and S. A. Umoren, "Experimental, DFT and QSAR models for the discovery of new pyrazines corrosion inhibitors for steel in oilfield acidizing environment", *Int J Electrochem Sci*, vol. 15, no. 9, pp. 9066–9080, Sep. 2020, doi: 10.20964/2020.09.72.
[23] A. H. Radhi *et al*., "HOMO-LUMO energies and geometrical structures effecton corrosion inhibition for organic compounds predict by DFT and PM3 methods", *NeuroQuantology*, vol. 18, no. 1, pp. 37–45, 2020, doi: 10.14704/nq.2020.18.1.NQ20105.
[24] M. Akrom, S. Rustad, H. K. Dipojono, "Prediction of Anti-Corrosion performance of new triazole derivatives via Machine learning", Comp. and Theoretical Chemistry, vol. 1236, pg. 114599, 2024, doi: 10.1016/j.comptc.2024.114599.
[25] I. B. Obot and N. O. Obi-Egbedi, "Theoretical study of benzimidazole and its derivatives and their potential activity as corrosion inhibitors", *Corros Sci*, vol. 52, no. 2, pp. 657–660, Feb. 2010, doi: 10.1016/j.corsci.2009.10.017.
[26] M. Akrom, "Green Corrosion Inhibitors for Iron Alloys: A Comprehensive Review of Integrating Data-Driven Forecasting, Density Functional Theory Simulations, and Experimental Investigation", *Journal of Multiscale Materials Informatics*, vol. 1, pg. 22-37, 2024, doi: 10.62411/jimat.v1i1.10495.
[27] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance", *Appl Soft Comput*, vol. 97, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.
[28] M. Akrom, S. Rustad, and H. Kresno, "Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors", Results in Chem., vol. 6, no. September, p. 101126, 2023, doi: 10.1016/j.rechem.2023.101126.
[29] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065. Royal Society of London, Apr. 13, 2016. doi: 10.1098/rsta.2015.0202.
[30] X. Ying, "An Overview of Overfitting and its Solutions", in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2019. doi: 10.1088/1742-6596/1168/2/022022.
[31] J. Qin, "Review of ansatz designing techniques for variational quantum algorithms", in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2634/1/012043.
[32] J. Cheng, H. Wang, Z. Liang, Y. Shi, S. Han, and X. Qian, "TopGen: Topology-Aware Bottom-Up Generator for Variational Quantum Circuits", Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.08190
[33] H. Hirai, "Application of quantum neural network model to a multivariate regression problem", Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.12559
[34] R. D. M. Simoes, P. Huber, N. Meier, N. Smailov, R. M. Fuchslin, and K. Stockinger, "Experimental Evaluation of Quantum Machine Learning Algorithms", *IEEE Access*, vol. 11, pp. 6197–6208, 2023, doi: 10.1109/ACCESS.2023.3236409.
[35] M. Akrom, S. Rustad, and H. K. Dipojono, "Development of quantum machine learning to evaluate the corrosion inhibition capability of pyrimidine compounds", Mater Today Commun, vol. 39, Jun. 2024, doi: 10.1016/j.mtcomm.2024.108758.
[36] M. Fernández-Pendás, E. F. Combarro, S. Vallecorsa, J. Ranilla, and I. F. Rúa, "A study of the performance of classical minimizers in the Quantum Approximate Optimization Algorithm", *J Comput Appl Math*, vol. 404, Apr. 2022, doi: 10.1016/j.cam.2021.113388.
[37] A. V Tatachar, "Comparative Assessment of Regression Models Based On Model Evaluation Metrics", *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net
[38] D. Chicco Corresp, M. J. Warrens, G. Jurman, and D. Chicco, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation analysis evaluation 4", PeerJ Comput Sci, vol. 7, Jul. 2021, doi: 10.7717/peerj-cs.623.