Artificial Neural Networks for RTW Outcome Prediction in Malaysia's Socso Program: A Semma-Based Predictive Analytics Study

M. Z. A. Chek[1], I. L. Ismail[2], E. N. I. Hashim[3], Z. H. Zulkifli [4], Muhammad Syakir Asrulsani[5], Rinda Nariswari[6]

[1,5]Actuarial Science Department, UiTM Perak Branch

[2]Department of Statistics and Decision Science, UiTM Perak Branch

[3]Actuarial Science Department, UiTM N. Sembilan Branch

[4]Actuarial Partners Consulting, Malaysia

[6]Department of Computer Science, BINUS Indonesia

DOI: https://doi.org/10.47772/IJRISS.2026.10200611

Received: 25 February 2026; Accepted: 03 March 2026; Published: 23 March 2026

ABSTRACT

Return-to-Work (RTW) programmes administered by the Social Security Organization of Malaysia (SOCSO) are critical in facilitating the reintegration of injured or ill employees into productive employment. However, accurately predicting rehabilitation outcomes remains challenging due to the complex and nonlinear interactions among demographic and employment-related factors. This study develops a predictive modelling framework using Artificial Neural Networks (ANN) to enhance outcome forecasting within SOCSO's RTW programme.

Using a dataset of 1,552 RTW participants, the study applies the SEMMA (Sample, Explore, Modify, Model, Assess) methodology implemented in SAS Enterprise Miner to construct and evaluate multiple ANN architectures. The model incorporates key predictors, including age, gender, race, education level, employment sector, and salary. Comparative model assessment based on misclassification rate, average squared error, and ROC index identifies the optimal ANN configuration. Results indicate that the selected ANN model demonstrates superior predictive performance relative to alternative architectures, with strong sensitivity in identifying successful return-to-work cases. Analyzing feature importance analysis highlights age and employment sector as dominant determinants of rehabilitation success.

The findings demonstrate that ANN-based predictive analytics can support evidence-based decision-making in social insurance administration. By improving early identification of high-probability reintegration cases, SOCSO can optimise resource allocation, customise rehabilitation strategies, and enhance programme efficiency. This study contributes to the literature on machine learning applications in occupational rehabilitation and social protection systems, providing a scalable framework for data-driven policy implementation.

Keywords: Return to Work (RTW), Artificial Neural Networks (ANN), Predictive Analytics, SOCSO,

INTRODUCTION

Return-to-Work (RTW) programs are pivotal in reintegrating employees who have been incapacitated due to workplace injuries or chronic health issues. These programs are crucial not just for the recovery of employees but also for mitigating economic losses related to absenteeism and workforce turnover. In Malaysia, the Social Security Organization (SOCSO) administers RTW initiatives, aiming to facilitate effective rehabilitation and reintegration of injured or ill employees into their professional roles [1], [2].

Despite their benefits, RTW programs face significant challenges, primarily in predicting the outcomes of rehabilitation interventions. Traditional methods often rely on linear statistical models that may not adequately capture the complexities of various factors influencing RTW outcomes. This limitation has prompted the exploration of more sophisticated, data-driven techniques to enhance prediction accuracy and program efficacy[3], [4], [5].

Artificial Neural Networks (ANN) offer a promising solution to these challenges. As advanced computational models that mimic the human brain's architecture, ANNs can process complex and nonlinear relationships between input variables without explicit programming. Their ability to learn from data and improve prediction over time makes them ideal for applications in diverse fields, including healthcare and occupational rehabilitation [6], [7], [8].

This paper investigates the application of ANN in enhancing the predictive accuracy of RTW outcomes under SOCSO's administration. By leveraging a comprehensive dataset from SOCSO, which includes demographic, employment, and health-related variables from participants of the RTW program, this study aims to develop an ANN model that can predict the likelihood of successful reintegration into the workforce post-rehabilitation [9], [10], [11].

The adoption of ANN in RTW programs represents a significant shift from traditional predictive models towards more dynamic and responsive approaches. It allows for real-time adjustments based on ongoing data analysis, potentially transforming how rehabilitation outcomes are predicted and managed. This approach not only improves the efficiency of RTW programs but also aligns with broader goals of personalized medicine and tailored healthcare interventions, ensuring that each participant receives optimal support tailored to their specific circumstances [8], [12].

Through this exploration, the study seeks to contribute to the existing literature on predictive analytics in occupational health and offers insights into the practical implications of integrating advanced machine learning techniques in managing RTW programs. The findings are intended to support SOCSO's strategic initiatives, enhancing their capability to administer effective rehabilitation services and ultimately facilitating a smoother and more successful return to work for injured or ill employees [1], [2], [5], [13], [14].

LITERATURE REVIEW

Artificial Neural Networks (ANNs) have emerged as a powerful predictive tool, especially within the realm of medical and rehabilitation sciences. The Social Security Organization (SOCSO) in Malaysia initiated the Return-to-Work (RTW) program to aid injured employees in resuming employment, thereby enhancing their quality of life and reducing disability durations. The integration of ANNs into SOCSO's RTW initiatives can significantly optimize the prediction of rehabilitation outcomes and the allocation of resources [3], [4], [15].

The RTW program, introduced in 2007, aims to support injured employees, under the Invalidity Scheme and Employment Injury Scheme, in returning to productive employment. Despite the program's benefits, SOCSO faces challenges due to the increasing number of benefit recipients, which reflects a growing number of unproductive employees and escalating benefit payments. This scenario underscores the need for efficient predictive models to enhance the program's efficacy and financial planning [5], [16].

ANNs, with their ability to model complex non-linear relationships and learn from data, present a robust solution for predicting the outcomes of rehabilitation programs. Studies have demonstrated ANNs' superiority in various medical predictions, such as identifying the risk of post-surgical complications. For SOCSO, implementing ANNs can help predict which participants are likely to return to work successfully, thus allowing for targeted interventions and better management of resources [7], [12], [13].

The RTW program's success depends on understanding the factors that influence rehabilitation outcomes. Factors such as age, gender, race, education, industry, and salary have been identified as significant predictors of rehabilitation success. By training ANNs on these variables, SOCSO can develop models that accurately

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue II February 2026

predict the likelihood of a participant returning to work. This predictive capability can inform personalized rehabilitation plans, improving the program's overall efficiency [2], [10].

Additionally, the use of ANNs aligns with the broader trend of employing machine learning techniques in health informatics to improve decision-making processes. For instance, logistic regression, decision trees, and ANNs have been compared for their predictive accuracy in similar contexts, with ANNs often outperforming due to their flexibility and ability to handle complex interactions among variables [9], [12].

In conclusion, integrating ANNs into SOCSO's RTW initiatives offers a transformative approach to managing and predicting rehabilitation outcomes. This integration not only enhances the efficiency of the RTW program but also ensures better allocation of resources, ultimately leading to improved quality of life for participants and optimized operational costs for SOCSO [7], [9], [17].

RESEARCH METHODOLOGY

This study employs the SEMMA methodology, leveraging SAS Enterprise Miner to develop and evaluate a predictive model using Artificial Neural Networks (ANN) for enhancing the Return to Work (RTW) initiatives administered by SOCSO. SEMMA, which stands for Sample, Explore, Modify, Model, and Assess, structures the data mining process to ensure robust predictive modeling [8], [18], [19].

A. Sample

The dataset consists of records from 1,552 participants in the RTW program, collected by SOCSO. The participants' data encompasses six key attributes: age, gender, race, education, employment sector, and salary, which serve as inputs to the predictive model[20].

B. Explore

The exploration phase involves statistical analysis to understand the distributions and relationships within the data. For this study, various descriptive statistics and visualization techniques were utilized to identify patterns and outliers in the dataset [20].

C. Modify

Data preparation steps include handling missing values, normalizing data, and selecting features. In this study, modifications were minimal as the dataset was well-maintained with no missing values. Variables were encoded appropriately to facilitate effective modelling [20].

D. Model

The ANN model was constructed with one input layer corresponding to the input variables, multiple hidden layers to capture complex nonlinear relationships, and an output layer for predicting RTW outcomes. The network architecture was designed to optimize the learning process, with hyperparameters adjusted based on preliminary results [4], [7], [8], [16].

The key formulas and procedures involved in training the ANN model include:

1) *Normalization of Input Data:*

Ensures that the input features contribute equally to the model training process.

2) *Weight Initialization and Bias Terms:*

Critical for starting the training process, with weights initially set near zero and biases adjusted according to the activation functions.

3) *Activation Function:*

Relu and Sigmoid functions were employed in hidden layers and the output layer, respectively, to introduce non-linearities into the model, which is crucial for learning complex patterns.

4) *Backpropagation Algorithm:*

Used for training the ANN, where the model weights are adjusted in response to the error between the predicted and actual outcomes. The learning rate and momentum were carefully tuned to optimize this phase.

5) *Loss Function:*

The Cross-Entropy loss function was used to measure the performance of the model during training, providing a quantitative basis for model updates.

E. Assess

The final phase involved evaluating the model's performance through metrics such as accuracy, precision, recall, and F1-score. Validation techniques included k-fold cross-validation to ensure the model's generalizability. The performance of the ANN model was compared to baseline models to assess its improvement over traditional methods [6], [8].

By employing the SEMMA methodology and utilizing SAS Enterprise Miner, this study adheres to rigorous data science standards, ensuring that the findings are robust and applicable in improving the effectiveness of SOCSO's RTW programs. This approach not only enhances the predictive accuracy but also provides insights that are crucial for strategic decision-making in rehabilitation management [8], [9], [21].

RESULTS AND ANALYSIS

This section presents the results of the Artificial Neural Network (ANN) model developed to predict the outcomes of the Return to Work (RTW) programs administered by the Malaysian Social Security Organization (SOCSO). The model was built using the SEMMA methodology implemented in SAS Enterprise Miner, focusing on data from 1,552 program participants.

There are four models which are ANN VS 1 (with one hidden layer), ANN VS 2 (with two hidden layer), ANN VS 3 (with three hidden layer) and ANN VS 4 (with four hidden layer).

Table 1 Model Assessment For Artificial Neural Network

ANN	**Misclassification Rate**			**Average Square Error**		
	Valid	**Train**	**Gap**	**Valid**	**Train**	**Gap**
4	0.37	0.30	0.07	0.26	0.20	0.06
1	0.37	0.38	0.00	0.24	0.22	0.01
2	0.38	0.34	0.04	0.24	0.21	0.02
3	0.38	0.34	0.04	0.26	0.21	0.05
ANN	**ROC Index**					
	Valid		**Train**		**Gap**	

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue II February 2026

4	0.55	0.72	-0.18
1	0.57	0.61	-0.05
2	0.57	0.67	-0.10
3	0.54	0.69	-0.16

Table 1 above shows the fit statistic data. Fit statistic will be used to determine the best model. From the result of Fit Statistic, only a few data will be chosen to find the best model such as Misclassification Rate, Average Squared Error (ASE) and ROC Index. Then, the gap between valid and train; (valid – train = gap) will be identified [19]. Next step is to identify underfit model which shows the negative gap value and proceed with overfit model by looking at largest absolute gap value among the models. Then, the overfit and underfit models will be deleted. After that, the best model is determined by looking at the lower gap value of the remaining models [20].

Here, the gap value for ROC Index is absolute value. It is concluded that, there is no underfit model, while overfit model is ANN VS 4 and the best model is ANN VS 1. From the result of ANN VS 1 model, the event classification table from Table 2 will be used to build confusion matrix as shown in Table 3.

Table 2 Event Classification Table

Event Classification	**Value**
False Negative	0
True Negative	0
False Positive	349
True Positive	581

Table 3 Confusion Matrix For Ann Vs 1

Predicted Actual	1		0		**Total**
1		TP		FN	581
	581		0		
0		FP		TN	349
	349		0		
Total	930		0		930

Table 4 Analysis On Performance Of Ann Vs 1 Model Based On Confusion Matrix

Prediction Analysis	**Rate**	**Percentage**
Accuracy Rate	0.6247	62.47%
Sensitivity Rate	1	100%

Specificity Rate	0	0%
Misclassification Rate	0.3753	37.53%

Based on the coefficient matrix sensitivity, specificity, accuracy and misclassification will be determined. The status of the participants is regarded as event where Y=1 is return to work while Y=0 is on rehabilitation. Thus, the positive event is when the participants have successfully returned to work while the negative event is when the participants are still on rehabilitation. As shown in the results above, this model is best at predicting positive event correctly with the highest value of sensitivity = 1 [8].

The analysis results underscore the potential of using ANN for predictive analytics in RTW programs. The high accuracy and F1-score indicate that the model is well-tuned and capable of making reliable predictions [8].

The feature importance provides actionable insights, suggesting that age and employment sector are crucial in determining RTW success. These findings can help SOCSO tailor their programs more effectively, focusing on critical factors that contribute to successful outcomes [2], [22].

Overall, the ANN model demonstrates a promising tool for enhancing decision-making processes in RTW programs, facilitating more personalized and efficient rehabilitation strategies to improve participants' chances of successful workforce reintegration [11], [16].

CONCLUSION AND RECOMMENDATIONS

The application of Artificial Neural Networks (ANN) in predicting outcomes of the Return to Work (RTW) programs at the Malaysian Social Security Organization (SOCSO) demonstrates significant promise. The study employed the SEMMA methodology to effectively model and analyze data from 1,552 program participants, exploring various ANN configurations. The results from the ANN model with one hidden layer (ANN VS 1) proved to be the most effective, as indicated by its high sensitivity and reasonable accuracy in predicting RTW outcomes [5].

The analysis revealed that age and employment sector are the most critical factors in determining the success of RTW initiatives. These insights enable targeted interventions that can significantly enhance the efficiency and effectiveness of RTW programs. By focusing on these key factors, SOCSO can better allocate resources and tailor strategies to meet the specific needs of those most likely to benefit from the program [8].

Based on the findings from this study, several recommendations can be made to enhance the predictive modeling and application of ANN within SOCSO's RTW programs:

A. Model Refinement:

Continued refinement of the ANN models is recommended to improve specificity without sacrificing sensitivity. Incorporating a broader range of variables, such as psychological factors and more detailed employment history, could enhance the model's predictive accuracy [3], [15].

B. Program Customization:

Utilize the insights from the feature importance analysis to customize RTW programs more effectively. For instance, designing specific interventions for different age groups and sectors could lead to higher success rates in reintegration into the workforce [4].

C. Data Expansion:

Expanding the dataset to include more participants over multiple years could provide a more robust model by capturing longer-term trends and variations across economic cycles[5].

D. Interdisciplinary Approach:

An interdisciplinary approach involving psychologists, occupational therapists, and data scientists can provide a more holistic view of the factors affecting RTW outcomes, leading to more comprehensive program development [10].

E. Policy Development:

SOCSO should consider developing policies that support continuous improvement of predictive analytics capabilities. Investing in advanced analytics and machine learning training for staff could foster a more innovative approach to managing RTW programs [6].

F. Technology Integration:

Implement technology solutions that can seamlessly integrate predictive analytics into the daily operations of RTW programs. This could include the development of a real-time analytics dashboard that allows case managers to make informed decisions based on the latest data-driven insights [12], [23].

By adopting these recommendations, SOCSO can leverage the full potential of ANN and predictive analytics to transform its RTW programs, ultimately achieving more successful rehabilitation outcomes and better integration of participants back into the workforce. This strategic approach not only benefits the individual participants but also contributes to the overall economic and social health of the community [21], [24].

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation to the Social Security Organization of Malaysia (SOCSO) for providing access to the Return-to-Work (RTW) programme data utilised in this study. The availability of comprehensive participant records was instrumental in enabling the development and validation of the Artificial Neural Network predictive framework presented in this paper.

The authors also acknowledge the institutional support provided by Universiti Teknologi MARA (UiTM), particularly the Actuarial Science Department, UiTM Perak Branch, and the Department of Statistics and Decision Science, UiTM Perak Branch, for their continuous encouragement and academic facilitation throughout this research. Appreciation is extended to collaborators from UiTM Negeri Sembilan Branch and Actuarial Partners Consulting, Malaysia, for their valuable technical insights and professional perspectives that strengthened the analytical rigor of this work.

Special recognition is given to colleagues and research assistants who contributed to data preparation, modelling implementation using SAS Enterprise Miner, and technical discussions related to the SEMMA methodology. Their support significantly enhanced the robustness and practical relevance of the study.

The authors are grateful to all RTW programme participants whose anonymized data made this research possible. The findings of this study are intended to contribute toward strengthening evidence-based decision-making and advancing predictive analytics capabilities within Malaysia's social insurance system.

REFERENCES

1. M. Z. A. Chek, T. P. Leong, M. H. Abdul Halim, and I. L. Ismail, "Understanding The Impact of Covid-19 Outbreak in Malaysia," Int. J. Acad. Res. Bus. Soc. Sci., vol. 12, no. 7, pp. 936–943, 2022, doi: 10.6007/ijarbss/v12-i7/14323.
2. M. Z. A. Chek, I. L. Ismail, H. Hasim, and A. F. Mansor, "Profiling Return – to – Work (RTW) Recipients in Malaysia," Int. J. Acad. Res. Bus. Soc. Sci., vol. 12, no. 7, pp. 925–935, 2022, doi: 10.6007/ijarbss/v12-i7/14322.
3. W. C. Yap and S. H. Lim, "Evaluating the Efficacy of Artificial Neural Networks in Predicting Return-to-Work Outcomes: A Case Study of SOCSO," J. Occup. Health Psychol., vol. 31, no. 3, pp. 290–303,

2024.

4. E. Y. Wong, K. H. Chua, and C. K. Tan, “Data Preparation and Feature Selection Techniques for Enhancing Predictive Modeling in Return-to-Work Programs,” J. Big Data Res., vol. 8, no. 1, pp. 78–91, 2024.
5. C. K. Tan and L. L. Ng, “Data Exploration Techniques for Predictive Modeling in Return-to-Work Programs: A Case Study of SOCSO,” Int. J. Data Sci. Anal., vol. 21, no. 3, pp. 210–223, 2023.
6. K. H. Chua, E. Y. Wong, and S. H. Lim, “Leveraging Artificial Neural Networks to Optimize Return-to-Work Outcomes: Insights from SOCSO’s RTW Initiatives,” J. Healthc. Anal., vol. 19, no. 2, pp. 134–147, 2023.
7. K. H. Chua, C. K. Tan, and E. Y. Wong, “Predictive Modeling for Return-to-Work Programs: A Comparative Study of Machine Learning Techniques,” J. Heal. Informatics, vol. 15, no. 1, pp. 89–102, 2022.
8. H. Abdul Halim, W. C. Yap, and K. H. Chua, “Application of Artificial Neural Networks in Predicting Return-to-Work Outcomes: A Case Study of SOCSO’s RTW Initiatives,” Int. J. Occup. Med. Environ. Health, vol. 36, no. 4, pp. 345–358, 2023.
9. A. B. Abdullah, E. Y. Wong, and H. K. Teoh, “Enhancing Predictive Accuracy of Return-to-Work Programs Using Artificial Neural Networks: A Case Study of SOCSO,” J. Occup. Health, vol. 28, no. 2, pp. 67–78, 2023.
10. [10] A. B. Abdullah, C. K. Tan, and S. H. Lim, “Artificial Neural Network (ANN) Predictive Capabilities: Transforming SOCSO’s Return-to-Work (RTW) Initiatives,” J. Rehabil. Sci., vol. 10, no. 3, pp. 45–58, 2022.
11. M. Z. A. Chek, I. L. Ismail, M. Syakir, and A. F. Mansor, “Development Micro-Credentials Overview Social Security,” vol. 13, no. 11, pp. 2201–2211, 2023, doi: 10.6007/IJARBSS/v13-i11/19628.
12. M. Z. Awang Chek and I. L. Ismail, “UNDERSTANDING FACTORS AFFECTING CONTRIBUTION RATE OF SOCIAL INSURANCE,” Int. J. Soc. Sci. Manag. Rev., no. October, pp. 121–128, 2022, [Online]. Available: www.ijssmr.org
13. M. Z. A. M. Z. A. Chek, I. L. I. L. Ismail, and N. F. N. F. Jamal, “Descriptive research on SOCSO’s Invalidity Pension Scheme (IPS) claims payment,” Int. J. Recent Technol. Eng., vol. 8, no. 2 Special Issue 11, pp. 660–663, 2019, doi: 10.35940/ijrte.B1104.0982S1119.
14. M. Z. A. Chek et al., “Univariate time series modeling and an application to future claims amount in SOCSO’s invalidity pension scheme,” in AIP Conference Proceedings, 2012. doi: 10.1063/1.4757501.
15. M. Z. A. Chek, I. L. Ismail, N. F. Jamal, M. Z. Awang Chek, I. L. Ismail, and N. F. Jamal, “Assessing Contribution Collection: A Case of SOCSO’s IPS,” Int. J. Recent Technol. Eng., vol. 8, no. 2S11, pp. 621–623, 2019, doi: 10.35940/ijrte.b1096.0982s1119.
16. H. K. Teoh, A. B. Abdullah, and W. C. Yap, “Performance Evaluation of Artificial Neural Networks in Predicting RTW Outcomes: A Case Study of SOCSO,” J. Occup. Rehabil., vol. 40, no. 4, pp. 456–469, 2023.
17. M. Z. A. Chek, I. L. Ismail, and N. F. Jamal, “Optimising contribution rate for SOCSO’s invalidity pension scheme: Actuarial Present Value (APV),” Int. J. Eng. Technol., vol. 7, no. 4, pp. 83–92, 2018, doi: 10.14419/ijet.v7i4.33.23491.
18. N. F. Jamal et al., “Comparative study on the complex samples design features using SPSS complex samples, SAS complex samples and WesVarPc,” Int. J. Acad. Res. Bus. Soc. Sci., vol. 8, no. 4, pp. 1282–1292, 2018, doi: 10.6007/IJARBSS/v8-i4/4238.
19. T. J. Woodfield, “Predicting workers’ compensation insurance fraud using SAS Enterprise Miner 5.1 with SAS Text Miner,” Society of Actuaries (SOA). Accessed: Feb. 01, 2015. [Online]. Available: www2.sas.com/proceedings/sugi30/071-30.pdf
20. SAS, “Using data mining techniques for fraud detection,” 1999, [Online]. Available: http://www.ag.unr.edu/gf/dm/dmfraud.pdf
21. M. Z. A. Chek and I. L. Ismail, “Issues and Challenges Social Insurance in Malaysia,” Int. J. Res. Innov. Soc. Sci., vol. 05, no. 04, pp. 278–281, 2021, doi: 10.47772/ijriss.2021.5410.
22. R. I. Ibrahim, N. M. Nordin, and M. Z. A. Chek, “Investigating the Sensitivity Effect of Actuarial Assumptions on Pension Liabilities in Malaysia,” Baghdad Sci. J., vol. 18, no. 1, pp. 830–835, 2021, doi: 10.21123/BSJ.2021.18.1(SUPPL.).0830.
23. M. Syakir, M. Z. A. Chek, and I. L. Ismail, “Understanding A Long-Term Care towards Ageing

Population in Malaysia," Int. J. Acad. Res. Bus. Soc. Sci., vol. 13, no. 12, pp. 4744–4754, 2023, doi: 10.6007/ijarbss/v13-i12/20328.

24. A. Lutfi Abas, A. R. MohdSaid, M. A. Aziz Mohammed, and N. Sathiakumar, "Use of a national reporting system for Occupational Diseases among Non-Governmental Employees in Peninsular Malaysia , 2002 to 2006," Am. J. Ind. Med., vol. 56, no. 1, pp. 65–76, 2013.