Journal of Multiscale Materials Informatics

Vol. 1, No. 2, August 2024
https://publikasi.dinus.ac.id/index.php/jimat/ ISSN 3047-5724
DOI: 10.62411/jimat.v1i2.11448

Quantum support vector regression for predicting corrosion inhibition of drugs

Akbar Priyo Santosa[1], Muhamad Akrom[2] *

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[2]Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received : August 23, 2024
Revised : August 27, 2024
Accepted : August 29, 2024

Keywords:
QSVR
Corrosion inhibitor
Quantum machine learning
Drug

ABSTRACT
This study evaluates the performance of Quantum Support Vector Regression (QSVR) in predicting material properties using limited data. Experimental results show that the QSVR model consistently produces superior prediction accuracy compared to previous conventional regression models. This improvement is especially evident in the prediction accuracy for small and complex datasets, where QSVR can better capture non-linear patterns. The superiority of QSVR in processing data with a quantum approach provides great potential in developing predictive models in materials science and computational chemistry.

Corresponding Author:
email: m.akrom@dsn.dinus.ac.id

1. INTRODUCTION (Times New Roman, 10 pt, bold)

An electrochemical process known as corrosion occurs when metal surfaces come into contact with corrosive surroundings. It is a major loss-causing factor in several sectors but is most prevalent in the oil and gas. Since corrosion-related problems account for a sizeable amount of the yearly expenses incurred by oil and gas production businesses globally, corrosion is a topic worth researching, particularly in oil field applications [1], [2], [3]. Furthermore, adequate corrosion prevention can aid in averting several possible catastrophes that can result in grave problems, including fatalities, detrimental effects on society, and contamination of the environment and water supplies [4], [5], [6].

In recent years, the development of predictive models to study material properties has become a major focus in materials science and computational chemistry. Accurate predictive models can accelerate the discovery of new materials by reducing the reliance on expensive and time-consuming laboratory experiments. One approach that has been widely used is Support Vector Regression (SVR), which is known to handle high-dimensional data and detect non-linear patterns [7], [8].

However, although SVR offers several advantages, its performance is often limited when applied to small or complex datasets. This has prompted researchers to explore more innovative approaches, such as using quantum computing principles in predictive models. Quantum Support Vector Regression (QSVR) has emerged as a new method that integrates the advantages of SVR with the capabilities of quantum computing to overcome the limitations in non-linear prediction [9], [10].

In this study, we evaluate the performance of QSVR in predicting material properties, especially on limited and complex datasets. We compare the prediction accuracy of the QSVR model with that of conventional regression models that have been used previously and examine the extent to which quantum approaches can improve the prediction results. The results of this research are expected to provide significant contributions to the development of predictive models based on quantum computing and pave the way for broader applications in materials science.

2. METHODS

2.1. Dataset

The method must explain the procedures for obtaining data and data analysis techniques. The research stages and analysis must be explained in detail. A published dataset with 260 data points and 14 molecular descriptors is used in this investigation [11]. Molecular weight (MW), acid dissociation constant (pKa), water solubility (log S), polar surface area (PSA), polarizability (α), the energy of lowest unoccupied molecular orbital (LUMO), the energy of highest occupied molecular orbital (HOMO), Ionization Energy (I), Electron Affinity (A), Electronegativity (eV), Electrophilicity (ω), Hardness (eV), and The Faction Electron Shared (ΔN) are some of these features.

2.2. Preprocessing

Preprocessing is done since some data points have missing values, leading to data cleaning with 78 clean data points. The Min-max scaler approach is then used to scale the features. The results are split into two variables: X and Y. Principal Component Analysis, or PCA, is used to variable X. By generating new, uncorrelated variables and optimizing data variance, PCA is a technique that minimizes information loss while reducing the dimension of a huge dataset to aid in interpretation [12], [13].

2.3. QSVR model

The QSVR developed in this study consists of several main components that work synergistically to improve the accuracy of material property prediction. This model utilizes the principles of quantum computing combined with the traditional SVR algorithm. The first step in the QSVR architecture is mapping classical features to the quantum feature space. This is done using a quantum feature map, where input data is converted into quantum states through a series of gates. This process allows the model to capture complex non-linear patterns in the data more efficiently [14], [15].

One key component in QSVR is the use of a quantum kernel. This quantum kernel is calculated based on the amplitude or probability of the measurement results from the quantum circuit. It measures the similarity between data pairs in the quantum feature space. Because kernel computation is performed in quantum space, QSVR can exploit the advantages of superposition and entanglement to measure similarity more accurately than traditional kernels [16], [17].

The quantum circuit used in QSVR is designed to implement the quantum feature map and quantum kernel optimally. This circuit consists of a series of quantum gates, such as Ry and Rz, applied to qubits to manipulate quantum states. This circuit design minimizes errors and increases the stability of quantum computing, resulting in more accurate prediction results [18], [19].

After the quantum kernel is calculated, the model training process is carried out similarly to conventional SVR by minimizing a loss function that measures the prediction error against the training data. However, optimization is performed in the quantum feature space, allowing for a wider and more diverse solution space exploration. Optimization algorithms such as the Quantum Approximate Optimization Algorithm (QAOA) or variational methods are often used to find the optimal parameters in this stage [20], [21].

After the model is trained, predictions are made by measuring the quantum circuit's results for new input data. These measurement results are converted to the classical world for the final prediction [22], [23]. This measurement process allows QSVR to make more accurate and faster predictions, especially on complex datasets.

The results of quantum computing are then further processed with classical methods to obtain the final prediction. This step involves calculating regression values based on the results of the quantum kernel and optimization. Combining quantum and classical computing provides a hybrid advantage that maximizes the model's predictive performance.

2.4. Model evaluation

The evaluation process is critical in determining the effectiveness of machine learning (ML) models, particularly in identifying the most suitable model for a given task. To achieve a thorough evaluation, several performance metrics are employed. Mean Absolute Deviation (MAD) measures the average of the absolute differences between the observed actual outcomes and the predictions made by the model. Unlike metrics that square the errors, MAD provides a straightforward measure of model accuracy by treating all errors equally without disproportionately penalizing larger errors. This makes MAD useful for obtaining a clear, interpretable understanding of the typical error magnitude. A lower MAD value indicates better predictive accuracy, highlighting the model's ability to produce predictions close to the actual values. Root Mean Squared Error (RMSE) is the square root of MSE and provides an interpretable error measure in the

same units as the output variable. RMSE is particularly valuable in understanding the typical size of the prediction errors and is often preferred because it is more sensitive to outliers. This study uses RMSE to gauge the robustness of the QSVR model's predictions. Mean Absolute Error (MAE) calculates the average of the absolute differences between predicted and actual values. Unlike MSE, which squares the errors, MAE provides a linear score that treats all errors equally, making it easier to interpret. MAE is often used alongside RMSE to give a more complete picture of model performance, with a lower MAE indicating better accuracy. These metrics collectively provide a robust framework for evaluating and comparing the performance of the QSVR model against traditional ML models. The study aims to highlight QSVR's strengths and potential limitations by analyzing these metrics, offering insights into its practical applicability and effectiveness in predictive modeling tasks [24], [25], [26], [27].

3. RESULTS AND DISCUSSION

Table 1. Comparison between QSVR and classical models.

Model	RMSE	MAE	MAD	Ref.
QSVR	4.40	3.33	3.17	This work
SVR	6.28	5.12	5,04	This work
ARX	7.03	-	-	[28]
GB	6.40	4.80	-	[29]

Table 1 presents a comparative analysis of the predictive performance of four models: QSVR, SVR, AutoRegressive with exogenous inputs (ARX), and Gradient Boosting (GB). The models are evaluated using three key metrics: RMSE, MAE, and MAD. The QSVR model achieves the lowest RMSE value of 4.40, indicating superior predictive accuracy compared to the other models. This suggests that QSVR is better at minimizing the squared differences between the predicted and actual values, making it a more reliable model for this dataset. The RMSE for SVR is 6.28, significantly higher than that of QSVR. This indicates that SVR produces larger errors on average, which affects its overall accuracy. The ARX model has an RMSE of 7.03, the highest among the models compared. This result implies that ARX struggles the most regarding prediction accuracy, likely due to its reliance on simpler linear relationships. With an RMSE of 6.40, GB performs better than ARX but is still less accurate than QSVR and SVR. This suggests that while GB captures some complex patterns, it doesn't match the predictive power of the quantum-enhanced QSVR.

QSVR also outperforms the other models regarding MAE, with a value of 3.33. This indicates that, on average, the absolute difference between the predicted and actual values is smaller, making QSVR the most precise model in this comparison. The MAE for SVR is 5.12, indicating that its predictions are generally less accurate than QSVR. This aligns with the RMSE results and reinforces QSVR's superior performance. GB shows an MAE of 4.80, slightly better than SVR but still significantly higher than QSVR, suggesting that while GB is reasonably accurate, it cannot match the precision of QSVR. The MAE for ARX is not provided in the table, which limits a direct comparison. However, given its high RMSE, it can be inferred that ARX likely has a higher MAE, indicating lower predictive accuracy.

The MAD for QSVR is 3.17, the lowest among the models. This further underscores the model's ability to produce consistent predictions with minimal deviation from the actual values. SVR has an MAD of 5.04, higher than QSVR, indicating greater prediction variability. This result highlights QSVR's advantage in producing more stable and accurate predictions. The MAD for GB is not provided, which limits a direct comparison. However, based on the RMSE and MAE values, it can be inferred that GB may have a MAD value higher than QSVR, further supporting the superior performance of QSVR. Similar to GB, the MAD for ARX is not provided. However, given ARX's high RMSE, it likely has a higher MAD, which is consistent with its lower overall accuracy.

The analysis demonstrates that the QSVR model outperforms the other models across all provided metrics. Its lower RMSE, MAE, and MAD values indicate that QSVR is better at capturing the underlying patterns in the data, leading to more accurate and stable predictions. This superior performance is likely due to the quantum-enhanced feature mapping and optimization processes in QSVR, which allow it to handle complex, non-linear relationships in the data. In contrast, traditional models like SVR, ARX, and GB exhibit higher error metrics, reflecting their limitations in predictive accuracy. While commonly used for regression tasks, SVR appears to fall short compared to QSVR, likely due to its reliance on classical optimization methods. ARX, with the highest RMSE, is particularly disadvantaged due to its simplistic

linear approach, making it less suitable for the complex relationships in the dataset. While performing better than ARX, GB still lags behind QSVR, suggesting that even advanced classical models may struggle to match the capabilities of quantum-enhanced methods. These results suggest that QSVR offers a promising advancement in predictive modeling, particularly for applications where high accuracy and consistency are critical. The findings of this study reinforce the potential of quantum machine learning techniques to surpass traditional methods, especially as quantum computing technology continues to evolve.

4. CONCLUSION

This study demonstrates the significant advantages of the QSVR model over traditional regression models such as SVR, ARX, and GB. The results show that QSVR consistently outperforms these models across multiple evaluation metrics, including RMSE, MAE, and MAD. The QSVR model achieved the lowest RMSE, MAE, and MAD values, indicating superior predictive accuracy and stability. This enhanced performance is attributed to the quantum-enhanced feature mapping and optimization processes that allow QSVR to capture complex, non-linear relationships in the data effectively. In contrast, traditional models like SVR, ARX, and GB exhibited higher error metrics, reflecting their limitations in handling complex datasets. The findings of this study highlight the potential of quantum machine learning techniques, particularly QSVR, in advancing predictive modeling capabilities. As quantum computing technology continues to mature, it is expected that QSVR and similar quantum-enhanced models will play a critical role in various applications, offering more accurate and reliable predictions than classical methods. In summary, QSVR presents a promising approach for tasks that require high precision and consistency, and its integration into practical applications could lead to significant improvements in predictive performance across a wide range of domains.

REFERENCES

[1] M. Akrom, T. Sutojo, A. Pertiwi, S. Rustad, H.K. Dipojono, Investigation of Best QSPR-Based Machine Learning Model to Predict Corrosion Inhibition Performance of Pyridine-Quinoline Compounds, J Phys Conf Ser, 2673(1), 012014 (2023), https://doi.org/10.1088/1742-6596/2673/1/012014.

[2] M. Akrom, Green corrosion inhibitors for iron alloys: a comprehensive review of integrating data-driven forecasting, density functional theory simulations, and experimental investigation. J Mult Mater Inf, 1(1), 22–37 (2024), https://doi.org/10. 62411/jimat.v1i1.10495

[3] M. Akrom, S. Rustad, H.K. Dipojono, A machine learning approach to predict the efficiency of corrosion inhibition by natural product-based organic inhibitors, *Phys Scr*, 99(3), 036006 (2024), https://doi.org/10.1088/1402-4896/ad28a9.

[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*, 10(2), (2023), https://doi.org/10.15294/sji.v10i2.43929.

[5] M. Akrom, S. Rustad, H.K. Dipojono, Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors, Results in Chemistry 6 (2023) 101126, https://doi.org/10.1016/j. rechem.2023.101126.

[6] M. Akrom, S. Rustad, A.G. Saputro, H.K. Dipojono, Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors, Comput. Theor. Chem. 1229 (2023) 114307, https://doi.org/10.1016/ J.COMPTC.2023.114307.

[7] M. Schuld, I. Sinayskiy, and F. Petruccione, The quest for a quantum support vector machine. Quantum Information Processing, 13(11), 2567-2586 (2014).

[8] V. Havlíček, A.D. Córcoles, K. Temme, A.W. Harrow, A. Kandala, J.M. Chow, and J.M. Gambetta. Supervised learning with quantum-enhanced feature spaces. Nature, 567(7747), 209-212 (2019).

[9] M. Akrom, S. Rustad, H.K. Dipojono, Prediction of Anti-Corrosion performance of new triazole derivatives via Machine learning, Comput. Theor. Chem. 1236 (2024), https://doi.org/10.1016/j.comptc.2024.114599.

[10] M. Akrom, Investigation of natural extracts as green corrosion inhibitors in steel using density functional theory, *Jurnal Teori dan Aplikasi Fisika*, 10(1), 89-102 (2022), https://doi.org/10.23960%2Fjtaf.v10i1.2927.

[11] 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. 23 (9) (2022), https://doi. org/10.3390/ijms23095086.

[12] M. Akrom, S. Rustad, H.K. Dipojono. Development of quantum machine learning to evaluate the corrosion inhibition capability of pyrimidine compounds. Materials Today Communications, 39, 108758 (2024), https://doi.org/10.1016/j.mtcomm.2024.108758.
[13] M. Akrom, S. Rustad, H.K. Dipojono, SMILES-based machine learning enables the prediction of corrosion inhibition capacity, MRS Commun 14 (2024) 379–387, https://doi.org/10.1557/s43579-024-00551-6.
[14] M. Akrom, S. Rustad, H.K. Dipojono. Variational quantum circuit-based quantum machine learning approach for predicting corrosion inhibition efficiency of pyridine-quinoline compounds. Materials Today Quantum, 2, 100007 (2024), https://doi.org/10.1016/j.mtquan.2024.100007.
[15] Aaronson, S., & Arkhipov, A. (2011). "The Computational Complexity of Linear Optics." Proceedings of the ACM Symposium on Theory of Computing (STOC).
[16] Wang, D., Guo, F., & Guo, Y. (2016). "A novel solution to multi-class classification problem using support vector machine." *Journal of Ambient Intelligence and Humanized Computing, 7*(4), 563-571.
[17] Chang, H., Liu, Y., & Bai, Y. (2017). "A new multi-category support vector machine algorithm." *Soft Computing, 21*(6), 1377-1389.
[18] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, 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. 35 (2023) 106402, https://doi.org/10.1016/J. MTCOMM.2023.106402.
[19] M. Akrom, et al., DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract, Appl. Surf. Sci. 615 (2023), https://doi.org/10.1016/j. apsusc.2022.156319.
[20] D. Alaminos, M.B. Salas, M.A. Fernández-Gámez, Quantum computing and deep learning methods for GDP growth forecasting, Comput. Econ. (2021) http://dx.doi.org/10.1007/s10614-021-10110-z.
[21] F.J. García-Peñalvo, Desarrollo de estados de la cuestión robustos: Revisiones sistemáticas de literatura, Educ. Knowl. Soc. (EKS) 23 (2022) http://dx.doi.org/10.14201/eks.28600, URL http://repositorio.grial.eu/handle/grial/2568.
[22] W. O'Quinn, S. Mao, Quantum machine learning: Recent advances and outlook, IEEE Wirel. Commun. 27 (3) (2020) 126–131, http://dx.doi.org/10.1109/MWC.001.1900341.
[23] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, Int. J. Surg. 8 (5) (2010) 336–341, http://dx.doi.org/10.1016/j.ijsu.2010.02.007.
[24] M. Petticrew, H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, vol. 11, 2006, http://dx.doi.org/10.1002/9780470754887.
[25] Y. Huang, H. Lei, X. Li, Q. Zhu, W. Ren, X. Liu, Quantum generative model with variable-depth circuit, Comput. Mater. Contin. 65 (1) (2020) 445–458, http://dx.doi.org/10.32604/cmc.2020.010390.
[26] M. Srikumar, C.D. Hill, L.C.L. Hollenberg, Clustering and enhanced classification using a hybrid quantum autoencoder, Quantum Sci. Technol. 7 (1) (2021) 015020, http://dx.doi.org/10.1088/2058-9565/ac3c53.
[27] D. Konar, S. Bhattacharyya, B.K. Panigrahi, E.C. Behrman, Qutrit-inspired fully self-supervised shallow quantum learning network for brain tumor segmentation, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–15, http://dx.doi.org/10. 1109/tnnls.2021.3077188, arXiv:2009.06767.
[28] 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. 23 (9) (2022), https://doi. org/10.3390/ijms23095086.
[29] T.H. Pham, P.K. Le, D.N. Son, A data-driven QSPR model for screening organic corrosion inhibitors for carbon steel using machine learning techniques, RSC Adv. 14 (16) (2024) 11157–11168, https://doi.org/10.1039/d4ra02159b.