Full Paper

QUALITY PREDICTION WITH NEURAL NETWORK TECHNIQUES FOR POLYPROPYLENE PRODUCTION VIA THE SPHERIPOL PROCESS

Chuan Heng Tan, Khairiyah Mohd-Yusof*, Sharifah Rafidah Wan Alwi

School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

*Corresponding author
khairiyah@cheme.utm.my

Graphical abstract

1. 1st loop hydrogen concentration
2. 2nd loop hydrogen concentration
3. Monomer to donor ratio
4. Total C3 flow
5. Prepoly reactor heat
6. 1st loop reaction heat
7. 2nd loop reaction heat
8. Calculated total production rate
9. Reactor pressure
10. 1st loop reactor pump kW
11. 2nd loop reactor pump kW
12. 1st loop reactor density
13. 2nd loop reactor density
14. Donor flow
15. 1st loop reactor temperature
16. 2nd loop reactor temperature

log sigmoid transfer function

Pure linear transfer function

MI

Abstract

In the polypropylene (PP) industry, melt index (MI) is the most important quality variable. Different grades of PP have their specific range of MI. Accurate prediction of MI is essential for efficient monitoring and off-grade reduction. Artificial Neural Network (ANN) models are proposed as the technique for MI estimation. It has powerful adaptive capabilities in response to nonlinear behaviour. In this research, ANN models for PP polymerization to predict the MI based on reactor parameters were developed. Three types of ANN models, the single hidden layer ANN (shallow ANN), stacked neural network (SNN) and deep learning are compared. The simulation results show that deep learning can perform better than shallow ANN and SNN by considering the accuracy of the prediction and detection of process fluctuation. All three model have proven that ANN are able to perform non-linear function approximation. Thus, ANN models are effective for supporting MI prediction such as for soft-sensors and process optimization in the polymer industry.

Keywords: Artificial Neural Network, Deep Learning, Melt Index, Polypropylene, Stacked Neural Network

Abstrak

Dalam industri polipropilena (PP), indeks leburan (MI) adalah kualiti yang paling penting. Ramalan MI yang tepat adalah penting untuk pemantauan yang cekap dan pengurangan produk tidak berkualiti. Pemodelan rangkaian saraf dicadangkan sebagai teknik untuk anggaran MI. Ia mempunyai keupayaan penyesuaian yang kuat dalam pemodelan kelakuan tidak linear. Matlamat penyelidikan ini adalah untuk membangunkan model rangkaian saraf untuk meramalkan MI berdasarkan parameter reaktor. Jenis rangkaian saraf yang berbeza seperti rangkaian saraf tiruan cetek (*shallow* ANN), rangkaian saraf bertindan (SNN) dan '*deep learning*' akan dibandingkan. Keputusan simulasi menunjukkan bahawa '*deep learning*' lebih baik daripada *shallow* ANN dan SNN dengan mengambil kira ketepatan ramalan dan pengesanan ketidakstabilan proses. Namun, ketiga-tiga pemodelan membuktikan bahawa rangkaian saraf mampu melakukan pemodelan bukan linear dan model rangkaian saraf adalah alat analisis yang berkesan yang boleh digunakan sebagai alat ramalan MI untuk pengoptimuman proses dalam industri polimer.

Kata kunci: *Deep learning,* indeks leburan, rangkaian saraf bertindan, rangkaian saraf tiruan, polipropilena

1.0 INTRODUCTION

Polypropylene (PP) is a versatile thermoplastic material with various applications, including food packaging, appliances, injection moulded parts for automobiles, textiles, labelling and toy making. It has intrinsic properties of good tensile strength, inert towards acid, high stiffness, and good processability. The polypropylene market is expected to have a compound annual growth rate (CAGR) of more than 6% expected over the forecast period (2021-2026).

Different grades of PP have their specific range of melt index (MI) that will be changed during product grade transition. Melt index (MI) is the weight of the polymer in grams extruded in 10 min under the application of a dead weight through a melt flow apparatus. It is the most widely used specification in the plastics industry to differentiate between different polymer grades. It has long been used by PP manufacturers to define the best end application for a particular grade of polymer. Material exceeding or lower than the intended MI can have a negative impact on the quality of the end product and receive customers' complaints and rejects. As a result, it is critical to guarantee that MI is within the specification.

Due to the high demand for different PP grades, the frequency of grade transition reaches almost four transitions per week. Higher transition frequency means the MI of PP produced needs to be adjusted frequently. Therefore, fast and efficient quality analysis is important to reduce the off-grade quantity. Yet, the problem is that MI can only be evaluated offline with the analytical procedure in a quality laboratory. The procedure is costly and also time-consuming, requiring between 2 to 4 hours [1]. Therefore, there is a need to come out with methods to estimate the product quality in real-time.

Product quality prediction is a much more complex issue for a polymerisation process compared to conventional short chain reactions because the chemical, physical, mechanical, thermal, and rheological properties of polymers are strongly influenced by its molecular and morphological properties [2]. Due to the high complexity and non-linearity of the PP process, first principle modelling contains many uncertainties resulting in difficulties in getting accurate prediction of PP MI. Therefore, instead of creating a first-principles driven predictor, a data-driven predictor is a better choice. In this work, artificial neural network (ANN) modelling is the data-driven model chosen for the MI prediction. ANN uses the past relationship between the process data and MI results to predict future MI values. The approach of using ANN models has become more popular recently due rapidly increasing computing power and the ability of ANN to simulate a wide variety of complex engineering and scientific problems.

ANN has been utilized in various ways in the PP industry. Li and Liu [3] presented an optimized radial basis function (RBF) neural network to predict the MI of polypropylene produced in the Hypol process. The RBF neural network is optimized by modified Particle Swan Optimization – Simulated Annealing (MPSO-SA) algorithm. Zhang and Liu [4] proposed a fuzzy neural network (FNN) and introduce support vector regression (SVR) for parameter tuning, where the output function is transformed into an SVR based optimization problem. Zhang *et al.* [5] research proposed a soft sensor based on aggregated RBF neural networks trained with chaotic theory. Xu and Liu [1] proposed a novel predictive model of MI, fuzzy functions with dynamic fuzzy neural networks (FF-D-FNN). FF-D-FNN is the combination of the universal approximation property of fuzzy functions and the parsimonious structure with high performance of dynamic fuzzy neural networks (D-FNN). Zhang *et al.* [6] researched MI estimation on Hypol process PP using a novel soft sensor based on extreme learning machine (ELM) and modified gravitational search algorithm (MGSA). Jumari and Mohd-Yusof [7] developed models for soft sensors to measure MI in industrial polypropylene loop reactors using artificial neural network (ANN) models, serial hybrid neural network (HNN) models and stacked neural network (SNN) models.

In recent years, deep learning has become more popular because the problem of vanishing gradient has been solved by the ReLU activation function. Deep learning networks with ReLU are more easily optimized because gradient can flow when the input to the ReLU function is positive. ReLU networks can perform favourable generalization behaviour in several practical applications and benchmark tests [8]. Ramachandran *et al.* [9] mentioned that the ReLU activation function is the most successful and widely-used activation function in deep networks. The abundance of automated acquired training data and higher computational power also facilitate the success of deep network training. Deep learning is widely used in the chemical industries, although currently none has been found in modeling polymer processes. Examples on the use of deep learning in the chemical industries include reconstructions for complex chemical processes [10], chemical production scheduling [11], fault detection and diagnosis in chemical processes [12, 13], analytical chemistry [14], optimizing textile chemical process [15] and risk assessment of occupational exposure to toxic chemicals in coal mine workplaces [16].

This work studies the effectiveness of deep learning in predicting the MI of polypropylene using the data set collected from the Spheripol Process Loop Reactors. The other two models created are shallow Artificial Neural Network (ANN) and Stacked Neural Network (SNN). The accuracy of the prediction is evaluated and discussed.

2.0 PROCESS DESCRIPTION

The Spheripol process consists of three liquid-phase loop reactors: the baby loop for pre-polymerisation

and two main polymerization reactors. In order to develop an effective model to predict MI, variables from the process were identified to construct the model for MI prediction. Figure 1 shows the reactor line-up of the PP Spheripol Process.
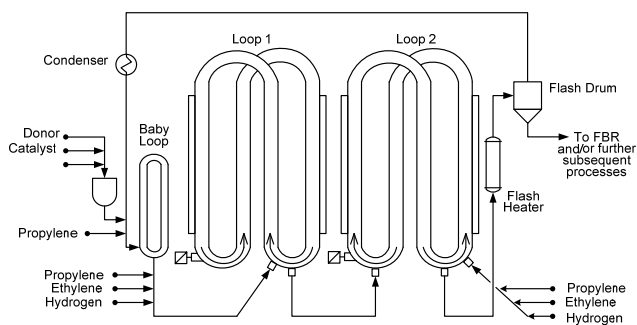

Figure 1 Spheripol Process

Propylene is the main feed because it is the monomer for PP. There are two types of propylene being fed to the reactor, fresh propylene and recycled propylene. Both of them are stored in a big drum (propylene feed drum) before being pumped to the reactor. Therefore, the *propylene flow* from the propylene feed drum is taken as one of the variables.

Hydrogen gas is the chain transfer agent injected into the loop reactors to initiate chain termination. It controls the length of the polymer and its molecular weights [17]. Hydrogen is the main component that alter the MI of the polymer. Another chemical added is the donor. Donor is a stereo regulating agent that affects the tacticity of the polymer [18], affecting the molecular weight of the polymer. Therefore, both *hydrogen concentration* and *donor flow rate* are taken as the input variables.

The next input variable for the ANN model is the *heat generated from the reactions in the loop reactors*. This variable is calculated using the temperature difference from the reactor jacket water. The '*density*' (proportion of solid polypropylene over the liquid propylene) inside the reactor is also taken into account. *Density* inside the reactors is directly proportional to the residence time of the polymer inside the reactor. High *density* means less propylene flows into the reactor, enabling higher residence time for the polymer inside the loop reactors before being replaced by the new incoming propylene flow. Both of these variables cause MI changes in the polymer.

Instead of taking the catalyst flow rate, the *calculated production rate* is taken as one of the variables. Although production rate is controlled by the catalyst flow rate, it is not suitable to take catalyst flow rate as a variable. The reason is the concentration of the catalyst cannot be made constant because it is manually mixed with white mineral oil and grease, causing variations in the concentration. Thus, it is better to take the *calculated production rate* compared to the catalyst flow rate. Catalyst type is kept constant throughout the research.

Pressure and *temperature* of the reactors are also taken into the account. Although *temperature* and *pressure* seldom deviate from their controlled value, MI is affected if they deviate from the normal value because polymerization can take place in non-favourable conditions. Table 1 shows the summary of the input and output variables used to train the neural networks.

Table 1 Neural networks input and output variables

Type	Variable	Justification
Input 1	1st loop hydrogen concentration	Hydrogen (chain transfer agent) concentration to control the length of the polymer chain.
Input 2	2nd loop hydrogen concentration	
Input 3	Propylene to donor ratio	Flow ratio of propylene to donor
Input 4	Propylene flow rate	Feed rate of monomer for polymerisation reaction
Input 5	Prepoly reaction heat	Calculated heat produced by the reaction of respective reactors
Input 6	1st loop reaction heat	
Input 7	2nd loop reaction heat	
Input 8	Production rate	Calculated PP production rate
Input 9	Reactor Pressure	Pressure of the reactors (Prepoly, 1st and 2nd loop reactors)
Input 10	1st loop reactor pump kW	Measurement of the reactor density by using 1st and 2nd loop reactor pump kW
Input 11	2nd loop reactor pump kW	
Input 12	1st loop reactor density	Measurement of the reactor density by using 1st and 2nd loop reactor gamma ray detection
Input 13	2nd loop reactor density	
Input 14	Donor flow rate	Flow rate of the stereoregulating agent to control atactic content of polymer
Input 15	1st loop reactor temperature	The temperature of the reactor (1st loop reactor and 2nd loop reactor)
Input 16	2nd loop reactor temperature	
Output	MI of polymer produced from reactor	Polymer quality specification of desired polymer grade

3.0 METHODOLOGY

Data collected for the neural networks must be sufficient and significant enough to cover the possible known variation in the problem domain [19]. The MI of the data collected is within the range of 1 to 17 g/10mins, with the input variables that correspond with the MI range. Collected data is divided into two distinct sets for training and validation.

3.1 Development, Simulation and Validation of Shallow ANN Models

The shallow ANN models with single hidden layer were developed to predict MI using reactor parameters collected. For the shallow ANN topology, the hidden layer is log sigmoid transfer function and the output layer was pure linear transfer function. Based on Hagan *et al.* [20], this type of topology can be trained to estimate most of the functions well. The model was trained using "nntools" function in MATLAB. Feed-Forward Back Propagation (FFBP) was used as the network architecture and Levenberg-Marquardt as the optimization algorithm for fast convergence of the weights and biases [7].

Data normalization is used before the training of the neural networks. The normalization formula to calculate *X normalized* is as follows.

$$X \ normalized = \frac{X - X \ minimum}{X \ maximum - X \ minimum} \qquad (1)$$

where *X* is Input data. *X minimum* is the lowest value within the input data range. *X maximum* is the highest value within the input data range.

Figure 2 shows the input variables and the output variable of the shallow ANN modelling.

1. 1st loop hydrogen concentration
2. 2nd loop hydrogen concentration
3. Propylene to donor ratio
4. Propylene flow rate
5. Prepoly reactor heat
6. 1st loop reaction heat
7. 2nd loop reaction heat
8. Production rate
9. Reactor pressure
10. 1st loop reactor pump kW
11. 2nd loop reactor pump kW
12. 1st loop reactor density
13. 2nd loop reactor density
14. Donor flow rate
15. 1st loop reactor temperature
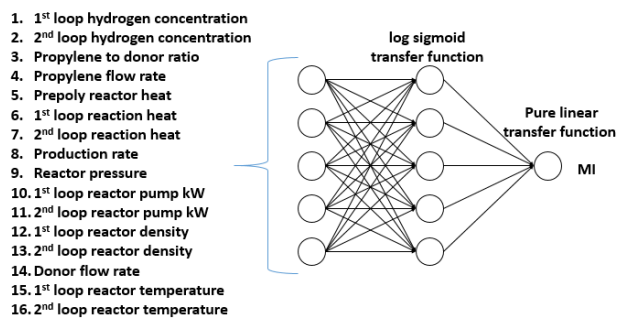16. 2nd loop reactor temperature

Figure 2 Shallow ANN topology

To determine the best shallow ANN model, the number of hidden nodes is systematically altered between 16 ~ 48 nodes during training. Prediction from each shallow ANN model with different numbers of nodes is cross-validated with the unseen data outside the training inputs. The number of nodes with the best prediction is presented.

3.2 Development, Simulation and Validation of SNN Models

Stacked neural network (SNN) was proposed to improve the accuracy of the model. Model accuracy is improved by combining several neural networks. Figure 3 shows the model for stacked neural networks.
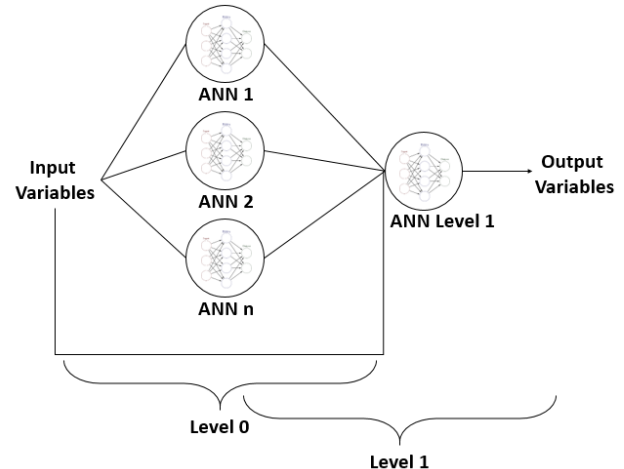

Figure 3 Stack neural networks topology

In order to produce SNN models, four different shallow ANN models (ANN 1, ANN 2, ANN 3 and ANN 4) are trained with different numbers of hidden layer nodes. ANN 1 and ANN 2 were trained by using the MI range of 1 – 5 and ANN 3 and ANN 4 were trained using MI range between 11 – 17. Subsequently, all the MI results predicted from each ANN are served as the training input data for the last shallow ANN (ANN Level 1), along with the original input variables. The reason for choosing MI ranges 1 – 5 and 11 – 17 is due to there are commercial products within these ranges and more input data are available for training and testing within these ranges.

For ANN 1 and ANN 2, the number of hidden nodes is systematically altered between 16 ~ 48 nodes and re-trained. The two best neural networks models which have the best prediction are chosen as ANN 1 and ANN 2. The method is then repeated for ANN 3 and ANN 4 for MI range 11 – 17. Table 2 shows the specification of the ANN n for the SNN models.

Table 2 Specification of ANN n

ANN n	Specification
ANN 1	ANN with 18 nodes trained for MI 1-5
ANN 2	ANN with 24 nodes trained for MI 1-5
ANN 3	ANN with 20 nodes trained for MI 11 - 17
ANN 4	ANN with 22 nodes trained for MI 11 - 17

Lastly, the number of hidden nodes for ANN Level 1 is systematically altered between 20 ~ 40 nodes and re-trained. Prediction from each SNN with different numbers of nodes is cross-validated with the unseen

data outside the training inputs. The number of nodes with the best prediction will be presented.

3.3　Development, Simulation and Validation of Deep Learning Models

Deep learning was proposed to solve the problem of vanishing gradients and overfitting. Different from shallow ANN, Rectified Linear Unit (ReLU) transfer function was chosen instead of the log sigmoid transfer function. ReLU transfer function is used to solve the problem of vanishing gradients, which arises when applying the chain rule in layered networks of sigmoidal units. The deep learning topology that is focused on in this study is deep neural networks. Figure 4 shows the deep neural networks topology.
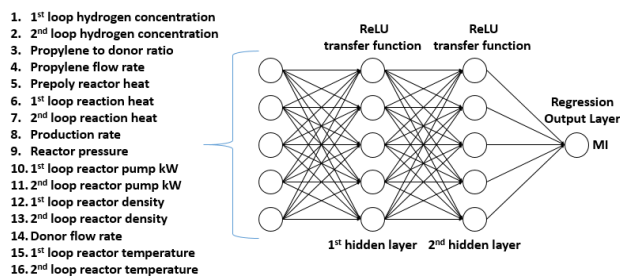
Figure 4 Deep Learning Topology

For deep learning training in MATLAB, instead of using "nntool", it trains using the "trianNetwork" function. Input and output variables are specified before the model was created. After that, the number of nodes for 1st, 2nd hidden layers and output layer are specified. Then, the deep neural network model is configured as the topology in Figure 4. In addition, training options have to be determined. Finally, training can be started once the input variables, output variables and neural network topology have been given to the "trianNetwork" function. The deep neural network has the full connection between all the input, hidden and output layers, and no connections between the same node in the same layer. The input and output variables are the same as those in shallow ANN model. After the training, the number of 1st and 2nd hidden nodes are systematically altered between 16 ~ 48 nodes and re-trained. Both 1st hidden and 2nd hidden layers have the same number of nodes to prevent the combination between the 1st and 2nd hidden nodes to be too complicated. Prediction from each deep learning with different numbers of nodes is cross-validated with the unseen data outside the training inputs. The number of nodes with the best prediction will be presented. Table 3 shows the summary of neural networks models used in this study.

Table 3 Summary of neural networks models

Model		Specification
Shallow ANN	-	Shallow neural networks model
	-	One hidden layer
	-	Log sigmoid transfer function for the hidden layer
SNN	-	Stacking of 4 different ANN models with different numbers of nodes as shown in Table 2
	-	Output from the ANN 1 – 4 serves as the input variables for ANN level 1 along with the input variables in Table 1
	-	All ANN in Level 0 and Level 1 use log sigmoid transfer function for the hidden layer
Deep Learning	-	Deep neural networks model
	-	Two hidden layers
	-	ReLU transfer function for the hidden layer

3.4　Performance Verification of Neural Networks

The model was validated using the cross-validation (unseen) data for all MI. The output of the network is compared to the target value of the testing data set by calculating the root mean square error (RMSE). The initial weights and biases that yield the smallest RMSE are recorded for each network being constructed. If the smallest RMSE cannot be achieved, the network construction is redeveloped by adjusting the number of hidden nodes. The formula for *RMSE* is as follows.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}(q' - q)^2}{n}} \qquad (2)$$

where *n* is the number of data. *q'* is the value from the data set (actual). *q* is the value from simulation (predicted).

4.0　RESULTS AND DISCUSSION

The neural network models (shallow ANN, SNN and deep learning) were developed by using data set collected from PP Spheripol Plant. 163 data sets for training and 49 data sets for testing with the MI ranges from 1 - 17 g/10mins.

Table 4 shows the performance comparison of different models on the reactor dataset. The best model developed to estimate the MI value would have the lowest RMSE value and the highest R^2 value. The RMSE result for shallow ANN, SNN and deep learning are 0.0586, 0.0624 and 0.0604 respectively, while the R^2 results are 0.9700, 0.9706 and 0.9717. Figure 5 to 7 are the graphs for the actual and predicted MI comparison.

Table 4 Performance comparison of different models (MI range of 1 – 17)

Model	RMSE	R^2
Shallow ANN	0.0586	0.9700
SNN	0.0624	0.9706
Deep Learning	0.0604	0.9717
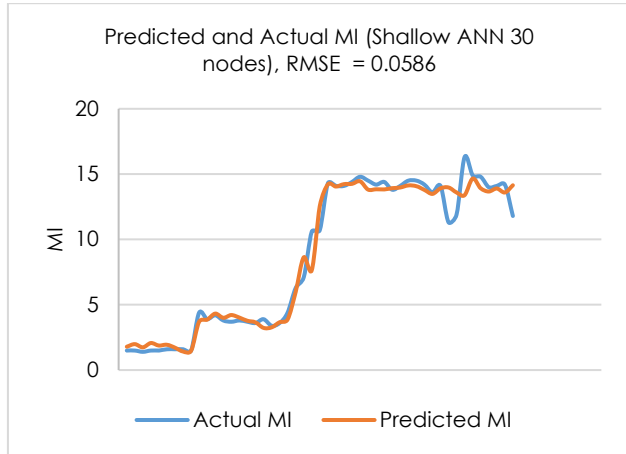
Figure 5 Predicted and actual MI (Shallow ANN MI 1 - 17)
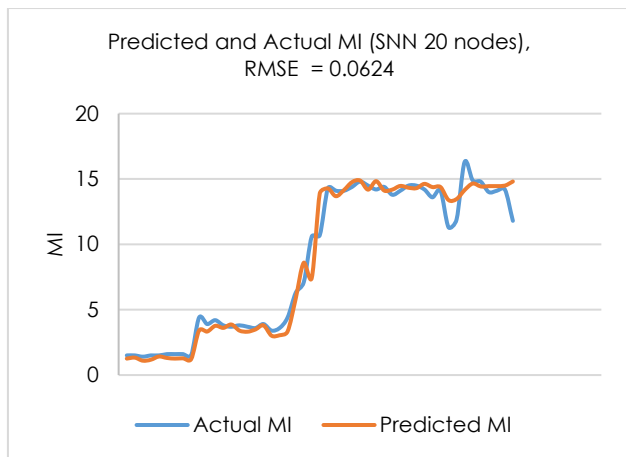
Figure 6 Predicted and actual MI (SNN MI 1 - 17)
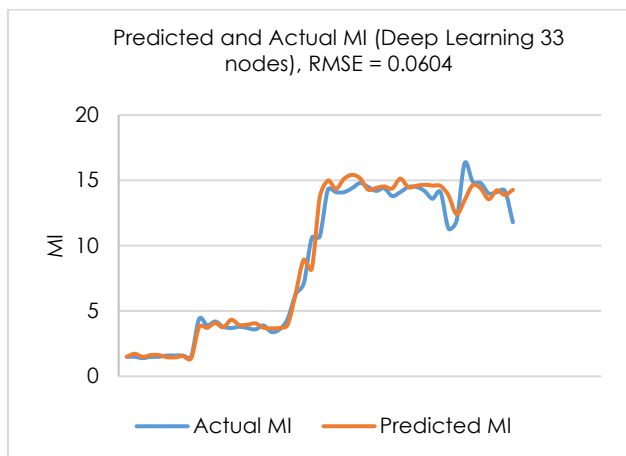
Figure 7 Predicted and actual MI (Deep Learning MI 1 - 17)

As shown in Table 4, the shallow ANN model predicted the MI with the lowest RSME, which is the highest in accuracy, followed by deep learning and then SNN model. On the other hand, deep learning prediction has the highest correlation with the actual data, with R^2 = 0.9717, followed by SNN model and then the shallow ANN model.

The shallow ANN model can represent complex and poorly understood behaviour very well. Although it does not use any structure that reflects the physical structure of the system, the shallow ANN model gives an input-output relation of the process and is useful in understanding the non-linear relationship between the variables involved. SNN and deep learning are the models improvised from the shallow ANN model. SNN targets to improve the generalisation capability of shallow ANN by combining several networks to improve model predictions on unknown data [21]. Deep learning aims to produce a better performance model by adding the number of hidden layers to increase the complexity and abstraction of the neural networks.

Table 4 shows that the performance of the 3 models are showing comparable results, and able to predict the MI of PP with great accuracy. Based on the result in Table 4, SNN model does not outperform shallow ANN models. This situation is similar to the result from Jumari and Mohd-Yusof [7], in which the SNN model prediction of a wide MI range does not outperform the shallow ANN model prediction. Results from Qazi and Yeung [22] also showed that the result from the SNN model was not significantly better than the shallow ANN model. This is because with the computational power nowadays, the shallow ANN models can be created with a large number of hidden nodes. Therefore, when the shallow ANN model is trained with a large number of data sets and high number of hidden nodes, the generalisation capability improves and the advantage of stacking the ANN become relatively small. In addition, the variability of the MI prediction is not wide enough to require different models to be stacked, which is the strength of SNN, resulting in similar results with the shallow ANN.

For deep learning, by adding one more layer in the ANN, deep neural networks model can detect the process fluctuation during MI dipped from 14 g/min to 12 g/min. Deep neural networks model has more complex architectures, where each layer performs specific operations on the inputs to provide various usable representations of the original signal. Neurons between consecutive layers are densely connected for the model to learn advanced characteristics and do hard tasks like resolving nonlinearly separable issues [23].

The research further investigates the training and testing of shallow ANN and deep learning modelling for a shorter range of MI. 77 data sets for training and 41 data sets for testing of MI range of 11 – 17.

Table 5 shows the performance comparison of different models on the reactor dataset. The RMSE result for shallow ANN and deep learning are 0.3213,
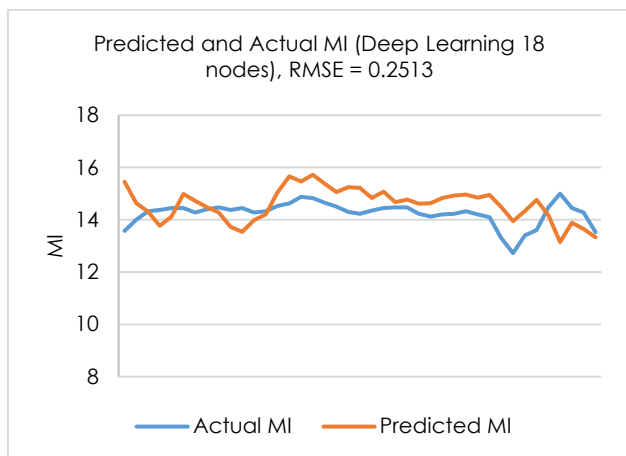
and 0.2513 respectively, while the R^2 results are 0.0248 and 0.0287. Figure 8 to 9 are the graphs for the actual and predicted MI comparison for the MI range of 11 - 17.

Table 5 Performance comparison of different models (MI range of 11 – 17)

Model	RMSE	R^2
Shallow ANN	0.3213	0.0248
Deep Learning	0.2513	0.0287
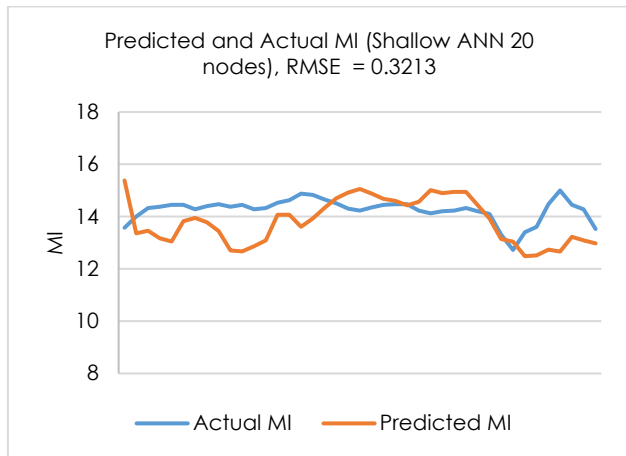
Figure 8 Predicted and actual MI (Shallow ANN MI 11 - 17)

Figure 9 Predicted and actual MI (Deep Learning MI 11 - 17)

The prediction comparison between Figure 8 and Figure 9 show that deep learning prediction with 18 nodes in the 1st and 2nd hidden layer is better compared to ANN with 20 nodes in the hidden layer. The difference between the actual and predicted MI in deep learning is mostly around ±1g/10min, which can be considered as a descent MI predictor. Compared to the sigmoid and tanh activation function in ANN, deep networks with ReLUs are more easily optimized because gradients can flow when the input to the ReLU function is positive.

$$Log\ sigmoid\ (x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Based on equation 3, the log sigmoid transfer function limits the output to a range of 0 to 1. The vanishing gradient problem results from the logarithmic sigmoid function's output becoming saturated for higher and lower inputs. The vanishing gradient problem describes a situation in which, while training a network using stochastic gradient descent, the gradient of the objective function with respect to a parameter approaches zero and almost completely disappears. Therefore, in the case of a vanishing gradient, training is virtually destroyed. Additionally, the lack of zero-centric output results in poor convergence for log sigmoid transfer function [24].

$$ReLU\ (x) = max(0, x) = \begin{cases} x, if\ x \geq 0 \\ 0, otherwise \end{cases} \quad (4)$$

From equation 4, deep networks with ReLU are easier to optimise than networks with log sigmoid because gradients can flow when the input to the ReLU function is positive [9]. Besides that, it is more computationally efficient to compute than Sigmoid functions since ReLU just needs to pick *max(0,x)* and not perform expensive exponential operations as in sigmoid function.

Oostwal *et al.* [8] found that for the sigmoidal transfer function, the neural network has a high likelihood to be trapped in an unfavourable configuration unless prior knowledge about the target is available. On the other hand, ReLU activation function does not have this issue and can produce good generalization from large training sets.

5.0 CONCLUSION

In this work, the shallow artificial neural network (ANN), stacked neural network (SNN) and deep learning were successfully developed to predict melt index (MI) of PP produced by Spheripol process. The results show that the three types of models are able to predict the MI of the polymer to a high degree of accuracy within a broad range of MI (1 – 17 g/10mins).

For prediction of MI range of 1 – 17g/10min, the shallow ANN model is able to predict the MI with the lowest RMSE, which is the highest in accuracy, followed by deep learning and then SNN model. Their RMSE are 0.0586, 0.0604 and 0.0624 respectively. Based on the analysis, although the shallow ANN has a slightly better prediction compared to deep learning, deep learning modelling is able to detect the process fluctuation when MI dipped from 14 g/10min to 12 g/10min, meaning that a more complex neural network is required to model a more complex problem by using the same training data.

The research continues with the prediction for MI range 11 – 17g/10min. Within this MI range, deep learning performed better with lower RMSE. The RMSE

result for shallow ANN and deep learning are 0.3213, and 0.2513 respectively.

Based on the results obtained in this study, it is proven that all the neural network models developed are able to perform non-linear function approximation for predicting MI. This indicates that neural network model (shallow ANN, SNN and deep learning) is an effective analytical tool that can be adopted to support MI prediction, such as for process optimization and model for soft sensors, in the polymer industry.

Acknowledgement

References

[1] Xu, S. and Liu, X. 2014. Melt Index Prediction by Fuzzy Functions with Dynamic Fuzzy Neural Networks. *Neurocomputing*. 142: 291-298. DOI: https://doi.org/10.1016/j.neucom.2014.03.056.

[2] Kiparissides, C. 1996. Polymerization Reactor Modeling: A Review of Recent Developments and Future Directions. *Chemical Engineering Science*. 51(10): 1637-1659. DOI: https://doi.org/10.1016/0009-2509(96)00024-3.

[3] Li, J. and Liu, X. 2011. Melt Index Prediction by RBF Neural Network Optimized with an MPSO-SA Hybrid Algorithm. *Neurocomputing*. 74(5): 735-740. DOI: https://doi.org/10.1016/j.neucom.2010.09.019.

[4] Zhang, M. and Liu, X. 2013. A Soft Sensor based on Adaptive Fuzzy Neural Network and Support Vector Regression for Industrial Melt Index Prediction. *Chemometrics and Intelligent Laboratory Systems*. 126: 83-90. DOI: https://doi.org/10.1016/j.chemolab.2013.04.018.

[5] Zhang, Z., Wang, T. and Liu, X. 2014. Melt Index Prediction by Aggregated RBF Neural Networks Trained with Chaotic Theory. *Neurocomputing*. 131: 368-376. DOI: https://doi.org/10.1016/j.neucom.2013.10.006.

[6] Zhang, M., Liu, X. and Zhang, Z. 2016. A Soft Sensor for Industrial Melt Index Prediction based on Evolutionary Extreme Learning Machine. *Chinese Journal of Chemical Engineering*. 24(8): 1013-1019. DOI: https://doi.org/10.1016/j.cjche.2016.05.030.

[7] Jumari, N. F. and Mohd-Yusof, K. 2016. Comparison of Product Quality Estimation of Propylene Polymerization in Loop Reactors using Artificial Neural Network Models. *Jurnal Teknologi*. 78: 6-13. DOI: https://doi.org/10.11113/jt.v78.9279.

[8] Oostwal, E., Straat, M. and Biehl, M. 2021. Hidden Unit Specialization in Layered Neural Networks: ReLU vs. Sigmoidal Activation. *Physica A: Statistical Mechanics and its Applications*. 564: 125517. DOI: https://doi.org/10.1016/j.physa.2020.125517.

[9] Ramachandran, P., Zoph, B. and Le, Q. V. 2017. Searching for Activation Functions. *arXiv preprint arXiv:1710.05941*.

[10] Li, H., Tang, X., Zhao, W. and Yang, B. 2021. Approaches to Deep Learning based Manipulating Strategy Reconstructions for Complex Chemical Processes. *Journal of Process Control*. 107: 127-140. DOI: https://doi.org/10.1016/j.jprocont.2021.10.009.

[11] Hubbs, C. D., Li, C., Sahinidis, N. V., Grossmann, I. E. and Wassick, J. M. 2020. A Deep Reinforcement Learning Approach for Chemical Production Scheduling. *Computers & Chemical Engineering*. 141: 106982. DOI: https://doi.org/10.1016/j.compchemeng.2020.106982.

[12] Agarwal, P., Tamer, M. and Budman, H. 2021. Explainability: Relevance based Dynamic Deep Learning Algorithm for Fault Detection and Diagnosis in Chemical Processes. *Computers & Chemical Engineering*. 154: 107467. DOI: https://doi.org/10.1016/j.compchemeng.2021.107467.

[13] Wang, Y., Wu, D. and Yuan, X. 2020. LDA-based Deep Transfer Learning for Fault Diagnosis in Industrial Chemical Processes. *Computers & Chemical Engineering*. 140: 106964. DOI: https://doi.org/10.1016/j.compchemeng.2020.106964.

[14] Debus, B., Parastar, H., Harrington, P. and Kirsanov, D. 2021. Deep Learning in Analytical Chemistry. *TrAC Trends in Analytical Chemistry*. 145: 116459. DOI: https://doi.org/10.1016/j.trac.2021.116459.

[15] He, Z., Tran, K.-P., Thomassey, S., Zeng, X., Xu, J. and Yi, C. 2021. A Deep Reinforcement Learning based Multi-criteria Decision Support System for Optimizing Textile Chemical Process. *Computers in Industry*. 125: 103373. DOI: https://doi.org/10.1016/j.compind.2020.103373.

[16] Fan, Z. and Xu, F. 2021. Health Risks of Occupational Exposure to Toxic Chemicals in Coal Mine Workplaces based on Risk Assessment Mathematical Model based on Deep Learning. *Environmental Technology & Innovation*. 22: 101500. DOI: https://doi.org/10.1016/j.eti.2021.101500.

[17] Niyomthai, T., Jongsomjit, B. and Praserthdam, P. 2017. Investigation of Alkoxysilanes in the Presence of Hydrogen with Ziegler-Natta Catalysts in Ethylene Polymerization. *Engineering Journal (Eng. J.)*. 21(7): 171-180. DOI: https://doi.org/10.4186/ej.2017.21.7.171.

[18] Alshaiban, A. and Soares, J. B. 2012. Effect of Hydrogen, Electron Donor, and Polymerization Temperature on Poly (Propylene) Microstructure. *Macromolecular Symposia*. 312(1): 72-80. DOI: https://doi.org/10.1002/masy.201100023.

[19] Funes, E., Allouche, Y., Beltrán, G. and Jiménez, A. 2015. A Review: Artificial Neural Networks as Tool for Control Food Industry Process. *Journal of Sensor Technology*. 5(01): 28. DOI: https://doi.org/10.4236/jst.2015.51004.

[20] Hagan, M. T., Demuth, H. B., Beale, M. H. and De Jesús, O. 1996. *Neural Network Design*. Pws Pub. Boston.

[21] Mukherjee, A. and Zhang, J. 2008. A Reliable Multi-objective Control Strategy for Batch Processes based on Bootstrap Aggregated Neural Network Models. *Journal of Process Control*. 18(7): 720-734. DOI: https://doi.org/10.1016/j.jprocont.2007.11.008.

[22] Qazi, N. and Yeung, H. 2014 Modeling of Gas–liquid Separation through Stacked Neural Network. *Asia-Pacific Journal of Chemical Engineering*. 9. DOI: https://doi.org/10.1002/apj.1777.

[23] Debus, B., Parastar, H., Harrington, P. and Kirsanov, D. 2021 Deep learning in analytical chemistry. *TrAC Trends in Analytical Chemistry*. 145: 116459. DOI: https://doi.org/10.1016/j.trac.2021.116459.

[24] Dubey, S. R., Singh, S. K. and Chaudhuri, B. B. 2022 Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing*. 503: 92-108. DOI: https://doi.org/10.1016/j.neucom.2022.06.111.