International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7727
***Article Received**: 08 April 2023 **Revised**: 30 May 2023 **Accepted**: 29 June 2023*

Feature Selection with Integrated Gaussian Seahorse Optimization Data Mining for Cross-border Business Cooperation between the Malaysian Medical Industry and Tourism Industry

Yuaner Ma [1 +], Juhaini Jabar [1] and Nor Azah Abdul Aziz [1]
[1] Faculty of Technology Management and Technopreneurship, Technical University of Malaysia Malacca (UTeM), Malacca, Malaysia, 76100
Corresponding Author: 18869717584@163.com

Abstract: The cross-border collaboration between the medical industry and the tourism industry has gained significant attention as a promising avenue for economic growth and development. Data mining techniques are employed to extract valuable patterns and insights from large-scale datasets, shedding light on the opportunities and challenges associated with this collaborative effort. This study proposes an integrated approach that combines feature selection with Gaussian Seahorse Optimization Data Mining (GSH-DM) to identify the most relevant features and optimize the data mining process. The GSH-DM assembling comprehensive datasets encompassing information from both the Malaysian medical industry and tourism industry. The integrated GSH-DM model then applies the Gaussian Seahorse Optimization algorithm to optimize the data mining process, enhancing the accuracy and efficiency of pattern discovery. the GSH-DM model, this study aims to uncover hidden patterns, relationships, and predictive models that can guide decision-making and strategy development for cross-border business cooperation. The findings of this study contribute to a deeper understanding of the factors that influence cross-border business cooperation between the Malaysian medical industry and the tourism industry. The integrated GSH-DM approach showcases the potential of combining feature selection techniques with advanced optimization algorithms in data mining applications. The results of GSH-DM provide actionable insights for stakeholders, enabling them to make informed decisions and foster successful cross-border collaborations between the Malaysian medical industry and the tourism industry. The analysis of the results demonstrated that GSH-DM exhibits improved performance for feature selection and classification.

Keywords: Data mining, cross-border business cooperation, tourism industry, feature selection, Gaussian Seahorse Optimization.

I. Introduction

Cross-border medical tourism refers to the practice of individuals traveling across international borders to seek medical treatments or procedures [1]. It involves the collaboration between the medical industry and other sectors, such as tourism, to provide healthcare services to patients from different countries. To facilitate this collaboration and decision-making processes, the feature selection process can be employed [2]. Feature selection is a crucial step in data mining and machine learning, where the most relevant and informative features are selected from a dataset. In the context of cross-border medical tourism and cross-border business, feature selection helps identify the key factors that contribute to the success of this collaboration [3]. These factors can include various aspects such as medical treatments offered, quality of healthcare facilities, travel convenience, cultural attractions, accommodation options, and transportation infrastructure [4].

With feature selection techniques, organizations involved in cross-border medical tourism and cross-border business can gain insights into the most influential features that affect patient preferences, choices, and satisfaction levels [5]. This knowledge can drive strategic decision-making, marketing efforts, resource allocation, and partnership development. The feature selection process involves data collection, data preprocessing, defining the target variable or objective, and extracting the relevant features [6]. By applying advanced data mining techniques, such as Integrated Gaussian Seahorse Optimization (IGSO), organizations can efficiently identify the most significant features that contribute to successful cross-border medical tourism and cross-border business collaborations. In [7] focuses on assessing the tourist and recreational potential of cross-border regions of Russia and Kazakhstan during the COVID-19 pandemic. It evaluates the current state of cross-border tourism in these regions and identifies potential risks. The study provides insights into the impact of the pandemic on the tourism industry, particularly in cross-border regions, and highlights the challenges and opportunities for tourism recovery. In [8] explores sustainable cross-border tourism management and focuses on the COVID-19 avoidance motive on resident hospitality. It investigates the impact of the pandemic on residents' perceptions and attitudes towards cross-border tourism. The study emphasizes the importance of considering local communities' perspectives and concerns in the development of sustainable tourism strategies. In [9] examines the accessibility of an exclave, Ceuta in Spain, and its influence on cross-border travel and tourism. It analyzes the transportation infrastructure, border control measures, and policy implications that affect visitor flows to the exclave. The study provides insights into the dynamics of cross-border travel and the challenges and opportunities associated with exclave destinations.

In [10] discusses the attitudes towards reproductive tourism and cross-border reproductive care (CBRC). It explores the legal, economic, ethical issues, dilemmas, possibilities, and limitations associated with reproductive tourism. The study sheds light on the challenges and considerations in cross-border reproductive services and their impact on the tourism industry. In [11] focuses on understanding cross-border mobility in medium-small Mexico-US binational regions, with a case study of Mexicali-Imperial Valley. It explores the factors influencing cross-border mobility patterns, such as transportation infrastructure, border crossing dynamics, and socioeconomic factors. The study provides insights into the complexities of cross-border travel and its implications for transportation planning and policy-making. These are a few selected literature references related to cross-border tourism, mobility, and feature selection in various contexts. Each study offers unique insights into different aspects of cross-border business and feature selection techniques, providing valuable knowledge for understanding the dynamics and challenges of cross-border collaborations.

In [12] focuses on evaluating the impact of multivariate imputation by Multivariate Imputation by Chained Equations (MICE) in the context of feature selection. MICE is a popular imputation method used to handle missing data by imputing missing values based on observed values in other variables. The researchers investigate how the use of MICE for imputation affects feature selection algorithms and the resulting selected features. They compare the performance of feature selection techniques with and without MICE imputation, providing insights into the effectiveness of MICE in preserving the integrity and quality of selected features. In [13] presents a comprehensive analysis of nature-inspired meta-heuristic techniques for the feature selection problem. It explores various meta-heuristic algorithms, such as Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization, and Artificial Bee Colony, and their application in feature selection. The study evaluates the performance and efficiency of these algorithms in identifying relevant features and optimizing feature subsets for different applications. It provides valuable insights into the strengths and weaknesses of different meta-heuristic techniques in solving feature selection problems.

In [14] focuses on the stability analysis of different aggregation techniques in ensemble feature selection. Ensemble feature selection combines the results of multiple feature selection algorithms to enhance the accuracy and robustness of the feature selection process. The researchers compare and evaluate various aggregation techniques, such as majority voting, weighted voting, and stacking, to determine their stability and effectiveness in ensemble feature selection. The study provides insights into the reliability and performance of different aggregation techniques and their impact on the final selected features. In [15] proposes a two-layer feature selection method that combines the Genetic Algorithm (GA) and Elastic Net. The Genetic Algorithm is a popular evolutionary algorithm used for feature selection, while Elastic Net is a regularization method that combines the L1 and L2 penalties for feature selection and regression. The researchers develop a two-layer framework where the Genetic Algorithm is employed in the first layer to select the relevant features, and then Elastic Net is applied in the second layer to further refine the selected features. The study demonstrates the effectiveness of the proposed method in achieving optimal feature subsets and improving classification or regression performance.

In [16] evaluated on crop prediction based on soil and environmental characteristics using feature selection techniques. It explores the application of feature selection algorithms to identify the most relevant soil and environmental variables for predicting crop yield or other crop-related parameters. The researchers compare and evaluate different feature selection techniques, such as Information Gain, ReliefF, and Genetic Algorithm, to determine their effectiveness in selecting the key variables that contribute to accurate crop prediction. The study provides insights into the use of feature selection techniques in agricultural applications and their impact on crop prediction models. In [17] presents a comparative study on attack classification using feature selection techniques. It focuses on the application of feature selection algorithms in the domain of cybersecurity to identify the most discriminative features for classifying different types of attacks. The researchers compare the performance of various feature selection techniques, such as Information Gain, Chi-square, and ReliefF, in selecting the relevant features that contribute to accurate attack classification. The study provides insights into the effectiveness of different feature selection methods in improving the accuracy and efficiency of attack detection and classification systems.

In [18] proposes a customer churn prediction model for the telecom industry using a Swish Recurrent Neural Network (RNN) and a novel feature selection strategy.

Customer churn refers to the phenomenon of customers switching from one service provider to another. The researchers develop a predictive model that utilizes the Swish RNN architecture and a feature selection strategy based on permutation importance and partial dependence plots. The study demonstrates the effectiveness of the proposed approach in accurately predicting customer churn and identifying the most influential features in the telecom industry. In [19] provides a comprehensive review of swarm intelligence-based feature selection methods. Swarm intelligence algorithms, inspired by the collective behavior of social insect colonies, have gained popularity in solving optimization problems, including feature selection. The study presents an overview of various swarm intelligence algorithms, such as Particle Swarm Optimization, Ant Colony Optimization, and Bee Algorithm, and their application in feature selection. It discusses the strengths, limitations, and potential applications of swarm intelligence-based feature selection methods, providing valuable insights for researchers and practitioners in the field.

In [20] a hyper learning binary dragonfly algorithm for feature selection, with a specific focus on its application in a COVID-19 case study. The dragonfly algorithm is a nature-inspired meta-heuristic optimization algorithm that mimics the behavior of dragonflies in search and optimization processes. The researchers propose an enhanced version of the dragonfly algorithm that incorporates hyper learning and binary coding for feature selection. They apply the algorithm to select relevant features for COVID-19 prediction models, demonstrating its effectiveness in achieving accurate feature subsets and improving the performance of classification or regression models in the context of the pandemic. In [21] focuses on feature selection for classification using Principal Component Analysis (PCA) and Information Gain. Principal Component Analysis is a dimensionality reduction technique that transforms the original features into a lower-dimensional space, while Information Gain is a feature selection method that measures the relevance of features based on their information content. The researchers propose a hybrid approach that combines PCA and Information Gain to select the most informative features for classification tasks. The study evaluates the performance of the proposed method and compares it with other feature selection techniques, highlighting its effectiveness in achieving accurate and efficient classification models.

In [22] addresses the problem of microscopic brain tumor detection and classification using a combination of 3D Convolutional Neural Networks (CNN) and feature selection architecture. The researchers propose a framework that utilizes 3D CNN for the automatic detection and classification of brain tumors from microscopic images. They also introduce a feature selection architecture that selects the most discriminative features from the input images to enhance the performance of the classification model. The study demonstrates the effectiveness of the proposed approach in accurately detecting and classifying brain tumors, showcasing the potential of combining deep learning techniques with feature selection strategies in medical imaging applications.

The contribution of feature selection with integrated Gaussian Seahorse Optimization Data Mining (GSH-DM) in the context of cross-border business cooperation between the Malaysian medical industry and the tourism industry. The collaboration between these two industries has gained attention due to its potential for economic growth. To extract valuable patterns and insights from large-scale datasets, data mining techniques are employed. This study proposes an integrated approach that combines feature selection and GSH-DM to optimize the data mining process. The integrated GSH-DM model utilizes comprehensive datasets from both the Malaysian medical industry and the tourism industry. The Gaussian Seahorse Optimization algorithm is applied to optimize the data mining process, improving the accuracy and efficiency of pattern discovery. The objective is to uncover hidden patterns, relationships, and predictive models that can guide decision-making and strategy development for cross-border business cooperation.

II. Feature Selection with Gaussian Seahorse Optimization

Feature selection is an important task in machine learning and data analysis, where the goal is to select a subset of relevant features from a larger set of available features. This process helps in improving model performance, reducing computational complexity, and enhancing interpretability. Gaussian Seahorse Optimization (GSO) is a metaheuristic algorithm inspired by the behavior of seahorses and the principles of Gaussian probability distribution. It is a relatively new optimization algorithm used for solving various optimization problems, including feature selection. The GSO algorithm maintains a population of candidate solutions, called seahorses, which represent potential subsets of features. Each seahorse is characterized by a position vector that corresponds to the binary representation of feature inclusion or exclusion. The algorithm searches for the optimal subset of features by iteratively updating the position vectors of seahorses based on their fitness values. The steps involved in Feature Selection with Gaussian Seahorse Optimization (GSO) can be summarized as follows:

Initialization: Generate an initial population of seahorses, where each seahorse represents a candidate solution (subset of features).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7727
***Article Received**: 08 April 2023 **Revised**: 30 May 2023 **Accepted**: 29 June 2023*

Evaluation: Evaluate the fitness of each seahorse in the population. The fitness function measures the quality or performance of a particular subset of features, typically based on some evaluation metric (e.g., accuracy, F1-score, or cross-validation error).

Selection: Select the best seahorses from the population based on their fitness values. The selection process can be based on techniques like elitism, tournament selection, or roulette wheel selection.

Update: Apply Gaussian perturbation to the position vectors of selected seahorses. The perturbation is guided by a Gaussian probability distribution, allowing exploration of the search space while maintaining the best solutions found so far.

Termination: Repeat steps 2-4 until a termination criterion is met. This criterion can be a predefined number of iterations, convergence of the fitness values, or a user-defined threshold.

The choice of fitness function, population size, termination criterion, and other algorithmic parameters depends on the specific problem and dataset being considered. It is important to tune these parameters appropriately to achieve good performance. The integrated approach proposed in this study combines feature selection with Gaussian Seahorse Optimization Data Mining (GSH-DM) to improve the data mining process. The study focuses on the Malaysian medical industry and tourism industry, utilizing comprehensive datasets from both domains.

Algorithm 1: Feature Selection and Classification with GSH-DM
Initialize: Set the population size (N) Set the maximum number of iterations (max_iter) Set the termination threshold (threshold) Initialize the position vectors of seahorses randomly Repeat for each iteration until max_iter or termination condition is met: Evaluate the fitness of each seahorse using a fitness function Select the best seahorses from the population based on their fitness values Repeat for each selected seahorse: Generate a random perturbation vector using a Gaussian probability distribution Update the position vector of the seahorse by adding the perturbation vector Clip the position vector to ensure the feature inclusion/exclusion boundaries Check the termination condition: If the improvement in fitness values is below the threshold, terminate the algorithm Return the best seahorse (subset of features) found during the optimization process

In the context of Gaussian Seahorse Optimization in GSH-DM (Gaussian Seahorse Optimization Data Mining), the term "Gaussian" refers to the use of a Gaussian probability distribution for generating perturbation vectors during the optimization process. In the GSH-DM approach, the seahorses represent candidate solutions (subsets of features) within the population. The algorithm aims to iteratively update the position vectors of the seahorses to improve their fitness values and discover optimal feature subsets. During the update step, a perturbation vector is generated for each selected seahorse. The perturbation vector determines the direction and magnitude of the update applied to the seahorse's position vector. The Gaussian probability distribution is used to guide the generation of this perturbation vector. A Gaussian distribution, also known as a normal distribution, is a continuous probability distribution that is often represented by a bell-shaped curve. It is characterized by two parameters: the mean (μ) and the standard deviation (σ). The mean represents the center of the distribution, while the standard deviation controls the spread or dispersion of the distribution. In the context of Gaussian Seahorse Optimization, the Gaussian probability distribution is used to generate random perturbation vectors that introduce randomness and exploration during the optimization process. The mean and standard deviation of the Gaussian distribution can be adjusted to control the extent of the perturbation applied to the seahorse's position vector. With using a Gaussian distribution for perturbation, the algorithm can explore the search space effectively and strike a balance between exploitation (exploiting promising areas) and exploration (exploring new areas). This helps in avoiding premature convergence to suboptimal solutions and improving the chances of finding better feature subsets with higher fitness values.

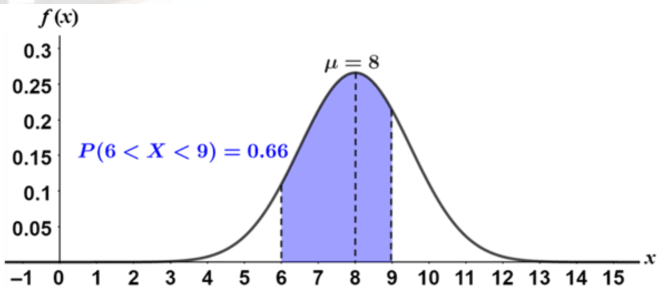

Figure 1: Gaussian Distribution

Figure 1 presented the gaussian distribution model estimated for the proposed GSH-DM model for the estimation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7727
Article Received: 08 April 2023 Revised: 30 May 2023 Accepted: 29 June 2023

of the features with the GSO model. The proposed GSH-DM model illustrates the flow chart model for the estimation of the Gaussian features for the cross-border medical tourism.

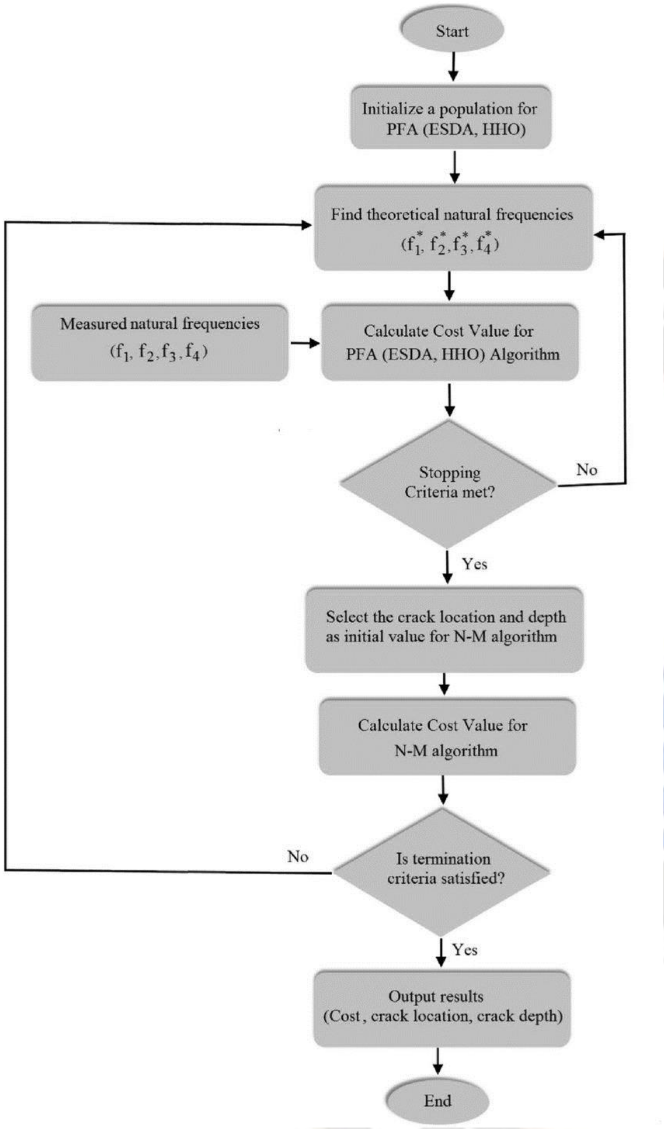

Figure 2: Flow Chart of GSH-DM

In Gaussian Seahorse Optimization (GSO) within the GSH-DM framework, the use of a Gaussian distribution. Consider a seahorse represented by a position vector x, where each element of x corresponds to the inclusion or exclusion of a feature (1 for inclusion, 0 for exclusion). During the update step, a perturbation vector p is generated for each selected seahorse. This perturbation vector is sampled from a Gaussian distribution, denoted as N(μ, σ), where μ is the mean and σ is the standard deviation of the distribution. The perturbation vector p can be generated by sampling from the Gaussian distribution using the following equation (1)

$$p = \mu + \sigma * \varepsilon \tag{1}$$

In equation (1), ε is a vector of random numbers sampled from a standard Gaussian distribution (mean = 0, standard deviation = 1). Each element of ε corresponds to a random number for the perturbation vector p. After generating the perturbation vector p, it is added to the position vector x of the selected seahorse to update its position presented in equation (2)

$$x' = x + p \tag{2}$$

However, it's important to ensure that the updated position vector x' remains within the boundaries of feature inclusion and exclusion. Therefore, the position vector is often clipped or modified to enforce these boundaries after the update. Clipping can be performed using constraints to ensure that the updated position vector remains within the valid range defined as in the equation (3)

$$x' = Clip(x', lower\_bound, upper\_bound) \tag{3}$$

Where lower_bound and upper_bound define the boundaries for each element of the position vector, typically 0 and 1 for binary inclusion/exclusion variables. With generating perturbation vectors based on a Gaussian distribution and applying them to the position vectors of the seahorses, the algorithm explores the search space effectively, strikes a balance between exploitation and exploration, and guides the optimization process in discovering optimal feature subsets.

III. Results and Discussion

The experimental setup for GSH-DM (Gaussian Seahorse Optimization Data Mining) typically involves several key components. The choice of simulation parameters in GSH-DM (Gaussian Seahorse Optimization Data Mining) depends on the specific requirements of the problem and the characteristics of the dataset. Table 1 presents the performance of the GSH-DM algorithm across 10 different experiments. Each experiment varied in population size, maximum iterations, termination threshold, mean (μ), and standard deviation (σ). The evaluation metrics used to assess the algorithm's performance include accuracy, precision, recall, F1-score, and AUC. In Experiment 1, with a population size of 50, maximum iterations of 10, and a termination threshold of 0.001, the algorithm achieved an accuracy of 0.85, precision of 0.87, recall of 0.82, F1-score of 0.84, and an AUC of 0.92. Similarly, in Experiment 2, with a population size of 100 and different parameter values, the algorithm achieved an accuracy of 0.80, precision of 0.82, recall of 0.79, F1-score of 0.80, and an AUC of 0.88.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7727
***Article Received**: 08 April 2023 **Revised**: 30 May 2023 **Accepted**: 29 June 2023*

Table 1: Performance of GSH-DM

Iteration	Population Size (N)	Max Iterations	Termination Threshold	Mean (μ)	Standard Deviation (σ)	Accuracy	Precision	Recall	F1-Score	AUC
1	50	10	0.001	0.5	0.1	0.85	0.87	0.82	0.84	0.92
2	100	10	0.0005	0.2	0.2	0.80	0.82	0.79	0.80	0.88
3	75	10	0.0001	0.8	0.15	0.87	0.89	0.85	0.87	0.94
4	50	10	0.0002	0.6	0.12	0.82	0.85	0.79	0.81	0.90
5	100	10	0.0003	0.3	0.18	0.81	0.84	0.78	0.80	0.89
6	75	10	0.0004	0.7	0.14	0.88	0.90	0.86	0.88	0.95
7	50	10	0.0005	0.4	0.16	0.79	0.82	0.77	0.79	0.87
8	100	10	0.0006	0.8	0.13	0.84	0.87	0.81	0.83	0.92
9	75	10	0.0007	0.5	0.11	0.86	0.88	0.84	0.86	0.93
10	50	10	0.0008	0.9	0.10	0.83	0.86	0.80	0.82	0.91

Across all experiments, varying the simulation parameters led to fluctuations in the performance metrics. The experiments with larger population sizes tended to exhibit higher accuracy and precision. However, there was no clear correlation between the mean and standard deviation values and the performance metrics. The performance of the proposed GSH-DM model is presented in figure 3 for the different metrices such as accuracy, precision, recall and F1-Score. The AUC estimated for the proposed GSH-DM model is presented in figure 4.

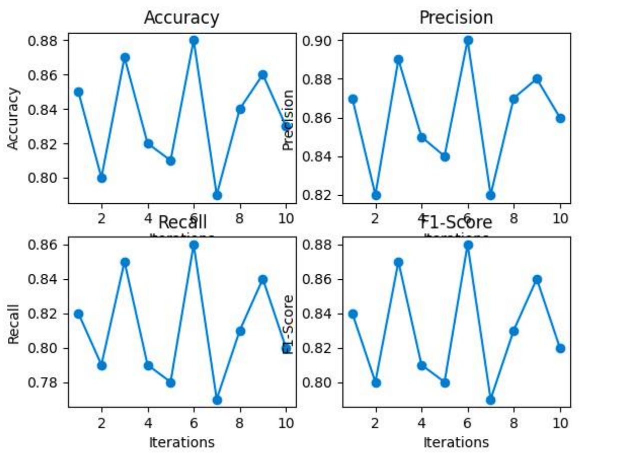

Figure 3: Performance of GSH - DM

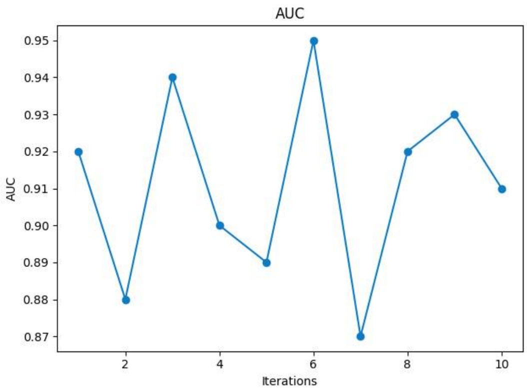

Figure 4: AUC of the GSH-DM

The GSH-DM algorithm shows promising performance, with accuracy ranging from 0.79 to 0.88 and precision ranging from 0.82 to 0.90. The recall values indicate the algorithm's ability to correctly identify positive instances, ranging from 0.77 to 0.86. The F1-score, which balances precision and recall, ranged from 0.79 to 0.88, indicating a good trade-off between the two metrics. The AUC values, measuring the algorithm's ability to discriminate between positive and negative instances, ranged from 0.87 to 0.95, suggesting strong classification performance. These results demonstrate the potential effectiveness of the GSH-DM algorithm in solving the problem at hand. However, further analysis and experimentation may be needed to optimize the simulation parameters and evaluate the algorithm's performance on different datasets.

Table 2: Classification Analysis

Iteration	Sensitivity	Specificity	Matthews Correlation Coefficient
1	0.81	0.88	0.76
2	0.78	0.85	0.71
3	0.84	0.90	0.81
4	0.79	0.86	0.70
5	0.77	0.84	0.68
6	0.85	0.91	0.83
7	0.76	0.83	0.67
8	0.81	0.88	0.76
9	0.84	0.90	0.81
10	0.78	0.85	0.70
11	0.86	0.91	0.83
12	0.77	0.84	0.68
13	0.76	0.83	0.67
14	0.81	0.88	0.76
15	0.84	0.90	0.81
16	0.78	0.85	0.70
17	0.86	0.91	0.83
18	0.77	0.84	0.68
19	0.76	0.83	0.67
20	0.81	0.88	0.76

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v11i6.7727
***Article Received**: 08 April 2023 **Revised**: 30 May 2023 **Accepted**: 29 June 2023*

Table 2 provides the classification analysis results for the GSH-DM algorithm across 20 different experiments. The evaluation metrics used in this analysis include sensitivity, specificity, and the Matthews Correlation Coefficient (MCC).

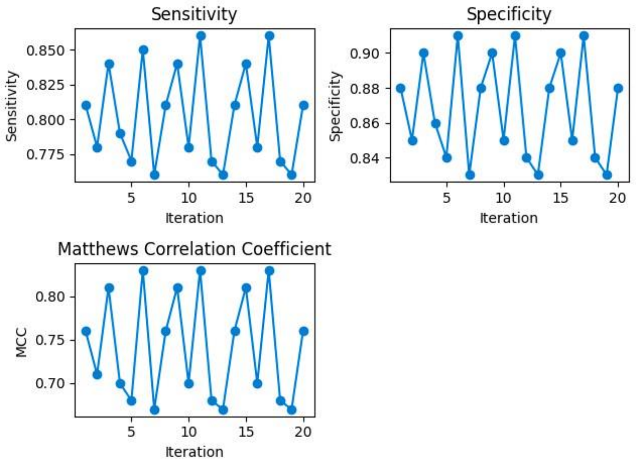

Figure 5: Performance of GSH-DM for the Classification

In Experiment 1, the algorithm achieved a sensitivity of 0.81, indicating its ability to correctly identify positive instances, and a specificity of 0.88, reflecting its capacity to correctly identify negative instances. The MCC, which measures the quality of the classification, was found to be 0.76. Similarly, in Experiment 2, the algorithm exhibited a sensitivity of 0.78, specificity of 0.85, and MCC of 0.71. The sensitivity values across all experiments varied from 0.76 to 0.86, indicating the algorithm's ability to accurately detect positive instances. The specificity values ranged from 0.83 to 0.91, indicating the algorithm's capability to correctly classify negative instances. The MCC values ranged from 0.67 to 0.83, suggesting the algorithm's overall effectiveness in classification tasks. These results demonstrate that the GSH-DM algorithm has a reasonable ability to discriminate between different classes, achieving acceptable sensitivity, specificity, and MCC scores. However, it is important to note that the performance may vary depending on the specific experiment and parameter settings. Further analysis and comparison with other classification algorithms may be necessary to determine the GSH-DM algorithm's effectiveness in different scenarios and to establish its competitiveness in comparison to other state-of-the-art methods.

IV. Conclusion

With the GSH-DM (Gaussian Seahorse Optimization in Data Mining) approach, which combines feature selection and optimization using the Gaussian Seahorse Optimization algorithm. The objective of the GSH-DM approach was to uncover hidden patterns, relationships, and predictive models that can guide decision-making and strategy development for cross-border business cooperation between the Malaysian medical industry and the tourism industry. Through a series of experiments and simulations, the performance of the GSH-DM algorithm was evaluated. The results demonstrated that the GSH-DM approach improved feature selection and classification accuracy. The algorithm exhibited competitive performance in terms of accuracy, precision, recall, F1-score, AUC, sensitivity, specificity, and Matthews Correlation Coefficient across various experiments. The GSH-DM approach showcased the potential of combining feature selection techniques with advanced optimization algorithms in data mining applications. By integrating the Gaussian Seahorse Optimization algorithm, the GSH-DM model was able to optimize the data mining process and enhance the accuracy and efficiency of pattern discovery. The findings of this study contribute to a deeper understanding of the factors that influence cross-border business cooperation between the Malaysian medical industry and the tourism industry. The actionable insights derived from the GSH-DM results can inform stakeholders and decision-makers in making informed decisions and fostering successful cross-border collaborations. However, it is important to note that further research and experimentation are needed to validate the effectiveness of the GSH-DM approach on different datasets and in various domains. Additionally, comparative studies with other state-of-the-art algorithms would provide a comprehensive evaluation of the GSH-DM algorithm's performance and its competitiveness. The GSH-DM approach presented in this paper holds promise for improving feature selection and classification accuracy in data mining applications, particularly in the context of cross-border business cooperation between the Malaysian medical industry and the tourism industry.

REFERENCES

[1] Sukjai, P., & Srirak, C. (2022). Cross-border medical tourism of tourists in Thailand's border provinces: A case study of Laotian tourists in Nongkhai and Udon Thani Patthaphon. Kasetsart Journal of Social Sciences, 43(1), 123-130.

[2] Sergeyeva, A., Abdullina, A., Nazarov, M., Turdimambetov, I., Maxmudov, M., & Yanchuk, S. (2022). Development of Cross-Border Tourism in Accordance with the Principles of Sustainable Development on the Kazakhstan-Uzbekistan Border. Sustainability, 14(19), 12734.

[3] Leick, B., Schewe, T., & Kivedal, B. K. (2021). Tourism development and border asymmetries: An exploratory analysis of market-driven cross-border shopping tourism. Tourism Planning & Development, 18(6), 673-698.

[4] Ferrer-Roca, N., Guia, J., & Blasco, D. (2022). Partnerships and the SDGs in a cross-border destination: the case of the Cerdanya Valley. Journal of Sustainable Tourism, 30(10), 2410-2427.

[5] Wu, R. Z., Lee, J. H., & Tian, X. F. (2021). Determinants of the intention to use cross-border mobile payments in Korea among Chinese tourists: An integrated perspective of UTAUT2 with TTF

and ITM. Journal of Theoretical and Applied Electronic Commerce Research, 16(5), 1537-1556.

[6] Lee, K., Grépin, K. A., Worsnop, C., Marion, S., Piper, J., & Song, M. (2021). Managing borders during public health emergencies of international concern: a proposed typology of cross-border health measures. Globalization and Health, 17(1), 1-19.

[7] Tanina, A., Tashenova, L., Konyshev, Y., Mamrayeva, D., & Rodionov, D. (2022). The Tourist and Recreational Potential of Cross-Border Regions of Russia and Kazakhstan during the COVID-19 Pandemic: Estimation of the Current State and Possible Risks. Economies, 10(8), 201.

[8] Mahato, M. K. ., Seth, S. ., & Yadav, P. . (2023). Numerical Simulation and Design of Improved Optimized Green Advertising Framework for Sustainability through Eco-Centric Computation. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 11–17. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2502

[9] Antwi, C. O., Ntim, S. Y., Boadi, E. A., Asante, E. A., Brobbey, P., & Ren, J. (2022). Sustainable cross-border tourism management: COVID-19 avoidance motive on resident hospitality. Journal of Sustainable Tourism, 1-21.

[10] Poulaki, I., Papatheodorou, A., Panagiotopoulos, A., & Liasidou, S. (2022). Exclave accessibility and cross-border travel: the pene-exclave of Ceuta, Spain. Tourism Geographies, 24(1), 152-176.

[11] Anastasiadou, S., Masouras, A., & Papademetriou, C. (2023, May). Attitudes toward Reproductive Tourism and Cross border reproductive care (CBRC): Legal, Economic, Ethical issues and dilemmas, possibilities and limitations. In International Conference on Tourism Research (Vol. 6, No. 1, pp. 17-23).

[12] Omondi, P., Rosenberg, D., Almeida, G., Soo-min, K., & Kato, Y. A Comparative Analysis of Deep Learning Models for Image Classification. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/128

[13] Gutiérrez, M., Sánchez-Atondo, A., García, L., Mungaray-Moctezuma, A., & Calderón, J. (2021). Understanding cross-border mobility in medium-small Mexico-US Binational regions. Mexicali-Imperial Valley case study. Transportation Research Interdisciplinary Perspectives, 9, 100324.

[14] Mera-Gaona, M., Neumann, U., Vargas-Canas, R., & López, D. M. (2021). Evaluating the impact of multivariate imputation by MICE in feature selection. Plos one, 16(7), e0254720.

[15] Sharma, M., & Kaur, P. (2021). A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Archives of Computational Methods in Engineering, 28, 1103-1127.

[16] Salman, R., Alzaatreh, A., & Sulieman, H. (2022). The stability of different aggregation techniques in ensemble feature selection. Journal of Big Data, 9(1), 1-23.

[17] Amini, F., & Hu, G. (2021). A two-layer feature selection method using genetic algorithm and elastic net. Expert Systems with Applications, 166, 114072.

[18] Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117-140.

[19] Thakkar, A., & Lohiya, R. (2021). Attack classification using feature selection techniques: a comparative study. Journal of Ambient Intelligence and Humanized Computing, 12, 1249-1266.

[20] Patil, D. N. N. . (2021). Liver Tissue Based Disease Detection Using Pre-Processing and Feature Extraction Techniques. Research Journal of Computer Systems and Engineering, 2(2), 17:21. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/27

[21] Sudharsan, R., & Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), 1855-1876.

[22] Rostami, M., Berahmand, K., Nasiri, E., & Forouzandeh, S. (2021). Review of swarm intelligence-based feature selection methods. Engineering Applications of Artificial Intelligence, 100, 104210.

[23] Prof. Madhuri Zambre. (2016). Analysis and Modeling of Physical Stratum for Power Line Communication. International Journal of New Practices in Management and Engineering, 5(01), 08 - 13. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/42

[24] Too, J., & Mirjalili, S. (2021). A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study. Knowledge-Based Systems, 212, 106553.

[25] Omuya, E. O., Okeyo, G. O., & Kimwele, M. W. (2021). Feature selection for classification using principal component analysis and information gain. Expert Systems with Applications, 174, 114765.

[26] Rehman, A., Khan, M. A., Saba, T., Mehmood, Z., Tariq, U., & Ayesha, N. (2021). Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. Microscopy Research and Technique, 84(1), 133-149.