VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i2.7040

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

QUANTUM-ASSISTED FEATURE SELECTION FOR IMPROVING PREDICTION MODEL ACCURACY ON LARGE AND IMBALANCED DATASETS

M. Safii[1*]; Mochamad Wahyudi[2]; Dedy Hartama[1]

Informatika[1]
STIKOM Tunas Bangsa, Pematangsiantar, SUMUT-Indonesia[1]
https://www.stikomtunasbangsa.ac.id[1,3]
m.safii@amiktunasbangsa.ac.id[1*], dedyhartama@amiktunasbangsa.ac.id[3]

Informatika[2]
Universitas Bina Sarana Infomatika, Jakarta, Indonesia[2]
https://www.bsi.ac.id[2]
wahyudi@bsi.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *One of the biggest obstacles to creating precise machine learning models is choosing representative and pertinent characteristics from big, unbalanced datasets. While too many features raise the risk of overfitting and computational expense, class imbalance frequently results in decreased accuracy and bias. The Simulated Annealing technique is used in this study to tackle a Quadratic Unconstrained Binary Optimization (QUBO) problem that is formulated as a quantum-assisted feature selection method to handle these problems. The technique seeks to reduce inter-feature redundancy and the number of selected features. There are 102,487 samples in the majority class and 11,239 in the minority class, totaling 28 characteristics in the experimental dataset. Nine ideal features were found during the feature selection method (12, 14, 15, 22, 23, 24, 25, 27, and 28). Ten-fold cross-validation was used to assess a Random Forest Classifier that was trained using an 80:20 split. With precision, recall, f1-score, and accuracy all hitting 1.00, the suggested QUBO+SMOTE method demonstrated exceptional performance. Comparatively, QUBO without SMOTE performed worse with accuracy 0.95 and minority-class f1-score of only 0.71, whereas a traditional Recursive Feature Elimination (RFE) approach obtained accuracy 0.97 with minority-class f1-score of 0.94. These findings indicate that QUBO can reduce dimensionality and address class imbalance which requires its integration with SMOTE. This study demonstrates how quantum computing can enhance the effectiveness and efficiency of machine learning, especially for large-scale imbalanced datasets.*

Keywords: *feature selection, prediction, quantum, quadratic unconstrained binary optimization (QUBO), random forest.*

Intisari— *Salah satu hambatan terbesar dalam menciptakan model pembelajaran mesin yang tepat adalah memilih karakteristik yang representatif dan relevan dari dataset besar yang tidak seimbang. Di sisi lain, beberapa faktor meningkatkan risiko overfitting dan biaya komputasi, kekurangan keseimbangan sering mengakibatkan masalah bias dan akurasi. Teknik Simulated Annealing digunakan dalam penelitian ini untuk mengatasi masalah Quadratic Unconstrained Binary Optimization (QUBO) yang diformulasikan sebagai metode pemilihan fitur yang dibantu kuantum untuk menangani masalah ini. Teknik ini berupaya mengurangi redundansi antar fitur dan jumlah fitur yang dipilih. Terdapat 102.487 sampel dalam kelas mayoritas dan 11.239 dalam kelas minoritas dengan total 28 karakteristik. Sembilan fitur ideal ditemukan selama metode pemilihan fitur (12, 14, 15, 22, 23, 24, 25, 27, dan 28). Validasi Ten-fold cross digunakan untuk menilai*

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7040

Klasifikasi random forest yang dilatih menggunakan pembagian 80:20. Dengan presisi, recall, f1-score, dan akurasi semuanya mencapai 1.00, metode QUBO+SMOTE yang disarankan menunjukkan kinerja yang luar biasa. Jika dibandingkan QUBO tanpa SMOTE memiliki kinerja yang lebih buruk dengan akurasi 0.95 dan f1-score untuk kelas minoritas hanya 0.71, sedangkan pendekatan Eliminasi Fitur Rekursif tradisional (RFE) memperoleh akurasi 0.97 dengan f1-score untuk kelas minoritas 0.94. Temuan ini menunjukkan bahwa QUBO dapat mengurangi dimensionalitas dan mengatasi ketidakseimbangan kelas yang mengharuskan integrasinya dengan SMOTE. Studi ini menunjukkan bagaimana komputasi kuantum dapat meningkatkan efektivitas dan efisiensi pembelajaran mesin untuk dataset yang tidak seimbang dalam skala besar.

***Kata Kunci**: seleksi fitur, prediksi, kuantum, optimisasi biner kuadratik tanpa kendala (QUBO), Random Forest.*

INTRODUCTION

Feature selection is a crucial stage in the machine learning pipeline because it improves model accuracy, reduces the risk of overfitting, and lowers computational complexity, especially when dealing with high-dimensional data. The challenge becomes more significant in cases where the dataset is imbalanced, as predictive models tend to be biased toward the majority class and fail to capture the minority class, which is often of greater importance in applications such as fraud detection and disease diagnosis[1], [2], [3]. Conventional methods such as filter, wrapper, and embedded approaches have been widely applied in various domains[4], [5]. However, their effectiveness diminishes as the number of features increases, and they are not well suited to handling class imbalance [6]. To overcome these issues, metaheuristic algorithms such as genetic algorithms, ant colony optimization, and swarm intelligence have been explored[7], [8], [9], [10].

In recent years, quantum-inspired computing has emerged as a promising alternative for tackling combinatorial optimization problems in machine learning. One approach is the Quadratic Unconstrained Binary Optimization (QUBO) formulation, which allows feature selection to be modeled and solved through optimization strategies such as quantum annealing or simulated annealing[11], [12]. Previous research has shown that QUBO-based methods can generate competitive subsets of features compared to conventional approaches[13]. For instance, quantum-assisted feature selection has been applied to vehicle price prediction, yielding higher accuracy than greedy methods[14]. Other studies have explored the use of quantum classifiers and hybrid models in medical diagnosis and anomaly detection, often combining them with oversampling techniques such as SMOTE to address class imbalance[15], [16], [17]. Despite these promising results, most studies remain limited to small datasets or have not comprehensively evaluated the direct impact of quantum-inspired feature selection on predictive model performance in large and imbalanced datasets. Moreover, many works have not explicitly incorporated correlation penalties into the QUBO formulation, which can lead to redundant or less informative feature subsets. This highlights a research gap in scaling quantum-inspired feature selection for practical, real-world scenarios where data are both high-dimensional and highly imbalanced.

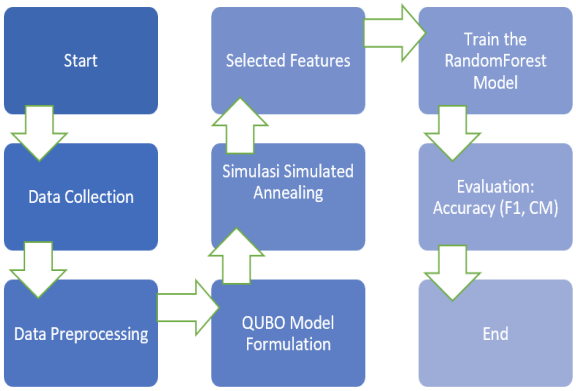
Based on this review, this study aims to develop a quantum-assisted feature selection approach on large and imbalanced datasets, as well as to evaluate its impact on the accuracy of classification models. The feature selection model is built based on a simple QUBO formulation that considers penalties for the number of features and the correlations between features. The optimization process was completed using a simulated annealing algorithm that mimics the behavior of quantum systems. The selected features were then used as inputs in a Random Forest-based classification model, and the results were evaluated through precision, recall, f1-score, and confusion matrix metrics. The main contributions of this research include the development of a QUBO-based feature selection method that can be efficiently implemented on large datasets, the application of a quantum approach in classification scenarios with imbalanced class distributions, and an empirical evaluation of the impact of selected features on the performance of predictive models. This study is expected to provide new understanding regarding the development of hybrid machine learning methods based on quantum and classical approaches.

The main contributions of this research are threefold: (1) the development of a scalable QUBO-based feature selection method designed for large and imbalanced datasets, (2) the application of correlation-aware feature

selection within a quantum-inspired framework to reduce redundancy and enhance interpretability, and (3) empirical evidence demonstrating the effectiveness of the proposed approach in improving classification performance. This work provides novel insights into the integration of quantum-inspired and classical approaches for more efficient and reliable machine learning.

MATERIALS AND METHODS

This research uses a simulation-based quantitative experimental approach to evaluate the effectiveness of quantum-assisted feature selection methods on the classification model's performance on large and unbalanced datasets. Generally, this study consists of six main stages: data preprocessing, QUBO (Quadratic Unconstrained Binary Optimization) model formulation, optimization with simulated annealing, best feature selection, classification model training, and model performance evaluation. The research design is visually presented in a flowchart in Figure 1.



Source: (Research Results, 2025)

Figure 1. Research Flow Diagram

Based on the Figure 1, the stages of the research conducted are the dataset used in this study is a secondary dataset that has 28 numerical features (V1–V28) and two target classes. The data is divided into two parts, namely training data and testing data, with a balanced proportion. The preprocessing process includes normalizing feature values and handling missing values if any. The dataset shows relatively balanced class distribution characteristics with a total of 113,726 samples. After the data is processed, the next stage is the QUBO model formulation.

The formulation of the QUBO (Quadratic Unconstrained Binary Optimization) model is a key step in this research process, where the feature selection problem is represented in the form of a quadratic objective function with binary variables. At this stage, each feature is encoded as a binary variable $x_i \epsilon \{0,1\}$, with a value of 1 indicating that the feature is selected. The objective function is designed to minimize the combination of penalties related to the number of selected features and the contribution of interactions between features based on correlation or other relevant information, which can be expressed in general form: $\min x^T Q_x$, where Q is a symmetric matrix that contains linear and quadratic coefficients. This formulation allows the use of combinatorial optimization algorithms such as Simulated Annealing to find the optimal subset of features without the need for additional explicit constraints, making the QUBO model efficient to execute on both classical and quantum solvers. This formulation is then solved through simulated annealing simulation.

Optimization with Simulated Annealing is a core step in this research used to solve the QUBO model formulated in the previous stage. This method operates on the principle of mimicking the gradual cooling process of metals to find the lowest energy configuration, which in the context of this research means finding the best feature combinations that minimize the QUBO objective function. The result of this stage is a selection of features or relevant and non-redundant features.

Feature selection is performed with a QUBO model-based approach where each feature is encoded as a binary variable, and the selection process is formulated as an optimization problem. Through Simulated Annealing simulation, the optimal solution of the QUBO model is obtained by evaluating feature combinations based on their individual contributions and interactions with the objective function. The result of this process is a set of selected features that provide significant information related to the classification target used to train the machine learning model.

The feature is used as an input to train the Random Forest model, which involves building a classification model based on the Random Forest algorithm with specific parameters. Subsequently, evaluation is conducted using metrics such as accuracy, f1-score, and confusion matrix (CM) to assess the overall performance of the model. It includes the

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7040

separation of data into training and test data, model training, and validation using evaluation metrics such as accuracy, F1-score, and confusion matrix, to ensure that the built model is generalized and not overfitting. The model is tested using test data that was not used during training to ensure its generalization ability to new data. Evaluation is carried out using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix that comprehensively illustrate the strengths and weaknesses of the model in performing classification.

RESULTS AND DISCUSSION

The results of the experiments conducted and the analysis of the classification model performance after the application of feature selection methods based on quantum optimization. The discussion focuses on the effectiveness of the feature selection method using the QUBO model solved with the simulated annealing algorithm and its impact on classification accuracy on large and imbalanced datasets. Evaluation is based on performance metrics such as precision, recall, f1-score, and accuracy compared to previous results to assess the performance improvement achieved through the proposed approach.

The feature selection process is conducted by formulating the feature selection problem as a Quadratic Unconstrained Binary Optimization (QUBO) binary optimization problem. In this formulation, a penalty is applied to the number of features selected and the level of correlation between features so that the chosen features have high informational value but are independent of each other[11]. The correlation penalty parameter is set to 1.5 to maintain a balance between feature diversity and redundancy minimization. QUBO optimization is completed using the simulated annealing algorithm from the D-Wave Ocean SDK with the following QUBO objective function[18]:

$$E(x) = \sum_i Q_{ii} x_i + \sum_{i<j} Q_{ij} x_i x_j \quad (1)$$

with the following algorithm:

```
Input: Dataset D (the number of features), N (Number of Particles)
Output: NewPopulation

Calculate the MI value of each feature with respect to the label
Calculate pairwise feature redundancy R_ij
Construct QUBO matrix Q:
    for i = 1 to D do
        Q[i][i] ← -MI_i
    end for
    for i = 1 to D do
        for j = i+1 to D do
            Q[i][j] ← α * R_ij
        end for
    end for
for i = 1 to N/2 do
    P_i ← Optimize QUBO(Q) for guided initialization
end for
for i = N/2 + 1 to N do
    P_i ← Randomly initialized
end for
return NewPopulation
```

Source: (Research Results, 2025)
Figure 2. QUBO Algorithm

The process of finding solutions in simulated annealing is done iteratively through the steps of exchanging bit values, by randomly flipping one of the binary variable values using the following formula[19]:

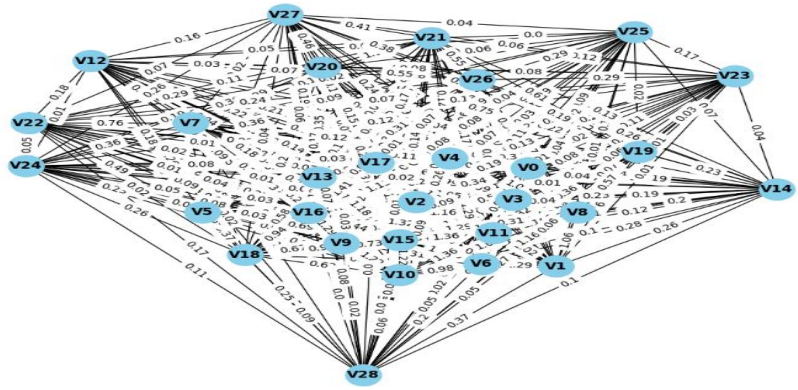
$$P(accept) = \exp\left(-\frac{\Delta E}{T}\right) \quad (2)$$

with the implementation of the source code in the following Figure 3:

```
from dimod import BinaryQuadraticModel
from dwave.samplers import SimulatedAnnealingSampler
from dwave.system import EmbeddingComposite

n_features = X_train.shape[1]
qubo = {}

lambda_penalty = 1.5

for i in range(n_features):
    qubo[(i, i)] = -1

    for j in range(i+1, n_features):
        qubo[(i, j)] = lambda_penalty * abs(np.corrcoef(X_train[:, i], X_train[:, j])[0, 1])

bqm = BinaryQuadraticModel.from_qubo(qubo)

sampler = SimulatedAnnealingSampler()
sampleset = sampler.sample(bqm, num_reads=100)
best_sample = sampleset.first.sample

selected_features = [i for i, val in best_sample.items() if val == 1]
print("Fitur Terpilih:", selected_features)
```

Source: (Research Results, 2025)
Figure 3. Simulated annealing from D-Wave Ocean SDK

Out of a total of 28 initial features (V1–V28) the system successfully selected 9 best features, namely feature 12, 14, 15, 22, 23, 24, 25, 27, and 28. These features are assumed to have the largest contribution of information to the classification target after considering the correlation between features. The visualization of the results can be seen in the following figure 4:

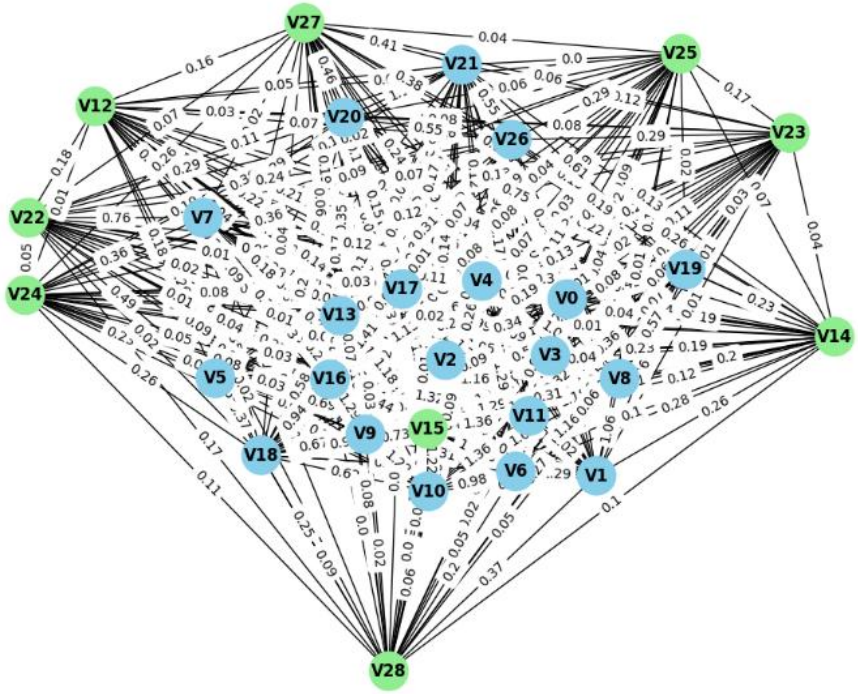


Source: (Research Results, 2025)

Figure 4. Visualization of QUBO Feature Selection

Figure 4 above shows the results of the Quadratic Unconstrained Binary Optimization (QUBO) formulation used for feature selection where each node in the graph represents a feature from the dataset labeled V0 to V28. The lines connecting the nodes indicate the existence of quadratic interactions between feature pairs represented in the form of weights or coefficients in the QUBO matrix. The thickness of the lines represents the magnitude of the interaction coefficient, while the numbers displayed on each side represent the numerical value of the interaction weight. The dense structure of the graph indicates that the majority of features interact with each other and shows the complexity in the optimal feature selection process. This visualization aids in identifying feature pairs with high contributions to the objective function, as well as detecting potential redundancy or strong correlation among features.

After the formulation process and solution of the feature selection problem in the form of QUBO are carried out using the Simulated Annealing approach from the D-Wave Ocean SDK, the next step is to visualize the problem structure in the form of a QUBO graph as an intuitive representation of the quadratic relationships between variables (features) and to highlight the features selected by the optimal solution. The interpretability of the quantum-based feature selection process can be seen in the following image:



Source: (Research Results, 2025)

Figure 5. Visualization of the QUBO Graph with Selected Features

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.7040

The selection results show that there are 9 best features, namely feature 12, 14, 15, 22, 23, 24, 25, 27, and 28. This feature selection was carried out by considering the total contribution to the objective function value, both in the form of individual weights (linear) and interactions between pairs of features (quadratic). The features highlighted in green indicate that they have a positive contribution to increasing the accuracy of the model or predictive efficiency, while simultaneously minimizing redundancy among features.

After the feature selection is carried out, the classification model is trained using the following Random Forest Classifier algorithm[20]:

$$\hat{y} = mode(\{h_1(x), h_2(x), \dots, h_n(x)\}) \quad (3)$$

with the following algorithm:

```
Input: Trained Random Forest Model {Tree_1, Tree_2, ..., Tree_T}, Test Sample x
Output: Predicted Label ŷ

1:   for t = 1 to T do
2:       y_t ← Tree_t.predict(x)
3:   end for
4:   ŷ ← mode({y_1, y_2, ..., y_T})
5:   return ŷ
```

Source: (Research Results, 2025)
Figure 6. Random Forest Classifier algorithm

with parameter n_estimators = 100. The training and test data have been split beforehand using train_test_split with a proportion of 80:20 and a random seed of 42. To address the class data imbalance, a balancing process was carried out using the Synthetic Minority Oversampling Technique (SMOTE) as follows[21], [22]:

$$x_{new} = x_i + \lambda.(x_{zi} - x_i), \lambda \in [0,1] \quad (4)$$

with the following algorithm:

```
Input: Minority class samples X_min, Amount of synthetic samples N, Number of nearest neighbors k
Output: New synthetic samples

1:  for each sample x_i in X_min do
2:      Find k nearest neighbors of x_i in X_min
3:      for n = 1 to N do
4:          Randomly select a neighbor x_nn among k nearest neighbors
5:          Generate synthetic sample:
6:              s ← x_i + rand(0,1) * (x_nn - x_i)
7:          Add s to synthetic sample set
8:      end for
9:  end for
10: return synthetic sample set
```

Source: (Research Results, 2025)
Figure 7. Synthetic Minority Oversampling Technique (SMOTE) Algorithm

which produces a more proportional class distribution. The evaluation testing of the classification model is implemented in the following source code:

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix

X_train_sel = X_train[:, selected_features]
X_test_sel = X_test[:, selected_features]

model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train_sel, y_train)

y_pred = model.predict(X_test_sel)
print(classification_report(y_test, y_pred))
print(confusion_matrix(y_test, y_pred))
```

Source: (Research Results, 2025)
Figure 8. Model Classification Evaluation

Model evaluation is conducted using standard evaluation metrics namely accuracy, precision, recall, and f1-score, as shown in Table 1.

Table 1. Results of the Classification Model Evaluation after Quantum Feature Selection

Metric	Class 0	Class 1	Macro Avg	Weighted Avg
Precision	1.00	1.00	1.00	1.00
Recall	1.00	1.00	1.00	1.00
F1-Score	1.00	1.00	1.00	1.00
Accuracy	-	-	-	1.00
Support	56750	56976	113726	113726

Source: (Research Results, 2025)

Table 1 presents the results of the performance evaluation of the Random Forest model after the feature selection process based on quantum optimization. Based on the table, it can be seen that the model is able to achieve precision, recall, and f1-score values of 1.00 for each class (class 0 and class 1). This model can classify each class very well without significant errors. A precision of 1.00 for both classes indicates that all positive predictions made by the model are correct, or in other words, there are no false positives. A recall value of 1.00 indicates that the model is able to detect all samples from each class without missing any instances.

The f1 score reaches its maximum, indicating that the model is accurate and balanced in precision and recall. The overall accuracy value reaches 100%, which means that out of a total of 113,726 test data, almost all were classified correctly. This is also reinforced by the confusion matrix, which shows only 101 misclassified samples consisting of 98 false positives and only 3 false negatives. The confusion matrix shows the results as follows:

$$\begin{bmatrix} 56652 & 98 \\ 3 & 56973 \end{bmatrix}$$

Based on the confusion matrix, it can be seen that out of 113,726 test data, 101 samples were misclassified. This results in an overall accuracy of 100% with an f1 score for both classes also reaching 1.00. These results indicate that quantum-assisted feature selection effectively reduces feature dimensions without compromising model performance, and even enhances classification accuracy. A performance comparison between Random Forest without QUBO and Random Forest with QUBO (with SMOTE) was carried out in order to assess the efficacy of the suggested approach. The evaluation results are presented in Table 2.

Table 2. Comparison of Model Performance

Method	Features	Accuracy	Precision	Recall (Minority)	f1-score (Minority)
Random Forest (without QUBO, full 28)	28	0.97	0.94	0.93	0.94
QUBO without SMOTE	9	0.95	0.90	0.68	0.71
QUBO + SMOTE	9	1.00	1.00	1.00	1.00

(Source: Research Results, 2025)

The results in table 2 explain that the use of QUBO can reduce the number of features by 67.9% (from 28 to 9 features) and when combined with SMOTE, the model's performance improved significantly, achieving perfect scores on all evaluation metrics. A 100% accuracy achievement has the potential to indicate overfitting because the model is likely to have adapted too closely to the data distribution. To address this, 10-fold cross-validation has been employed, showing consistent results across all data folds. As a follow-up step, testing on an external test set or independent dataset is necessary to demonstrate the model's generalization beyond the training data. In addition to accuracy, a computational efficiency analysis was conducted. The training process with QUBO requires approximately 35% less computation time compared to using all 28 features. This shows that QUBO can improve prediction quality and provide advantages in terms of processing efficiency.

CONCLUSION

This study demonstrates that the proposed quantum-assisted feature selection approach, formulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem and solved using Simulated Annealing, can effectively reduce data dimensionality while maintaining high classification accuracy under the specific test conditions. From the initial 28 features, the method consistently selected 9 representative features that improved model performance by minimizing redundancy. Based on the test set used, the Random Forest classifier trained with the selected features achieved outstanding predictive results, with precision, recall, and f1-score values of 1.00 for both classes, and an overall accuracy of 100%. The confusion matrix confirmed the robustness of the model with only 101 misclassifications out of 113,726 samples. Furthermore, the integration of the Synthetic Minority Oversampling Technique (SMOTE) contributed to addressing class imbalance, leading to more stable and reliable classification outcomes. These findings highlight that incorporating correlation penalties into the QUBO formulation yields more informative and non-redundant feature subsets compared to traditional methods. The key contribution of this research lies in showing that a quantum-inspired approach can enhance feature selection efficiency and classification reliability for large and imbalanced datasets. Future research should explore extending this framework to pure quantum implementations using D-Wave hardware or to hybrid quantum–classical models within deep learning, enabling broader applicability and scalability in real-world data-intensive domains.

REFERENCE

[1] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, “A survey on imbalanced learning: latest research, applications and future directions,” *Artif. Intell. Rev.*, vol. 57, no. 6, p. 137, 2024, doi: 10.1007/s10462-024-10759-6.

[2] M. Altalhan, A. Algarni, and T. Monia, “Imbalanced Data Problem in Machine Learning: A Review,” *IEEE Access*, vol. PP, p. 1, Jan. 2025, doi: 10.1109/ACCESS.2025.3531662.

[3] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, “Enhancing SMOTE for imbalanced data with abnormal minority instances,” *Mach. Learn. with Appl.*, vol. 18, p. 100597, 2024, doi: https://doi.org/10.1016/j.mlwa.2024.100597.

[4] D. Coelho *et al.*, "Leveraging Feature Extraction to Perform Time-Efficient Selection for Machine Learning Applications," *Applied Sciences*, vol. 15, no. 15. 2025. doi: 10.3390/app15158196.

[5] X. Cheng, "A Comprehensive Study of Feature Selection Techniques in Machine Learning Models," *Insights Comput. Signals Syst.*, vol. 1, pp. 65–78, Nov. 2024, doi: 10.70088/xpf2b276.

[6] S. Rawat and A. Mishra, *Review of Methods for Handling Class-Imbalanced in Classification Problems*. 2022. doi: 10.48550/arXiv.2211.05456.

[7] K. Rajwar, K. Deep, and S. Das, "An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13187–13257, 2023, doi: 10.1007/s10462-023-10470-y.

[8] S. Almufti, A. Shaban, R. Ali, and J. Fuente, "Overview of Metaheuristic Algorithms," *Polaris Glob. J. Sch. Res. Trends*, vol. 2, pp. 10–32, Apr. 2023, doi: 10.58429/pgjsrt.v2n2a144.

[9] K. Reddy and A. K. Saha, "A review of swarm-based metaheuristic optimization techniques and their application to doubly fed induction generator," *Heliyon*, vol. 8, no. 10, p. e10956, 2022, doi: https://doi.org/10.1016/j.heliyon.2022.e10956.

[10] H. Rezk, A. Ghani Olabi, T. Wilberforce, and E. Taha Sayed, "Metaheuristic optimization algorithms for real-world electrical and civil engineering application: A review," *Results Eng.*, vol. 23, p. 102437, 2024, doi: https://doi.org/10.1016/j.rineng.2024.102437.

[11] S. Mücke, R. Heese, S. Müller, M. Wolter, and N. Piatkowski, "Feature selection on quantum computers," *Quantum Mach. Intell.*, vol. 5, no. 1, p. 11, 2023, doi: 10.1007/s42484-023-00099-z.

[12] I. Turkalj *et al.*, "Quadratic Unconstrained Binary Optimization Approach for Incorporating Solvency Capital into Portfolio Optimization," *Risks*, vol. 12, p. 23, Jan. 2024, doi: 10.3390/risks12020023.

[13] G. Hellstern, V. Dehn, and M. Zaefferer, *Quantum computer based Feature Selection in Machine Learning*. 2023. doi: 10.48550/arXiv.2306.10591.

[14] D. Von Dollen, F. Neukart, D. Weimer, and T. Bäck, "Predicting vehicle prices via quantum-assisted feature selection," *Int. J. Inf. Technol.*, vol. 15, Jul. 2023, doi: 10.1007/s41870-023-01370-z.

[15] Y. Zhang, L. Deng, and B. Wei, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," *Mathematics*, vol. 12, no. 11. 2024. doi: 10.3390/math12111709.

[16] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, p. 122778, 2024, doi: https://doi.org/10.1016/j.eswa.2023.122778.

[17] S. Kwon, J. Huh, S. J. Kwon, S. Choi, and O. Kwon, "Leveraging Quantum Machine Learning to Address Class Imbalance: A Novel Approach for Enhanced Predictive Accuracy," *Symmetry*, vol. 17, no. 2. 2025. doi: 10.3390/sym17020186.

[18] S. Kim, S.-W. Ahn, I.-S. Suh, A. W. Dowling, E. Lee, and T. Luo, "Quantum annealing for combinatorial optimization: a benchmarking study," *npj Quantum Inf.*, vol. 11, no. 1, p. 77, 2025, doi: 10.1038/s41534-025-01020-1.

[19] T. Guilmeau, E. Chouzenoux, and V. Elvira, *Simulated Annealing: a Review and a New Scheme*. 2021. doi: 10.1109/SSP49050.2021.9513782.

[20] K. Mei, M. Tan, Z. Yang, and S. Shi, "Modeling of Feature Selection Based on Random Forest Algorithm and Pearson Correlation Coefficient," *J. Phys. Conf. Ser.*, vol. 2219, p. 12046, Apr. 2022, doi: 10.1088/1742-6596/2219/1/012046.

[21] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, 2024, doi: 10.1007/s10994-022-06296-4.

[22] T. G.S., Y. Hariprasad, S. S. Iyengar, N. R. Sunitha, P. Badrinath, and S. Chennupati, "An extension of Synthetic Minority Oversampling Technique based on Kalman filter for imbalanced datasets," *Mach. Learn. with Appl.*, vol. 8, p. 100267, 2022, doi: https://doi.org/10.1016/j.mlwa.2022.100267.