Heterogeneous Transfer Learning Optimization for Greenhouse Gas Emissions Prediction Using Quantum Annealing

Wahyu Hidayat, Nur Ulfa Maulidevi and Kridanto Surendro

September 23, 2024

Heterogeneous Transfer Learning Optimization For Greenhouse Gas Emissions Prediction Using Quantum Annealing

1st Wahyu Hidayat
[a]School of Electrical Engineering and Informatics,
Institut Teknologi Bandung
Bandung, Indonesia
[b]School of Applied Science,
Telkom University
Bandung, Indonesia
33223004@std.stei.itb.ac.id

2nd Nur Ulfa Maulidevi
School of Electrical Engineering and Informatics,
Institut Teknologi Bandung
Bandung, Indonesia
ulfa@itb.ac.id

3rd Kridanto Surendro
School of Electrical Engineering and Informatics,
Institut Teknologi Bandung
Bandung, Indonesia
endro@itb.ac.id

***Abstract*— Despite its urgency, the prediction of greenhouse gas (GHG) emissions at the city level is hampered by limited quality and quantity of training data so most predictions of GHG emissions are carried out at the country level, using different feature spaces. Heterogeneous Transfer Learning (HeTL) is considered capable of getting around this limitation due to its ability to facilitate the transfer of knowledge between domains that have different feature spaces and distributions. However, the implementation of HeTL is still haunted by the potential of negative transfer in the knowledge transfer process. Current studies on mitigating negative transfer in HeTL still rely heavily on classical optimization techniques and focus solely on either feature-level or instance-level optimization. In this paper, a method is proposed to optimize the knowledge transfer process in HeTL using quantum annealing. The proposed optimization is carried out in three stages: (1) feature alignment, (2) common feature optimization, and (3) data instance optimization. The proposed method seeks to optimize the knowledge transfer process at both the feature and instance levels. It utilizes a combination of classical computing and quantum computing, thereby combining the advantages of the classical approach and the quantum approach to obtain optimal results.**

Keywords — heterogeneous transfer learning, quantum annealing, greenhouse gas emissions

I. Introduction

Greenhouse gas emissions refer to the total emissions of greenhouse gases (GHG) produced by an individual, product, service, place, event, or organization, expressed in units of carbon dioxide equivalent (CO2e) [1]. These emissions include various gases such as carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and fluorinated gases, all of which play a role in global warming and climate change. Globally, urban areas have become major contributors to greenhouse gas (GHG) emissions [2]. Even though the urban area is only 2.4% of the global area, it produces a staggering 80% of emissions related to energy consumption [3] [4]. In addition, 60% of global CO2 emissions from fossil fuels are produced by urban areas. There is an urgent need for high-quality city-level emissions inventories to aid international climate mitigation efforts [5]. Thus, urban areas play an important role in efforts to reduce greenhouse gas emissions globally [2].

Predicting greenhouse gas emissions is very important in reducing greenhouse gas emissions and mitigating climate change. With the ability to accurately predict future greenhouse gas emissions, steps for intervention can be formulated through changes in behavior or policies related to emission trigger factors.

One of the challenges of predicting GHG emissions with machine learning is the limited training data [6]. GHG emission data at the city level is generally hampered by a lack of quantity or low quality of statistical data related to energy, especially for some less developed areas [2]. The concentration of emissions in urban areas highlights the critical need for accurate and detailed city-level emissions inventories. High-quality data at the city level can support international climate mitigation efforts by providing a clear picture of where emissions are highest and where interventions can be most effective. Unfortunately, a global, accessible, and standardized dataset of city-level emissions inventories is still unavailable [5]. Currently available city-level GHG emissions data such as CEADS and MEIC are considered low in area and time coverage [5]. Developing cities in South America, Southeast Asia, the Middle East & Africa are some of the areas where data is most difficult to obtain. In contrast, most CO2 emissions inventories are performed at the national level, as it is more challenging to obtain fossil fuel consumption data at the city level [5].

Transfer learning (TL) is considered a feasible solution to get around the limited training data for predicting GHG emissions [6]. Transfer learning is motivated by the observation that humans can leverage prior knowledge to address new challenges more efficiently or effectively. This approach permits differences between the domains, tasks, and distributions used in training and those used in testing, allowing knowledge to be drawn from one or more source tasks and applied to a target task [7][8].

It should be noted that there are differences in the feature space used in predicting greenhouse gas emissions at the city level and the country level. Traditional Transfer Learning (TL) is only applicable in cases where the source and target domains share identical feature spaces and distributions [9]. In this case, the difference in feature space between city-level GHG emissions data and country-level GHG emissions data opens up opportunities for a specific type of transfer learning, i.e. Heterogeneous Transfer Learning (HeTL). HeTL enables the transfer of knowledge between domains that have different feature spaces and distributions, allowing machine learning algorithms to be used in a variety of applications without being limited by the need for identical distributions of training data and test data [9]. Generally, HeTL has various advantages over traditional transfer learning, i.e. better flexibility, the ability to handle data with complex connections, smaller domain adaptation costs, smaller risk of overfitting, and better scalability [9].

However, in the knowledge transfer process of HeTL, there is the potential for negative transfer, a phenomenon where the knowledge transfer process worsens model performance due to cross-domain noise. Most contemporary HeTL methods have not explicitly addressed this negative transfer problem [9]. The knowledge transfer process can be carried out at the feature level or instance level. Both aim to find features and instances that can be transferred from the source domain (SD) to the target domain (TD) to maximize the closeness between SD and TD to avoid negative transfer [8] which can be seen as an optimization problem.

On the other hand, Quantum Annealing (QA) is designed to find the global minimum of a cost function to solve optimization problems [10]. QA is a quantum computing approach that utilizes the principles of quantum mechanics to search for the minimum energy state of a system [10]. QA has been successfully implemented for optimization tasks in various cases and problem domains, including machine learning.

These successful implementations of QA have highlighted its promising potential for optimizing HeTL. Therefore, studies that present methods of implementing QA to optimize the knowledge transfer process in HeTL for city-level GHG emission prediction models are needed. This paper proposes a method to optimize the knowledge transfer process in heterogeneous transfer learning using quantum annealing. The rest of this paper is organized as follows: Section II discusses related studies about GHG emissions prediction, transfer learning, and quantum annealing; Section III presents a detailed proposed method to optimize heterogeneous transfer learning using quantum annealing; and Section IV highlights the conclusion and future research opportunities.

II. Related Work

Predicting GHG emissions using machine learning can be done at the macroscopic level using data on economic development, population, urbanization energy consumption, industrial structure, and technological advancement [6], as well as through predictions of building energy needs. The five most popular models used in macroscopic GHG emissions prediction are Long Short Term Memory (LSTM), Back Propagation Neural Network (BPNN), Support Vector Machine (SVM), Extreme Learning Machine (ELM), and Random Forest (RF) [6]. Meanwhile, three architectural groups of artificial neural networks, i.e. FFNN, CNN, and RNN are the most popular models used to predict the energy needs in buildings [11]. GHG emission predictions have also been carried out on a smaller scale, for example at universities, both based on building energy use [12] and based on human behavior and activities [13].

Several studies on GHG prediction at the city level mention obstacles related to the quality and quantity of data, limited global datasets available, as well as the limited area and time coverage of the available datasets [2] [5] while other studies focus on identifying the key predictors that can be used to predict city-level GHG emission [14]. Transfer learning is mentioned in [6] as a feasible solution to the problem of limited data for predicting GHG emissions. The development, categorization, and implementation of transfer learning has been widely discussed in [7] [8] and [15].

Heterogeneous Transfer Learning (HeTL) is a type of transfer learning that has the special ability to carry out knowledge transfer between domains that have different feature spaces. Various implementations and recent developments of HeTL are discussed in [9]. HeTL is more complex than traditional transfer learning but has the potential to significantly improve model efficiency [9]. Several studies have tried to optimize HeTL [16] [17] [18] [19] [20] [21]. Most of these studies focus on optimizing HeTL specifically for classification tasks [16] [17] [18] [19] [20]. Some of them carry out optimization at the feature level that aims to overcome the domain differences between the source domain and target domain feature spaces [16] [18] [20]. Other studies carry out optimization that focuses on knowledge transfer from source domain to target domain at the instance level [17] [19]. The optimization techniques used are still dominated by classical optimization techniques such as Hybrid PCA & Correlation Effect Analysis [16], Adaptive Clustering Transfer Learning [17], Particle Swarm Optimization [18], Successive Convex Approximation [19], alternating direction method of multipliers (ADMM) [20], and Centroid Distribution [21]. The feature-based HeTL is still relatively new and not widely used [9].

Generally, any GHG prediction model can be optimized both at the data level and algorithm level using several optimization techniques, although some models still have tendencies to fall into local optima [6]. Currently, classical swarm intelligence-based optimization techniques are prominent in the optimization of GHG emissions prediction models [6] [11]. However, quantum-based optimization techniques also have shown their potential. Quantum annealing (QA) has been implemented in various optimization tasks related to GHG emissions issues, for example, to optimize real-time traffic light control [22], assist bike-sharing operators in optimizing bike load balancing processes [23], optimize virtual machine allocation and tasks in cloud infrastructure [24] and optimize energy consumption in buildings [25]. Additionally, QA has been shown to optimize ensemble learning [26] and the implementation of QA for machine learning has been thoroughly discussed in [27]. The wide application scenarios indicate that QA has advantages and flexibility in various use cases and has great potential to be used to optimize GHG emission prediction models.

III. Proposed Method

To reduce the risk of negative transfer and increase the accuracy of city-level greenhouse gas emission prediction models, this paper proposes a knowledge transfer process optimization for heterogeneous transfer learning (HeTL) using quantum annealing (QA). This optimization process involves three interrelated steps: (1) identifying shared features between the source and target domains through feature alignment, (2) optimizing these common features to determine the best combination, and (3) optimizing data instances to select the ideal set of data instances for use as training data. An overview of the proposed method is shown in Figure 1.

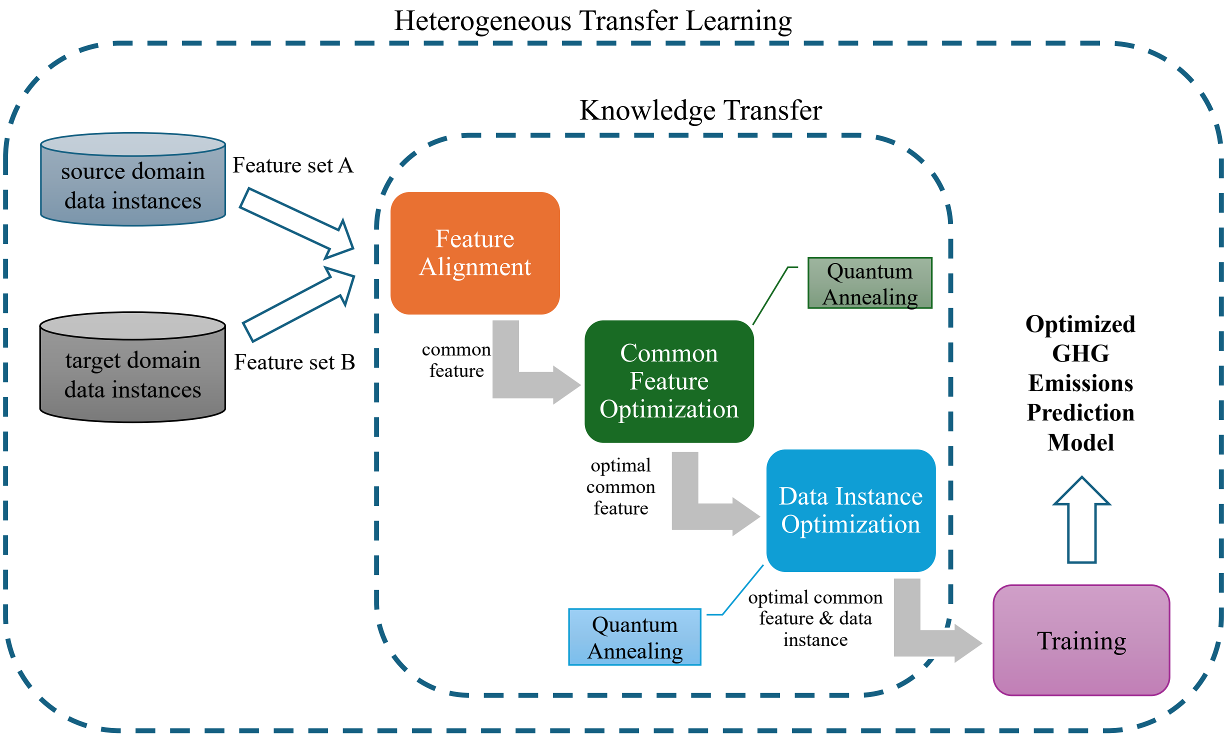

Figure 1. The proposed method

A. Feature Alignment

Considering the differences in feature space x between the source domain D_S and the target domain D_T, the feature alignment stage aims to find common features from both domains. After going through the pre-processing stage to improve data quality, the data from the source domain D_S and target domain D_T go through a feature transformation process. The flow of the feature alignment stage is shown in Figure 2.

Figure 2. The feature alignment stage (classical computing)

The feature transformation can be done using 2 approaches, namely asymmetric transformation (AST) and symmetric transformation (ST) as described [9]. To discover a common latent feature space and then convert the source and target features into new representations, ST takes the target feature space X_T and the source feature space X_S and learns to transform them into a shared space X_C, facilitating adaptation.

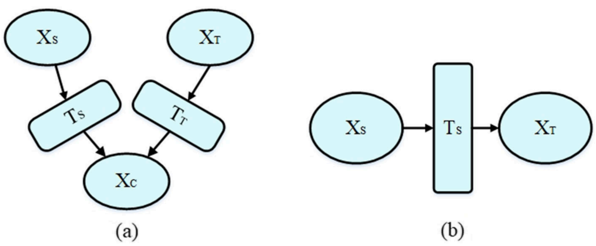

Figure 3. Symmetric (a) and asymmetric (b) feature transformation [9]

On the other hand, AST applies a transformation function search T_S, which transforms the source features X_S to match the target features X_T. The illustration of ST and AST is depicted in Figure 3.

B. Common Feature Optimization

Generally, the feature alignment stage produces common features whose number increases rapidly compared to the number of original features from the source domain and target domain. However, not all of these common features are suitable for inclusion in the training process because there is the potential for negative transfer. The main goal of the common feature optimization stage is to find the optimal combination of common features using quantum annealing. An overview of the steps for the common feature optimization stage is shown in Figure 4.

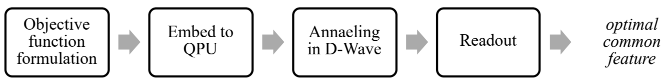

Figure 4. The common feature optimization stage (quantum computing)

To formulate the objective function, a Quadratic Unconstrained Binary Optimization (QUBO) objective function can be used. Basic QUBO formulation can be expressed as [28] :

$$f(x) = \sum_{i} Q_{i,i} x_i + \sum_{i<j} Q_{i,j} x_i x_j \tag{1}$$

where x is a vector of binary decision variables and Q is an upper triangle matrix of the real weight.

C. Data Instance Optimization

Previously, the common feature optimization stage produces an optimal common feature that maximizes the similarity between the source domain and the target domain. However, not all instances from the source domain are suitable to be included in the training process due to the potential of negative transfer, so it is necessary to select data instances to get an optimal subset of the source domain. Figure 4 shows an overview of the steps for optimizing instance data using quantum annealing.

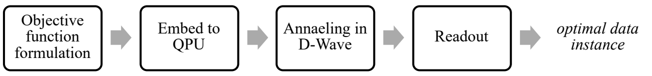

Figure 5. The data instance optimization stage (quantum computing)

Quantum annealing is used in the common feature optimization and data instance optimization steps. In these two stages, there are four main steps taken; 1) formulating a classical approach to the QUBO objective function; 2) embedding in the Quantum Processing Unit (QPU); 3) carrying out the annealing process on the D-Wave quantum annealer device; and 4) reading the results (readout). The main goal of this data instance optimization stage is to find the optimal subset of the source domain using quantum annealing. This optimal subset will act as training data in the training phase to produce the optimized GHG emissions prediction model.

D. Performance Metrics

The performance of the greenhouse gas emission prediction models can be measured using the following metrics[29][30]:

1) RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2} \tag{2}$$

2) Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|X_i - Y_i| \tag{3}$$

3) R-Squared (R2)

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{\sum_{i=1}^{N}(\bar{Y} - Y_i)^2} \tag{4}$$

4) MAPE (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{N} - \sum_{i=1}^{N}\left|\frac{Y_i - X_i}{Y_i}\right| \tag{5}$$

where X_i is the actual value for the i^{th} trial, Y_i is the predicted value for the i^{th} trial, and N is the number of prediction trials.

Other evaluation metrics can also be used, including Adjusted R2, Coefficient of Variation of Root Mean Square Error (CV-RMSE), Normalized Root Mean Square Error (NRMSE), Mean Absolute Deviation (MAD), etc [30].

IV. Conclusion

This paper presents a method to optimize the knowledge transfer process in heterogeneous transfer learning using quantum annealing. The optimization is carried out to minimize negative transfer in the knowledge transfer process through three interconnected steps; (1) feature alignment: finding common features between the source domain and target domain, (2) common feature optimization: finding the optimal combination of common features, and (3) data instance optimization: finding the optimal combination of data instances to be included as training data. The common feature optimization stage and the data instance optimization stage utilize quantum computing, while other stages utilize classical computing. Unlike existing methods that focus solely on either the feature or the instance level, the proposed method seeks to optimize the knowledge transfer process at both the feature and instance levels. By utilizing a combination of classical computing and quantum computing, a hybrid approach combines the strength of the classical approach and the quantum approach to obtain optimal results. Since this paper focuses on regression tasks using tabular data, there are opportunities for further studies about the potential of utilizing quantum annealing to optimize the knowledge transfer process in heterogeneous transfer learning used for different tasks using various data types.

Acknowledgment

The authors gratefully acknowledge the Lembaga Pengelola Dana Pendidikan (LPDP) for their support through the Beasiswa Pendidikan Indonesia (BPI) for doctoral study at Institut Teknologi Bandung (grant no. 202327091777). The authors also extend their appreciation to Institut Teknologi Bandung and Telkom University for providing financial assistance and research facilities.

References

[1] L. A. Wright, S. Kemp, and I. Williams, "'Carbon footprinting': towards a universally accepted definition," *https://doi.org/10.4155/cmt.10.39*, vol. 2, no. 1, pp. 61–72, Feb. 2014, doi: 10.4155/CMT.10.39.

[2] Y. Li and Y. Sun, "Modeling and predicting city-level CO2 emissions using open access data and machine learning," *Environmental Science and Pollution Research*, vol. 28, no. 15, pp. 19260–19271, Apr. 2021, doi: 10.1007/S11356-020-12294-7/FIGURES/7.

[3] X. Zhang, F. Yan, H. Liu, and Z. Qiao, "Towards low carbon cities: A machine learning method for predicting urban blocks carbon emissions (UBCE) based on built environment factors (BEF) in Changxing City, China," *Sustain Cities Soc*, vol. 69, p. 102875, Jun. 2021, doi: 10.1016/J.SCS.2021.102875.

[4] X. Sun, Z. Mi, A. Sudmant, D. M. Coffman, P. Yang, and R. Wood, "Using crowdsourced data to estimate the carbon footprints of global cities," *Advances in Applied Energy*, vol. 8, p. 100111, Dec. 2022, doi: 10.1016/J.ADAPEN.2022.100111.

[5] D. Huo *et al.*, "Carbon Monitor Cities near-real-time daily estimates of CO2 emissions from 1500 cities worldwide," *Sci Data*, vol. 9, no. 1, p. 533, Sep. 2022, doi: 10.1038/s41597-022-01657-z.

[6] Y. Zhao, R. Liu, Z. Liu, L. Liu, J. Wang, and W. Liu, "A Review of Macroscopic Carbon Emission Prediction Model Based on Machine Learning," *Sustainability (Switzerland)*, vol. 15, no. 8, 2023, doi: 10.3390/su15086876.

[7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[8] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[9] S. Khan, P. Yin, Y. Guo, M. Asim, and A. A. Abd El-Latif, "Heterogeneous transfer learning: recent developments, applications, and challenges," *Multimed Tools Appl*, 2024, doi: 10.1007/s11042-024-18352-3.

[10] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, "Quantum annealing: A new method for minimizing multidimensional functions," *Chem Phys Lett*, vol. 219, no. 5–6, pp. 343–348, Mar. 1994, doi: 10.1016/0009-2614(94)00117-0.

[11] C. Lu, S. Li, and Z. Lu, "Building energy prediction using artificial neural networks: A literature survey," *Energy Build*, vol. 262, p. 111718, 2022, doi: 10.1016/j.enbuild.2021.111718.

[12] N. U. Maulidevi, B. S. K. Aji, E. Hikmawati, and K. Surendro, "Modelling Integrated Sustainability Monitoring System for Carbon Footprint in Higher Education Buildings," *IEEE Access*, pp. 1–1, 2023, doi: 10.1109/ACCESS.2023.3333890.

[13] M. Hans, E. Hikmawati, and K. Surendro, "Predictive Analytics Model for Optimizing Carbon Footprint From Students' Learning Activities in Computer Science-Related Majors," *IEEE Access*, vol. 11, pp. 114976–114991, 2023, doi: 10.1109/ACCESS.2023.3324725.

[14] C. Franco *et al.*, "Key predictors of greenhouse gas emissions for cities committing to mitigate and adapt to climate change," *Cities*, vol. 137, p. 104342, Jun. 2023, doi: 10.1016/J.CITIES.2023.104342.

[15] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst Appl*, vol. 242, p. 122807, May 2024, doi: 10.1016/J.ESWA.2023.122807.

[16] H. T. Gidey, X. Guo, L. Li, and Y. Zhang, "Heterogeneous Transfer Learning for Wi-Fi Indoor Positioning Based Hybrid Feature Selection," *Sensors*, vol. 22, no. 15, p. 5840, Aug. 2022, doi: 10.3390/s22155840.

[17] Z. Huan *et al.*, "Learning to Select Instance: Simultaneous Transfer Learning and Clustering," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA: ACM, Jul. 2021, pp. 1950–1954. doi: 10.1145/3404835.3462992.

[18] G. Castillo-García, L. Morán-Fernández, and V. Bolón-Canedo, "Feature selection for domain adaptation using complexity measures and swarm intelligence," *Neurocomputing*, vol. 548, p. 126422, Sep. 2023, doi: 10.1016/j.neucom.2023.126422.

[19] M. Askarizadeh, M. Hussien, A. Morsali, and K. K. Nguyen, "Modeling and Optimizing Resource-Constrained Instance-Based Transfer Learning," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, IEEE, Dec. 2022, pp. 6366–6371. doi: 10.1109/GLOBECOM48099.2022.10001716.

[20] Y. Yan *et al.*, "Transferable Feature Selection for Unsupervised Domain Adaptation," *IEEE Trans Knowl Data Eng*, vol. 34, no. 11, pp. 5536–5551, Nov. 2022, doi: 10.1109/TKDE.2021.3060037.

[21] J. Zhang, W. Zhou, X. Chen, W. Yao, and L. Cao, "Multi-Source Selective Transfer Framework in Multi-Objective Optimization Problems," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2020, doi: 10.1109/TEVC.2019.2926107.

[22] A. Singh, C.-Y. Lin, C.-I. Huang, and F.-P. Lin, "Quantum Annealing Approach for the Optimal Real-time Traffic Control using QUBO," in *2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, IEEE, Nov. 2021, pp. 74–81. doi: 10.1109/SNPD51163.2021.9704985.

[23] R. Harikrishnakumar, S. F. Ahmad, and S. Nannapaneni, "Comparing Quantum Optimization Solvers for Rebalancing Analysis of Bike Sharing System," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2022, pp. 753–755. doi: 10.1109/QCE53715.2022.00106.

[24] V. Uotila and J. Lu, "Quantum Annealing Method for Dynamic Virtual Machine and Task Allocation in Cloud Infrastructures from Sustainability Perspective," in *2023 IEEE 39th International Conference on Data Engineering Workshops (ICDEW)*, IEEE, Apr. 2023, pp. 105–110. doi: 10.1109/ICDEW58674.2023.00023.

[25] Z. Deng, X. Wang, and B. Dong, "Quantum computing for future real-time building HVAC controls," *Appl Energy*, vol. 334, p. 120621, Mar. 2023, doi: 10.1016/j.apenergy.2022.120621.

[26] L. P. Yulianti, J. Santoso, A. Trisetyarso, and K. Surendro, "Hybrid Classical-Quantum Optimization for Ensemble Learning," *2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2022*, 2022, doi: 10.1109/ICAICTA56449.2022.9932950.

[27] L. P. Yulianti and K. Surendro, "Implementation of Quantum Annealing: A Systematic Review," *IEEE Access*, vol. 10, no. June, pp. 73156–73177, 2022, doi: 10.1109/ACCESS.2022.3188117.

[28] F. Glover, G. Kochenberger, and Y. Du, "Quantum Bridge Analytics I: a tutorial on formulating and using QUBO models," *4OR*, vol. 17, no. 4, pp. 335–371, Dec. 2019, doi: 10.1007/s10288-019-00424-y.

[29] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[30] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review," *Eng Appl Artif Intell*, vol. 115, no. June, p. 105287, 2022, doi: 10.1016/j.engappai.2022.105287.