Training Machine Learning Algorithms for the Controller of VSLLIM Microreactor using Artificial Neural Networks

Mohamed S. El-Genk, Timothy M. Schriener, Ahmad N. Shaheen
Institute for Space and Nuclear Power Studies and Nuclear Engineering Department, University of New Mexico, Albuquerque, NM, USA

Technical Work Scope: DOE Office of Nuclear Energy's Nuclear Energy University Programs contract DE-NE0009268 subaward from Purdue University to the University of New Mexico (UNM).

Performance Period: 10-01-2023 to 09-30-2024

Report No. UNM-ISNPS-02-2024
Institute for Space and Nuclear Power Studies, The University of New Mexico,
Albuquerque, NM, USA, http://isnps.unm.edu/reports/

October 2024

Training Machine Learning Algorithms for the Controller of VSLLIM Microreactor using Artificial Neural Networks

Mohamed S. El-Genk, Timothy M. Schriener, Ahmad N. Shaheen

Institute for Space and Nuclear Power Studies and Nuclear Engineering Department

The University of New Mexico, Albuquerque, NM

Abstract

This report presents the progress made in the second year towards Demonstrating Autonomous Control, Remote Operation, and Human Factors for Microreactors. The University of New Mexico's Institute for Space and Nuclear Power Studies (UNM-ISNPS) investigated training Artificial Neural Networks (ANN) for remote control of the Very-Small, Long-LIfe, Modular (VSLLIM), a 1.0 – 10 MW_{th} microreactor. The two investigated and implemented algorithms for Supervised Learning and Reinforcement Learning Machine Learning paradigms to train the ANNs to control the position of the control rods during a simulated startup of the VSLLIM microreactor. The trained neural networks are incorporated into a controller program of a Programmable Logic Controller (PLC) for real-time control of the VSLLIM's Simulink model to evaluate their performance for the reactor startup. The Supervised Learning (SL) algorithm using a Long Short-Term Memory (LSTM) network showed excellent accuracy of up to 99.95% for predicting the correct control rod position when testing against the pre-generated VSLLIM startup data. However, the LSTM networks trainiing using supervised learning showed poor results when incorporated into PLC coupled to the transient VSLLIM Simulink model. We also investigated the Reinforcement Learning (RL) using the Soft Actor Critic (SAC) and Asynchronous Advantage Actor Critic (A3C) algorithms for training the ANNs. The SAC algorithm was able to train models that successfully completed the startup scenario, accurately predicting the control rods positions to within ±1.6% of the target values. The A3C algorithm failed in training a neural network to complete the startup, with the network predictions diverging to extreme rod position predictions. The SAC trained ANNs demonstrated superior performance when integrated into the real-time controller program compared to the LSTM networks trained using SL and were able to closely follow the target curve for the VSLLIM reactor power during startup. This research successfully demonstrated that the SAC ML algorithm can train artificial neural networks capable of controlling the startup of the Simulink model of the VSLLIM microreactor. The developed PLC controller with trained ANN is connected to a remote-control setup at UNM-ISNPS for further testing and demonstration of remote and autonomous control of the VSLLIM microreactor.

Nomenclature

A3C	Asynchronous Advantage Actor Critic
AI	Artificial Intelligence
ANN	Artificial Neural Network
ESD	Emergency Shut Down
FNN	Feed-forward Neural Network
FPY	Full Power Years
HEX	Heat Exchanger

LOBO NCS	LOBO Nuclear CyberSecurity
LR	Learning Rate
LSTM	Long Short-Term Memory
$\dot{m}$	Na mass flow rate (kg/s)
ML	Machine Learning
P_{sp}	Power setpoint (MW_{th})
P_{sp1}	Initial, low power setpoint (MW_{th})
P_{sp2}	Final, high power setpoint (MW_{th})
PD	Proportional-Differential
PI	Proportional-Integral
PLC	Programmable Logic Controller
Q_{Rx}	Reactor thermal power (MW_{th})
RC	Reactor Control
RL	Reinforcement Learning
RMSE	Root Mean Square Error
SAC	Soft Actor Critic
SL	Supervised Learning
T_{in}	Reactor Na inlet temperature (K)
T_{ex}	Reactor Na outlet temperature (K)
UNM-ISNPS	University of New Mexico’s Institute for Space and Nuclear Power Studies
VSLLIM	Very-Small, Long-LIfe, Modular

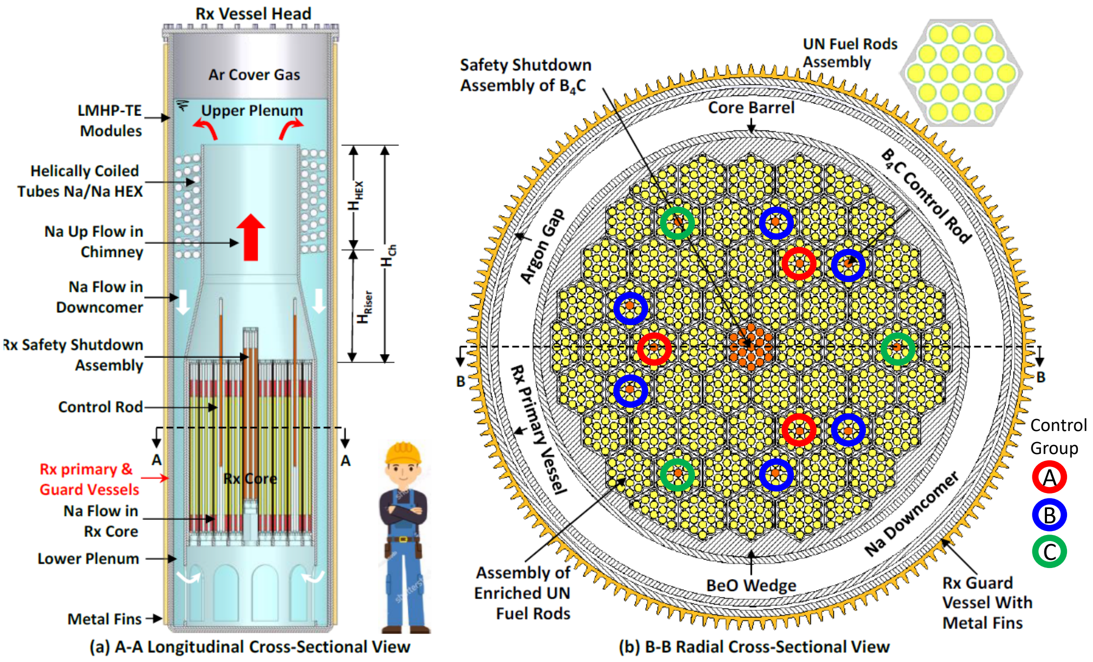
Greeks

α	Controller scaling coefficient
λ_e	Estimated effective decay constant for a delayed neutron group
ρtotal	Total reactivity ($)
τ	Reactor period (s)

1. Introduction

Microreactor designs envision deployment and potential operation remotely with a high degree of local autonomy [Agarwal et al., 2021]. The DOE Fission Battery initiative aims to develop microreactors capable of ‘plug-and-play’ operation with autonomous Instrumentation and Control (I&C) systems [Agarwal et al., 2021]. Safe and autonomous operation and control of microreactors requires developing resilient, dependable, and fault tolerant I&C systems to achieve fail-safe performance. Artificial Intelligence (AI) and Machine Learning (ML) algorithms, which can monitor, control, and diagnose anomalous operating conditions and take corrective actions, are being explored for partial and total autonomous control of nuclear power plants [Cetiner, et al., 2016]. The ML algorithms being investigated are within the supervised learning and reinforcement learning paradigms [Tang et al., 2022]. The Supervised Learning (SL) paradigm trains models using preexisting data sets that include both input state variables and the desired control action. The advantages include the ability to train models using historical plant operating data, as well as high quality pregenerated simulation data [Wang et al., 2019]. On the other hand, the Reinforcement Learning (RL) paradigm trains an intelligent agent to take actions while interacting with a dynamic environment. The algorithm trains the agent to maximize the cumulative reward for making the correct control actions [Sutton, 2018]. RL attempts to balance exploration (of the action space) and exploitation (of current knowledge of the responses to the controllers’ actions) to train the ML model and maximize the long-term reward [Kaelbling, et al., 1996].

The objective of this research is to investigate training ML algorithms of both the SL and the RL paradigms to remotely control the Very-Small, Long-LIfe, Modular (VSLLIM) microreactor developed at UNM-ISNPS [El-Genk and Palomino 2019; El-Genk, Schriener, and Palomino 2021]. The trained ML algorithms perform the function of the control rods’ Programmable Logic Controller (PLC) during the startup transient of VSLLIM to full power steady state operation. First, the SL paradigm is investigated using Long Short-Term Memory (LSTM) neural networks capable of processing sequential time series input data. The LSTM networks are trained using transient operating data sets generated using a MATLAB Simulink [The MathWorks, 2022] dynamic simulation model of the VSLLIM, developed at UNM-ISNPS.



(a) A-A Longitudinal Cross-Sectional View

(b) B-B Radial Cross-Sectional View

Fig. 1: Longitudinal and radial cross section views of the VSLLIM microreactor for generating 1-10 MW_{th} showing sodium flow path and locations of reactor control rod groups.

Parametric analyses are performed for the SL algorithm varying the model parameters to identify the combination for achieving high accuracy and low variation in the predicted positions of the control rods during the startup scenario. We investigated the RL paradigm for training Artificial Neural Networks (ANNs) in control of the Simulink VSLLIM dynamic model. This training investigated both the Soft Actor Critic (SAC) and Asynchronous Advantage Actor Critic (A3C) algorithms. In addition, we evaluated the performance of the trained neural networks for the three algorithms investigated while integrated into a PLC program for real-time control of the VSLLIM microreactor. The next section provides highlights of the VSLLIM microreactor design and the developed transient model in MATLAB Simulink used to train and evaluate the neural networks investigated.

2. Description of the VSLLIM Microreactor Design and the Developed Simulink Model

The walk-away safe VSLLIM microreactor design developed at UNM-ISNPS offers many passive operation and safety features. It is cooled by natural circulation of in-vessel liquid sodium (Na) during nominal operation and after shutdown and for decay heat removal with the aid of in-vessel helically coiled tubes Na/Na heat exchanger and 1-2 m tall chimney (Fig.1) [El-Genk and Palomino 2019; El-Genk, Schriener, and Palomino 2021]. Additional means for removing decay heat in the reactor after shutdown, and in case of a mal-function of the helical coiled tube Na-Na heat exchange placed at the top of the downcomer, is by natural circulation of ambient are along the surface of the reactor guard vessel [Palomino, El-Genk, Schriener 2019]. The reactor can continuously generate 10 - 1.0 MW of thermal power for ~5.9 - 92 Full Power Years (FPY), respectively, without refueling, thus eliminating the need for onsite storage of either fresh or spent nuclear fuel.

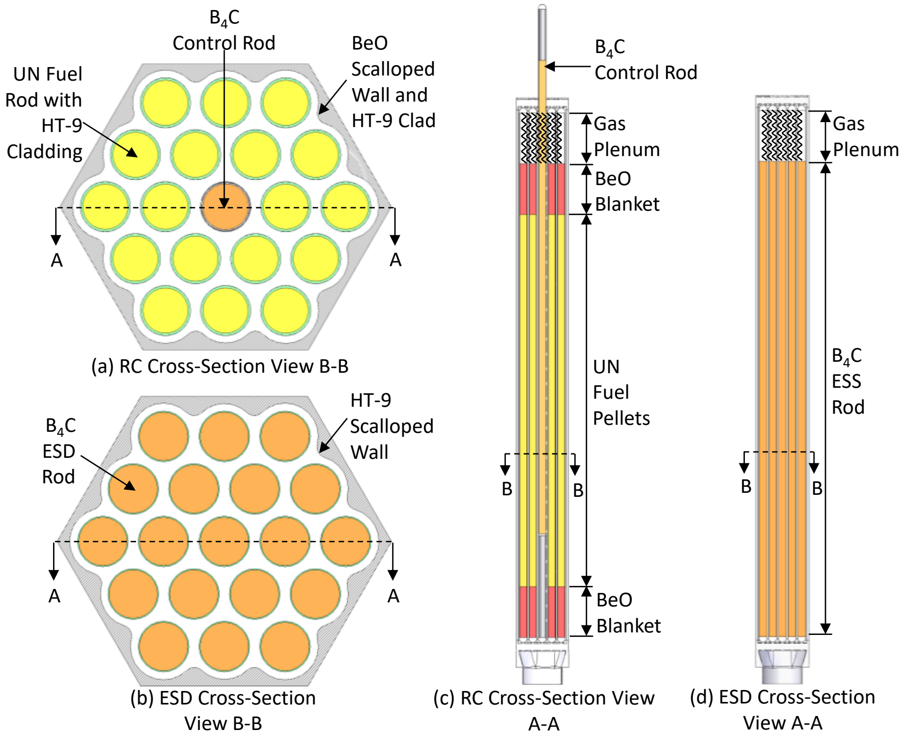

B_4C Control Rod
UN Fuel Rod with HT-9 Cladding
BeO Scalloped Wall and HT-9 Clad
A
A
(a) RC Cross-Section View B-B
B_4C ESD Rod
HT-9 Scalloped Wall
A
A
(b) ESD Cross-Section View B-B
B_4C Control Rod
Gas Plenum
BeO Blanket
UN Fuel Pellets
B
B
BeO Blanket
(c) RC Cross-Section View A-A
Gas Plenum
B_4C ESS Rod
B
B
(d) ESD Cross-Section View A-A

Fig. 2: Cross section and elevation views of the VSLLIM UN fuel assembly containing the B_4C reactor control rods and the center ESD assembly.

The VSLLIM reactor core is loaded with scalloped wall BeO hexagonal bundles of UN fuel rods each clad in HT-9 steel cladding Fig. 1b, 2a). The scalloped BeO walls help achieve a laterally uniform liquid sodium flow across the bundles [El-Genk and Palomino 2019]. A total of fifty-four full hexagonal bundles and six partial corner bundles of UN fuel rods in the reactor core are arranged in four concentric rings (Fig. 1b). Each full bundle with 19 UN fuel rods and the partial bundles each have twelve rods (Fig. 2a). The fuel bundles loaded in the reactor core are radially surrounded by BeO wedges and enclosed within the HT-9 steel core barrel, which also serves as a radial neutron reflector (Fig. 2b).

The VSLLIM microreactor has two independent means for reactor control. The first is the 12 B_4C Reactor Control (RC) rods located in selected fuel bundles or assemblies within the second and third rings of the core (Fig. 1b, 2a and c). The HT-9 clad control rods replace the centermost UN fuel rod in these assemblies (Fig 2a). The 12 B_4C control rods are divided into three groups identified as Groups A, B, and C and each has a separate drive motor (Fig. 1b). The control rods of naturally enriched B_4C pellets within HT-9 cladding have upper gas plenums to contain the helium generated by the neutron absorption in the Be during reactor operation. Group A is the three B_4C rods located in the second ring of fuel assemblies in the reactor core. Group B is the six B_4C rods in the third ring of the fuel assemblies adjacent to those of the Group A rods. Group C is the three B_4C rods in the fuel assemblies in the third ring of reactor core. For redundancy, the VSLLIM reactor core has a central Emergency Shut Down (ESD) assembly of nineteen HT-9 clad B_4C rods, 80% enriched in ^{10}B, with an HT-9 steel scalloped wall (Fig. 1, 2b and d). This assembly provides an independent means for shutting down the VSLLIM reactor in case of emergency. The next sections detail the dynamic model of the VSLLIM microreactor and reactor controller used to generate training and testing data of simulated startup transients.

2.1 VSLLIM MATLAB-Simulink Transient Model

The VSLLIM transient model had been developed previously at the UNM-ISNPS and has been used in the first year of the project. This model that couples 6-group point reactor kinetics and reactor thermal-hydraulics sub-models and is based on the versatile MATLAB Simulink platform [The MathWorks, 2022] for simultaneously solving the constituent equations of these sub-models for the physics-based operation parameters of the reactor as functions of time during simulated startup transients. These parameters are the reactor thermal power, the average temperatures of the UN fuel and cladding in the VSLLIM reactor core, the mass flow rate and the inlet and exit temperatures of the in-vessel circulating liquid sodium coolant in the core, the chimney, the upper and lower plenums, the downcomer, and helically collide tubes Na-Na heat exchanger (Fig. 1a). The transient model of the VSLLIM reactor also calculates the spatial temperature distributions of the Na/Na HEX's solid structure and the primary and the secondary liquid sodium flows and temperatures. The VSLLIM transient Simulink model uses the ode23s modified Rosenbrock solver within Simulink with a timestep size of twenty ms. Further details on the reactor kinetics and thermal-hydraulics models can be found in El-Genk, Schriener, and Shaheen [2024].

2.2. VSLLIM Microreactor Controller

The investigated ANNs of the three ML algorithms are trained to perform the actions of the Reactor Controller during startup transients of the VSLLIM microreactor. The Reactor Controller determines the movement rates of the ESD bundle and the control rod Groups A, B, and C (Figs. 1b, 2). This PLC receives commands from a remote operator to startup or shutdown the VSLLIM reactor as well as to set the desired values or setpoints of the reactor thermal power, P_{s1} and P_{s2}. The positions of the control rods in the separate groups in the VSLLIM reactor core and for the ESD bundle are used to calculate the external reactivity insertion in the reactor core as functions of time during the simulated startup transients. The reactivity worth of each of the control rod groups in the VSLLIM reactor core is determined as a function of the position and temperature using the MCNP6 code [Goorley 2014].

During a simulated startup, the VSLLIM reactor Controller regulates the positions of the control rods in the core and the ESD assembly to bring the reactor from an initial subcritical state to full operation at an operator specified power level. During startup it withdraws the ESD assembly and Group B control rods at constant rates of 5.833 mm/s and 4.28 mm/s, respectively, to bring the reactor to criticality. The PLC's logic then changes the position of the Group A and C control rods to increase the reactor thermal power to the desired power setpoint, P_{SP2}. When the reactor thermal power, Q_{Rx}, < 100 kW_{th} the controller withdraws Group A and C control rods at a constant rate of 0.75 mm/s. When Q_{Rx} reaches or exceeds 100 kW_{th} the PLC uses a Proportional-Differential (PD) controller to manage the control rods position at a variable rate ±0.125 mm/s for increasing the reactor thermal power. The PLC manages2 the withdrawal of the control rods from the reactor core to avoid a rapid increase in total reactivity and maintaining a smoother startup without spiking the reactor thermal power and temperatures. This is down using a criterion derived from a control scheme proposed by Bernard, Lanning, and Ray [1984] as:

$$\rho_{\text{total}} < \frac{1}{\alpha}\left[\frac{\left|\frac{d\rho}{dt}\right|}{\lambda_e} + \left|\frac{d\rho}{dt}\right| \tau \ln \frac{P_{SP}}{P_{Rx}}\right] \qquad (1)$$

In this expression, α is a scaling coefficient, $\frac{d\rho}{dt}$ is the rate of change in reactivity, τ is the reactor period, and λ_e is the estimated effective decay constant for a delayed neutron group. The scaling coefficient, α, is adjusted to increase or decrease the total reactivity before the PLC halts the withdrawal of the control rods to provides time for the delayed negative temperature reactivity feedback in the VSLLIM reactor by the thermal inertia of the system, to drop the total reactivity before further displacing the control rods. The value of α was set to be equal to 25 based on the results of testing using the VSLLIM Simulink model [El-Genk, Schreiner, Shaheen 2024].

Figure 3 shows the results of the startup sequence for the VSLLIM reactor with the Reactor Controller from initial subcriticality to full power nominal operation at 10 MW_{th}. The reactor startup sequence begins with the reactor subcritical with the in-vessel sodium thawed to a temperature of 500 K. The reactor controller then fully withdraws the ESD assembly from the reactor core at a constant speed of 5.833 mm/s over a period of 240 s (Point 1 in Fig 3a). At this point, the reactor is still subcritical. The reactor controller next partially withdraws the six Group B control rods 0.77 m from the reactor core at a constant rate of 4.28 mm/s over a period of 180 s until the reactor achieves criticality (Point 2 in Fig. 3a). Next, the PLC Controller withdraws simultaneously the Group A and C control rods at a constant rate of 0.75 mm/s until the reactor power reaches a level of 100 kW_{th} (Point 3 in Figs. 3a and b).

At this point the reactor controller switches to using the PD controller to regulate the position of the control rods in the core to bring the reactor thermal power to an initial $P_{SP,1}$ = 0.5 MW_{th}. The controller restricts the movement of the Group A and C control rods to limit the reactivity rise in the core during the startup transient. The HEX Na flow controller increases the flow rate of sodium through the tube side of the in-vessel Na/Na HEX when the inlet temperature of the reactor exceeds 600 K, with the flow rate regulated by a PI controller to maintain the rector inlet temperature constant at T_{in} = 610 K (Fig. 3c and d). Fig. 3b shows that the VSLLIM reactor reaches steady state operation at 0.5 MW_{th} at time t = 2.38 hr into the startup sequence.

The remote operator sends the command to the reactor controller to increase the power setpoint from 0.5 MW_{th} to 10 MW_{th} (Point 6 in Fig. 3). The controller continues to withdraw the Group A and C control rods simultaneously to insert external reactivity and increase the reactor power (Figs. 3a and b). The reactor power, core inlet and exit temperatures, and Na mass flow rate increases steadily over a period of 4.75 hrs until the reactor is leveled off at a power of 10 MW_{Th} (Point 7 in Fig. 3). After this point the system reaches steady state operating conditions at 10 MW_{th}, reactor inlet and outlet temperatures of 610.0 K and 780.6 K, and mass flow rate of 46.0 kg/s.

2.3. Generated Data for Machine Learning Training

The VSLLIM Simulink transient model generated the training and target data sets for the SL and RL paradigms during many simulated startup transients. Simulink saves the generated data for the reactor's state variables and controller actions with time throughout the simulated startup transients (Fig. 3). During the generation of training data, the reactor controller uses the PD controller described in Section 2.2 for the withdrawal of the Group A and C control rods in the reactor core (Fig. 1b). The generated training data sets are for a wide range of the low power setpoint $P_{SP,1}$ = 0.5 - 9.75 MW_{th}, and the higher power setpoint $P_{SP,2}$ = 1.0 - 10.0 MW_{th} in 0.25 MW_{th} increments. The generated data sets are also for various times for changing the reactor thermal power setpoints from $P_{SP,1}$ to $P_{SP,2}$. A total of 797 startup transient data sets with more than 956 million data points. The next section presents the implementation of the SL algorithm and the training results using the generated data sets.

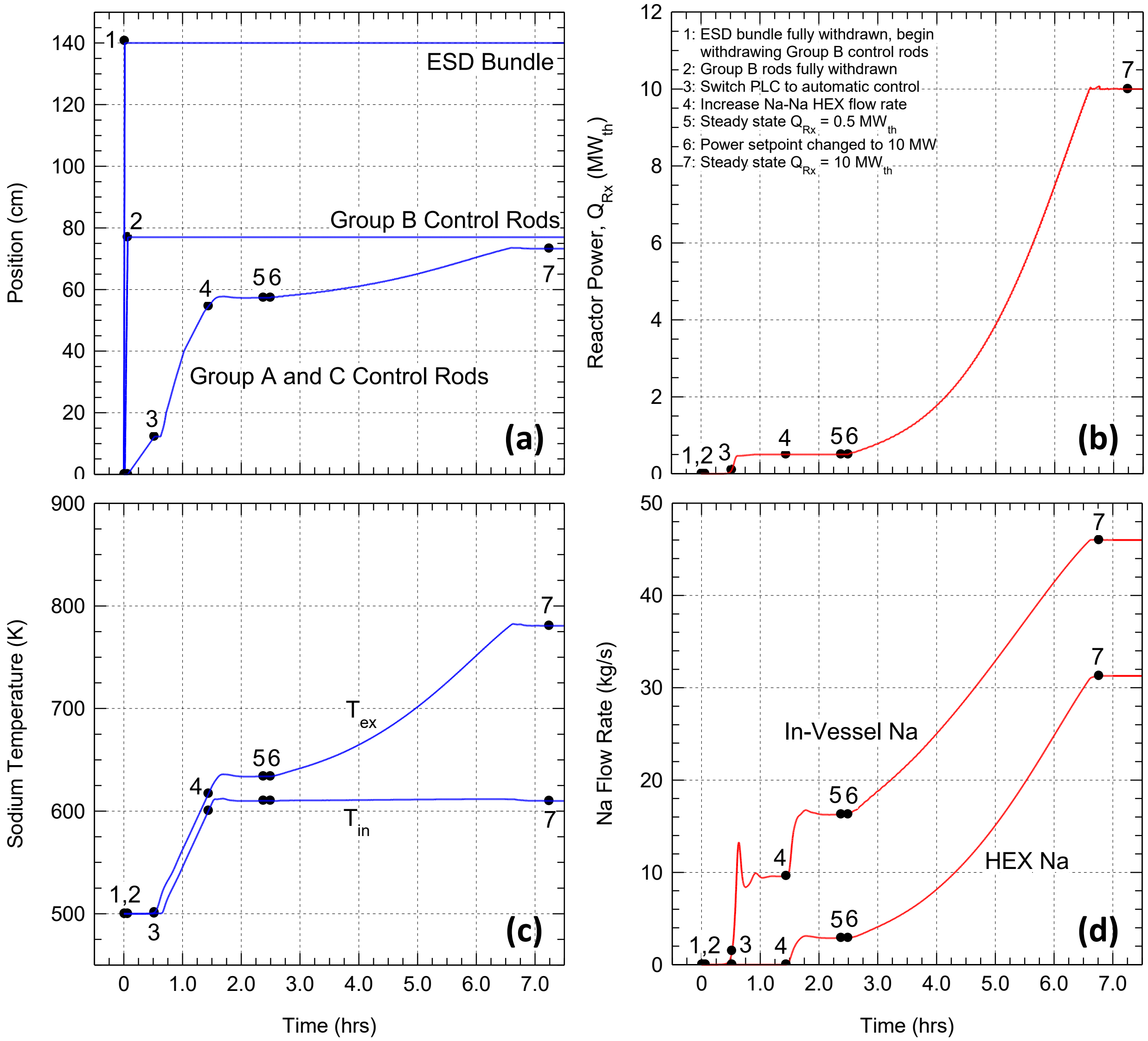


Fig. 3: Operation parameters of in-vessel sodium inlet and exit temperatures and mass flow rates in startup transients of the VSLLIM reactor with different values of scaling coefficient, α, in the Reactor Control PLC.

3. Supervised Learning Algorithm

The implemented SL algorithm trains the LSTMs using the generated transient data sets for numerous startup transients of the VSLLIM microreactor. Each data set represents a log file of the VSLLIM state parameters with time for a temporal discretization of data point every 0.2 s. For each training case, the implemented SL algorithm selects five of the primary input state parameters, or features, to train the LSTMs. These state variables are the Reactor thermal power Setpoints, P_{sp1} and P_{sp2}, the reactor Thermal Power, Q_{Rx}, the reactor Core sodium and core inlet temperature, T_{in}, the reactor core sodium outlet temperature, T_{out}, and the in-vessel liquid sodium mass flow rate, $\dot{m}$. The neural network is trained to estimate a single target value, which is the position of the Group A and C Control Rods in the VSLLIM reactor core (Fig. 1b). The SL algorithm attempts to learn the patterns and relationships between the five input features or state variables to accurately predict the target value of the position of the Group A and C control rods. The implemented LSTM model uses the Root Mean Square Error (RMSE) as the loss function and the AdamW optimizer with a weight decay constant = 0.1 [Paszke et al. 2017].

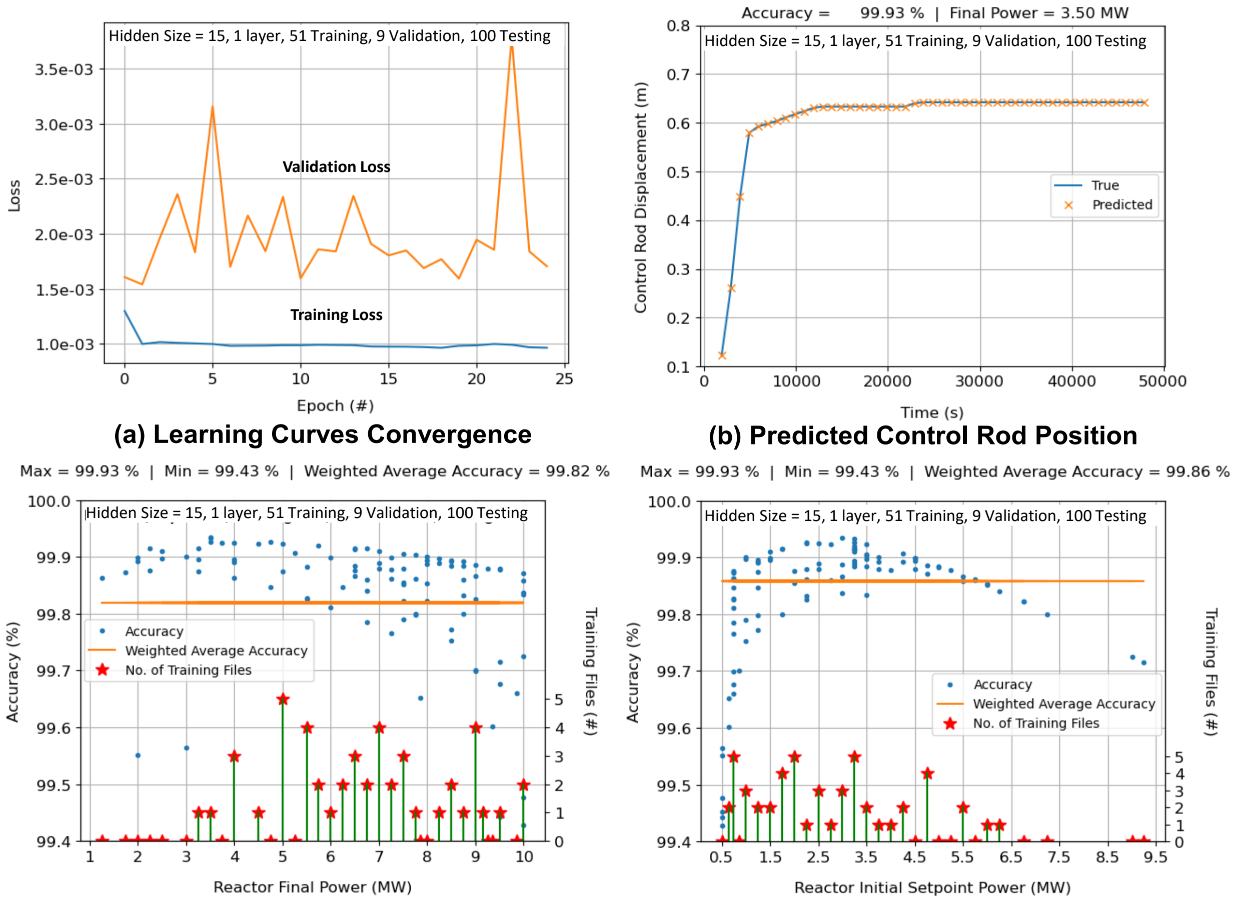
The generated data sets for the simulated startup transients are divided into three groups, one for Training, one for Validation, and one for Testing. During the Training, the weights and biases for the artificial neurons are updated based on the calculated RMSE training loss of the predicted control rod position relative to the 'true' values in the generated datasets. During the Validation phase, the model calculates the RMSE validation loss for the control rods position but does not update the weight and bias values. During each training cycle, or epoch, the SL algorithm goes once through all the Training and Validation data sets. The performance of the trained neural network is analyzed in the Testing phase by determining the RMSE of the testing sets. These sets are not included in either the training or validation data sets. The Testing phase evaluates how well the neural network predicts groups' A and C control rod position in the VSLLIM reactor core for simulated transients the network that is not exposed to previously.

3.1 Results of the Supervised Learning Algorithm

The optimization of the ML algorithms strongly depends on the model parameters, referred to as the hyperparameters. The performed parametric analyses investigate the effect of different hyperparameters on the weighted average accuracy of the algorithm and the accuracy of the VSLLIM controller. This work investigated the effects of the Learning Rate (LR) of the SL algorithm, the length of lookback window sequence, the size of the training and validation data sets, using equal numbers of trainings sets for each final power setpoint $P_{SP,2}$, the order of the training sets used in the training of the SL algorithm, the size and the number of hidden layers, and including additional parameters as features. The learning curves help to quantify the progress of the training process. The learning curves measure the change in the loss function with sequential epochs. The SL algorithm calculates the learning curves for the training and the RMSE calculates the validation losses calculated.

Figure 4 presents samples of the testing results of the case of randomly shuffled training data sets with one layer of neurons, hidden size of 15, 51 randomly selected training cases, 9 validation data sets, and 100 testing sets, and learning rate, LR= 0.00. The red star symbols along the bottom axes of Figs. 3c and 3d indicate the number of test cases for the values of $P_{SP,2}$ and $P_{SP,1}$. The training loss decreases to ~$1x10^{-3}$ after 3 epochs with slight change thereafter. The validation loss oscillates, and the predictions of the control rod position for one of the testing cases are in good agreement of the values determined by the trained neural network. Figs. 4c and 4d show the accuracy values plotted in increasing order for the final and initial power setpoints $P_{SP,two}$ and $P_{SP,1}$, respectively. The accuracy values show a spread between a minimum of 99.43% and a maximum of 99.93% for the one hundred testing data sets cases. The weighted average accuracies of the two power setpoints of 99.82 and 99.86% are comparable.

The performed parametric analyses for the SL algorithm by varying the Learning Rate determined that the validation loss does not converge when the initial learning rate is set to a high value of 0.1. A learning rate scheduler that varied the learning rate decreased the validation loss and increased the training loss with increasing numbers of epochs. The best training results are those for a constant learning rate of 0.001. Investigating the effect of the lookback window length showed that the testing accuracy increases as the size lookback window increases up to twenty, with difference in accuracy for longer sequence lengths se negligible. With large lookback windows of 250 and 500 points the SL algorithm failed during the backpropagation step of the training process.



(a) Learning Curves Convergence

(b) Predicted Control Rod Position

(c) Accuracy with Final Power $P_{SP,2}$

(d) Accuracy with Initial Power Setpoint $P_{SP,1}$

Fig. 4. Sample results for SL algorithm with LSTM for hidden size of 15, one layer of neurons, LR = 0.001 and randomly shuffled training data sets.

The analyses investigating the effect of increasing the hidden size for the LSTMs determined that increasing the hidden size from 5 to 10 increased the weighted average accuracy but decreased the accuracy spread. A hidden size of ten gave the highest accuracy values. Larger hidden sizes to up to thirty decreased the accuracy. Increasing the number of layers from 1 to 2 improved the testing accuracy of the models, however the results for networks with three layers are like those for two layers. Varying the number of data sets used for training the models increased the testing accuracy as number of sets increased to five. There were small further improvements as the number of training data sets increased from fifty up to 626.

Investigating the effect of ordering of the training data sets showed that ordering the data sets by final power setpoint, $P_{SP,2}$, from low-to-high resulted in the highest weighted average accuracy and the lowest accuracy spread. Ordering the data sets by $P_{SP,2}$ from high-to-low gave poor testing accuracy and random ordering accuracy values slightly below those for ordering the data from low-to-high. Random shuffling of the training data sets for $P_{SP,1}$ resulted in the highest testing accuracy values. The parametric analyses investigated additional input parameters beyond the five primary features (P_{SP}, Q_{Rx}, T_{ex}, T_{in}, $\dot{m}$). These included the time derivatives of the features in addition to their scalar values, and the total, external, and feedback reactivity values. Adding these features had little to no effect on the calculated weighted average accuracy values for the testing cases.

The SL algorithm trained the LSTM networks to predict the position of the Group A and C control rods in the VSLLIM reactor core (Fig. 1b) for the testing cases with a high degree of accuracy. Testing results for the SL algorithm showed that the trained models have average weighted accuracy for the testing sets up to >99.9% for the LSTMs with hidden size of 10 and 2 layers of neurons.

4. Reinforcement Learning Paradigm

The Reinforcement Learning (RL) paradigm is an ML process where the model interacts directly with the environment and learns based on feedback from its actions. The method attempts to optimize the actions to maximize a reward given to the actor based on the responses of the environment to those actions. A RL algorithm employs an Actor Network which interacts directly with the process and makes control actions, and a Critic Network which evaluates the performance of the Actor Network and adjusts the values of its weights and biases to improve its performance. Fig. 5 shows a block diagram of a basic RL process. The Actor takes control actions which are passed to the environment, in this case the VSLLIM Simulink model. The model calculates the change in the reactor operation parameters in response to the control action and passes these values to the Critic. The Critic evaluates the parameters and calculates a Reward based on how close the parameters are to the target values. The operation parameters are passed to the Actor to calculate another control action while the Reward is used to update the network parameters of the Actor. The present work investigated the performance of two RL algorithms, namely: (a) the Soft Actor Critic (SAC), and (b) the Asynchronous Advantage Actor Critic (A3C), for controlling the VSLLIM Simulink model.

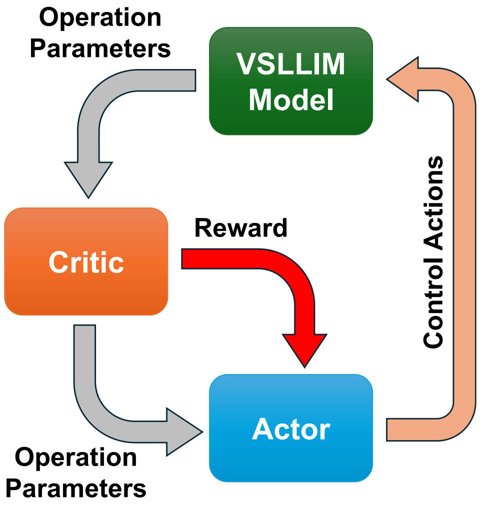


Fig. 5. Block diagram of a Reinforcement Learning Process

The implemented SAC algorithm uses Feed-forward Neural Networks (FNNs). These ANNs process information in one direction, where the output values of each layer of neurons are passed to the inputs of the subsequent layer of neurons. The implemented SAC algorithm, based on that developed by Bae, Kim, and Lee [2023], is incorporated into a Python program using the Tensorflow [Abadi, et al. 2016] and Keras ML libraries [Chollet 2015]. The SAC algorithm stores data in a replay buffer during each training episode and uses this data to update the actor and critic networks at the end of each episode. A developed POSIX shared memory interprocess communication function couples the Python ML algorithm to the transient Simulink model of the VSLLIM reactor. The Python code launches the VSLLIM Simulink model at the start of each episode using the MATLAB engine for python [Mathworks, 2022].

The A3C algorithm is implemented in a Python program using the Tensorflow [Abadi, et al. 2016] and Keras ML libraries [Chollet 2015]. The same shared memory interprocess communication link couples the Python program to the VSLLIM Simulink model as above. The implemented A3C algorithm uses an LSTM network with two layers, and a hidden size of ten. The algorithm updates the networks during the episode after each period of ten twenty ms timesteps. The same neural network is used for both the actor and the critic functions. The training process begins with randomly selected values of the weight and bias matrices in the ANN.

4.1. Results of SAC Reinforcement Learning

The implemented SAC algorithm uses a batch size of 256, actor and critic learning rates of 0.000001, a discount factor of 0.99, and a buffer size of two million experiences. This work conducted twenty-four separate training cases, labeled Cases F through AE. The performed parametric analyses for the SAC algorithm varied the number of neurons in the hidden layers of the actor and critic networks and the initial values for the weights of the actor and critic networks. Testing cases are investigated with 256, 64, 32, and 16 neurons in each layer of the actor network. The initial weights for the actor and critic networks are either selected randomly or are taken from a successful model from previous training case. All the cases for the models successfully completing the training scenario began with randomly determined initial weight and bias values.

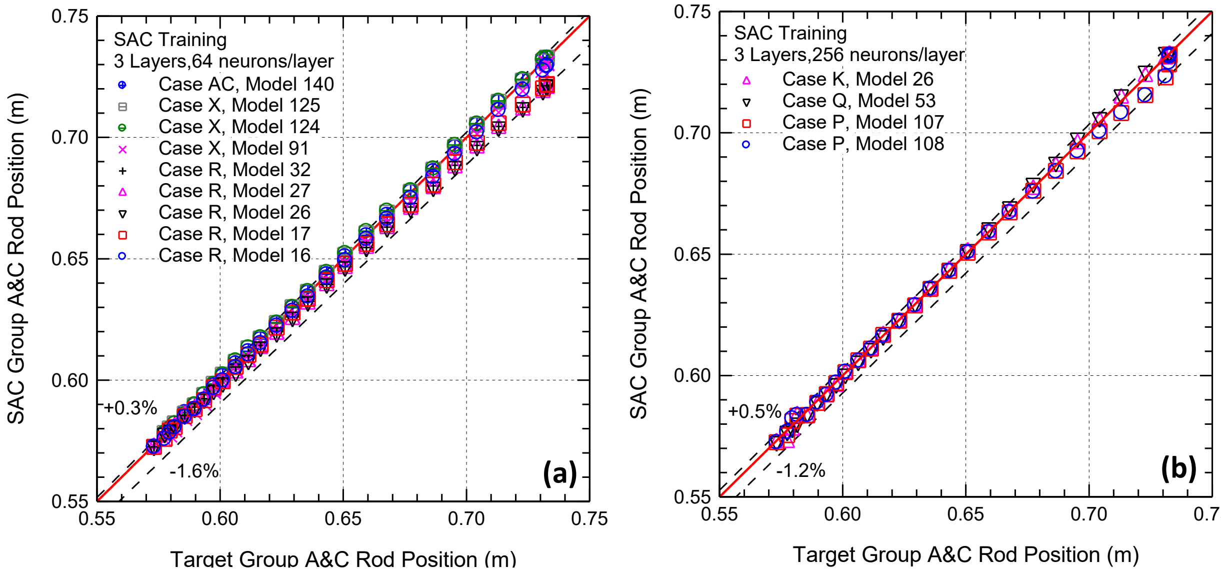


Fig. 6. Comparison of the predicted Group A and C control rods position by the SAC algorithm during training versus the target values: (a) models with three layers, sixty-four neurons/layer, and (b) models with three layers, 256 neurons/layer.

The twenty-four training cases generated a total of thirteen models for the episodes which completed the startup scenario. Fig. 6 compares the predicted Group A and C control rods' position by the FNN during training with the SAC algorithm to the targets in the training data set. Four of these models are for cases where the actor FNN has 256 neurons/layer (cases K, P, Q in Fig. 6a) and for nine networks with sixty-four neurons/layer (case R, X, and AC in Fig. 6b). The nine models with 64 neurons/layer accurately determined the control rods position to within +0.3 % and -1.6% of the target value throughout the startup scenario (Fig. 6a). The four models with 256 neurons/layer also showed only small deviations, with the predictions are within +0.5% and -1.2% from the target value.

While the SAC algorithm successfully trained the neural networks to complete the startup scenario, most training cases failed to produce trained models. The highest percentage of success is for Case R, with

10.2% of the episodes completing the training scenario. The exploration function of the SAC algorithm frequently caused the critic network to adjust the actor network in ways that resulted in successful models which appeared sporadic during training.

4.2. Results of A3C Reinforcement Learning

The training with the A3C algorithm used LSTM networks with two layers and a hidden size of ten. The python program is coupled to the VSLLIM Simulink model with the critical updating of the actor network every ten, 20 ms simulation timesteps. The implemented A3C algorithm failed in training a network to complete the startup scenario without termination. The algorithm calculated large gradients between the old and values for the policy updates to the model weight and bias values. This caused large swings in the predictions of the control rod positions by the LSTM networks as the number of episodes increased, trending to extremes of the rods being fully withdrawn or fully inserted. The A3C algorithm only considers recent data for the last ten timesteps when updating the policy, which could have resulted in it overcorrecting based short-term changes in the reactor operating variables, as opposed to the SAC algorithm which updated its policy based on data sampled throughout the startup sequences in the episodes.

5. Real Time Testing of VSLLIM Microreactor Controller

The SL and RL algorithms are both able to train neural networks with demonstrated good testing accuracy during the training process. This shows that the networks can accurately predict the Group A and C control rods' position under the testing conditions. However, the networks' performance needs evaluation for the intended application of real time reactor control of the VSLLIM microreactor. To accomplish this, the trained neural networks using the different ML algorithms are integrated into a developed python Reactor Control PLC program (Fig. 7). This program is coupled to the LOBO Nuclear CyberSecurity (NCS) platform, developed by UNM-ISNPS in collaboration with Sandia National Laboratory [El-Genk and Schriener 2022; Schriener and El-Genk 2022]. The LOBO NCS platform synchronizes the timing of the VSLLIM Simulink model to a real-time clock such that each timestep takes 20 ms physical wall time.

The controller has two data communication channels, a Modbus TCP channel which communicates with the LOBO NCS platform, and a TCP/IP channel for communicating with the remote operator (Fig. 7). The Modbus communication channel with the LOBO NCS data broker receives the values of the state variable from the VSLLIM Simulink model and stores them within its Modbus holding registers. It transmits back the control signals stored in the Modbus holding register for the movement of the control rods and ESD assembly in the VSLLIM reactor core. The measured network latency between the two computers is ~0.2 ms. The remote operator sends commands to startup and shutdown the reactor, or individually moves the control rod groups. The PLC transmits back status monitoring data of its present actions and store state variable values within its Modbus holding registers. The remote operator station has a large screen with a Graphical User Interface (GUI) for monitoring the state of the VSLLIM simulation (Fig. 7). During each scan cycle, the PLC reads the most recent values of the state variable from the holding registers and passes them to the control logic program to determine the appropriate control action. The movement rate for the Group A and C control rods in the VSLLIM reactor core is calculated from the difference between the present rod position and the desired position determined by the ANN. The movement rate is limited to ±0.125 mm/s. It is the same used with the PD controller to generate the training data. The controller then writes the commanded movement rates to its Modbus holding registers to them sent back to the VSLLIM Simulink model. At the end of the scan cycle the timing function of the Reactor Control PLC checks if the elapsed time is less than the user specified scan cycle time. If so, it activates a wait function to attempt to maintain a regular cycle time.

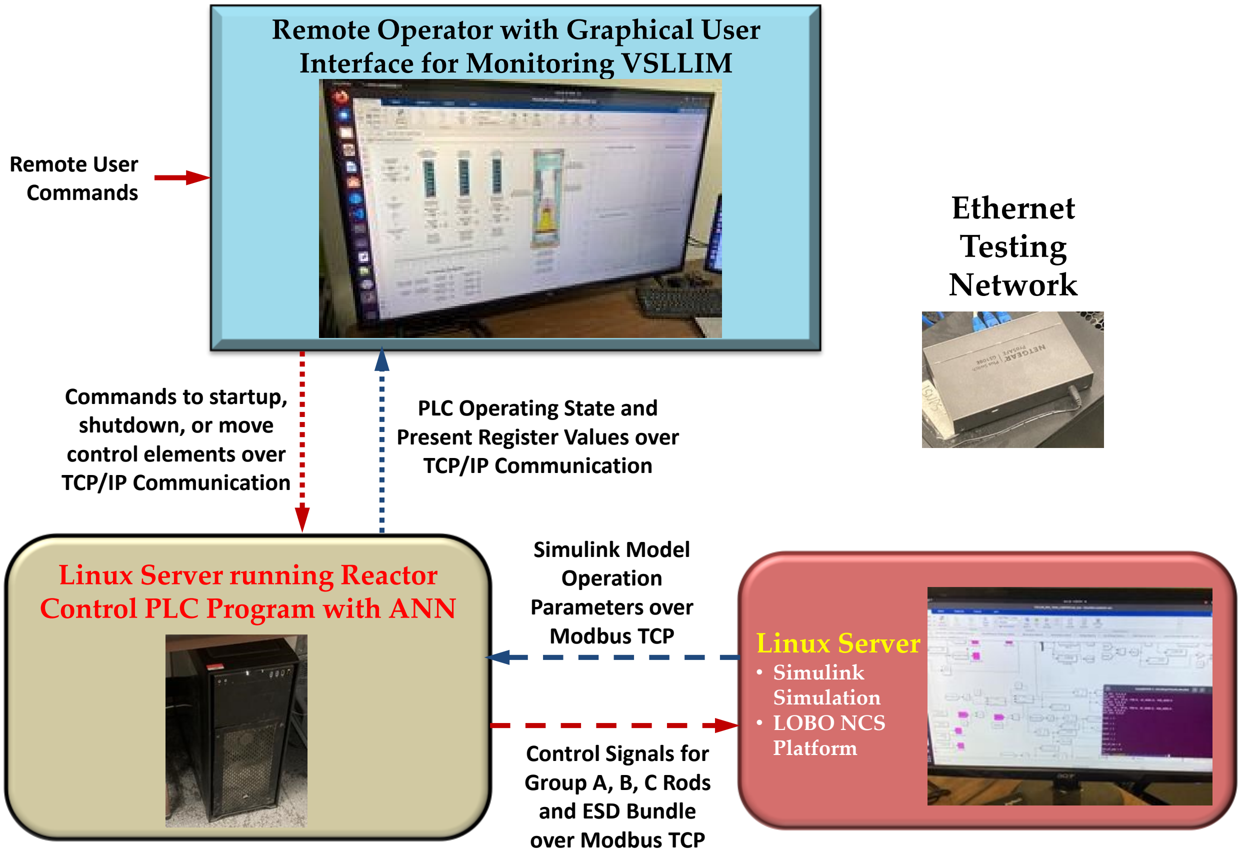

Remote Operator with Graphical User Interface for Monitoring VSLLIM
Remote User Commands
Ethernet Testing Network
Commands to startup, shutdown, or move control elements over TCP/IP Communication
PLC Operating State and Present Register Values over TCP/IP Communication
Linux Server running Reactor Control PLC Program with ANN
Simulink Model Operation Parameters over Modbus TCP
Linux Server
• Simulink Simulation
• LOBO NCS Platform
Control Signals for Group A, B, C Rods and ESD Bundle over Modbus TCP

Fig. 7. Remote control testing setup for the Control PLC with trained ANN coupled to a real-time MATLAB Simulink model of VSLLIM microreactor.

5.1. Results of Real-Time Testing using LSTMs trained by Supervised Learning

Figure 8 plots the predicted thermal power, Q_{Rx}, of the VSLLIM reactor by four trained networks using the supervised learning paradigm. These are for a startup scenario where the PLC is commanded to increase the reactor thermal power from an initial steady state value of 1.0 MW_{th} to 10 MW_{th}. The red curve in this figure is of the target transient response of the reactor thermal power in the training data sets. The LSTM networks showed mediocre performance, withdrawing the Group A and C control rods too quickly or too slow, and leveling off at the wrong steady state thermal power of the reactor (Fig. 8). In cases (a), (c), and (d) in Fig. 8 the trained controllers approximately reached the correct final reactor thermal power of 10 MW_{th}, however, they rapidly withdraw the Group A and C control rods causing the reactor thermal power to increase faster than the reference target curve. For case (b) in Fig. 8, the Reactor Control PLC closely tracks the correct reactor thermal power during the first 6,000s of the simulated startup transient, following the change in the power setpoint. However, the reactor thermal power levels off at a low steady state value of only 6.37 MW_{th}, well below the setpoint of 10 MW_{th}. These results show that achieving high ML testing accuracy with SL from the pre-generated data sets does not necessarily produce a good real-time control performance.

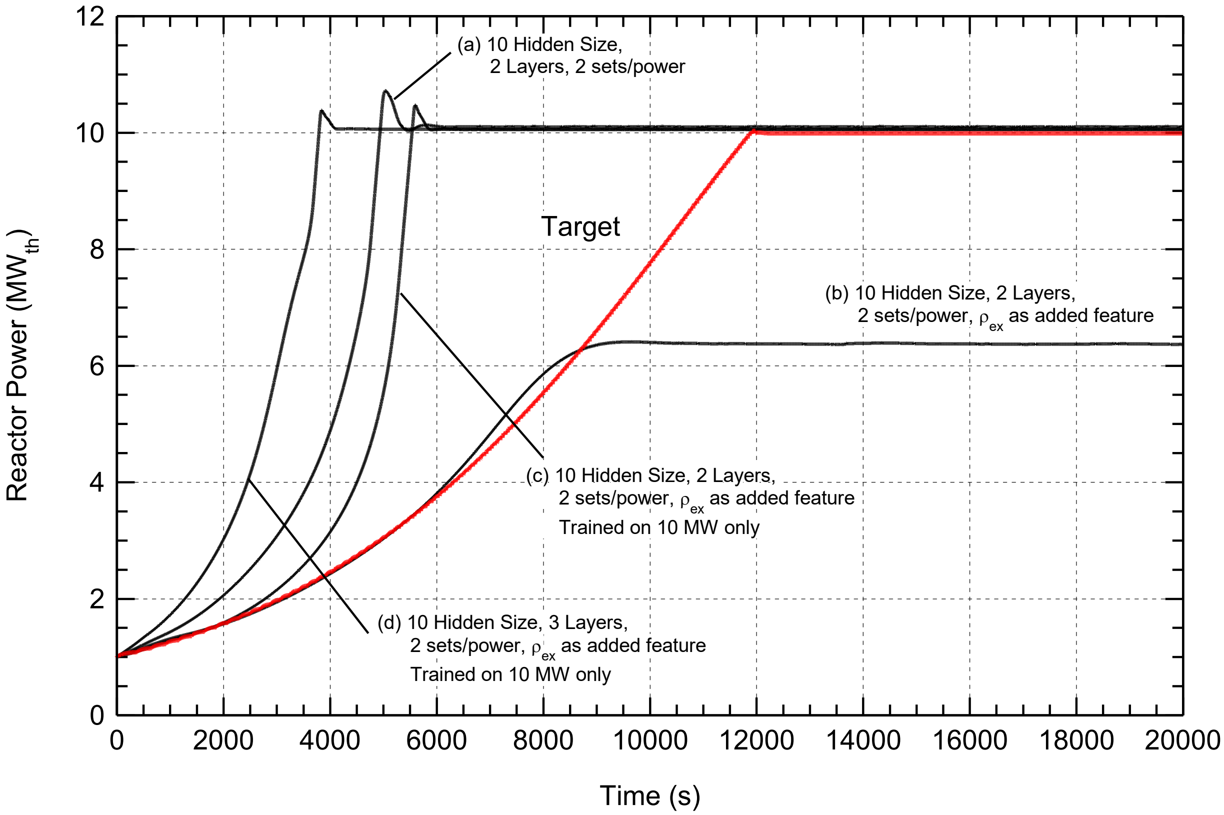


Fig. 8. Comparisons of the predictions of the thermal power of the VSLLIM reactor for selected LSTM trained networks using the SL algorithm of the to the target values (red curve), following a change in the reactor power setpoint from 1.0 to 10 MW$_{th}$.

5.2. Results of Real-Time Testing using FNNs trained by Reinforcement Learning

The trained FFNs using the SAC algorithm (Section 4.1) are integrated into the controller PLC and evaluated for the real-time startup transient results of the VSLLIM Simulink model. The results shown in Fig. 9 are for the trained neural networks that manage the movement of the Group A and C control rods in the VSLLIM reactor core (Fig. 1b) to increase the thermal power from a steady state value of 0.5 MW$_{th}$ up to a final reactor thermal power of 10 MW$_{th}$. These are network model sixteen, for Case R with three layers and sixty-four neurons/layer (Figs. 9a and b), and network model twenty-six for Case K with 3 Layers and 256 neurons/layer (Figs. 9c and d).

Figures 9a and c compare the predicted values of the control rods position by the FNN in the PLC to that of the target training case. Figs. 9b and d plot the change in the reactor thermal power during the startup transient. The FFN model sixteen for Case R slightly underpredicted the control rod positions to within -0.6% of the correct target values (Fig. 9a). As a result, the predicted values of the reactor thermal power are slightly lower than the target throughout the startup transient compared to that for the target case (9.8 MW$_{th}$ compared to 10.0 MW$_{th}$) (Fig. 9b). Model twenty-six for Case K predicted values of the position of the control rods positions differed only slightly from the target values, to within +0.7% and -0.5% (Fig. 9c). The controller with this model leveled off the final steady state reactor thermal power at 9.93 MW$_{th}$, slightly lower than target of 10 MW$_{th}$. These results demonstrate the predictions of the trained neural networks using the SAC reinforcement learning algorithm are reasonably accurate and better than those trained using SL for real-time control of the VSLLIM microreactor (Figs. 8 and 9).

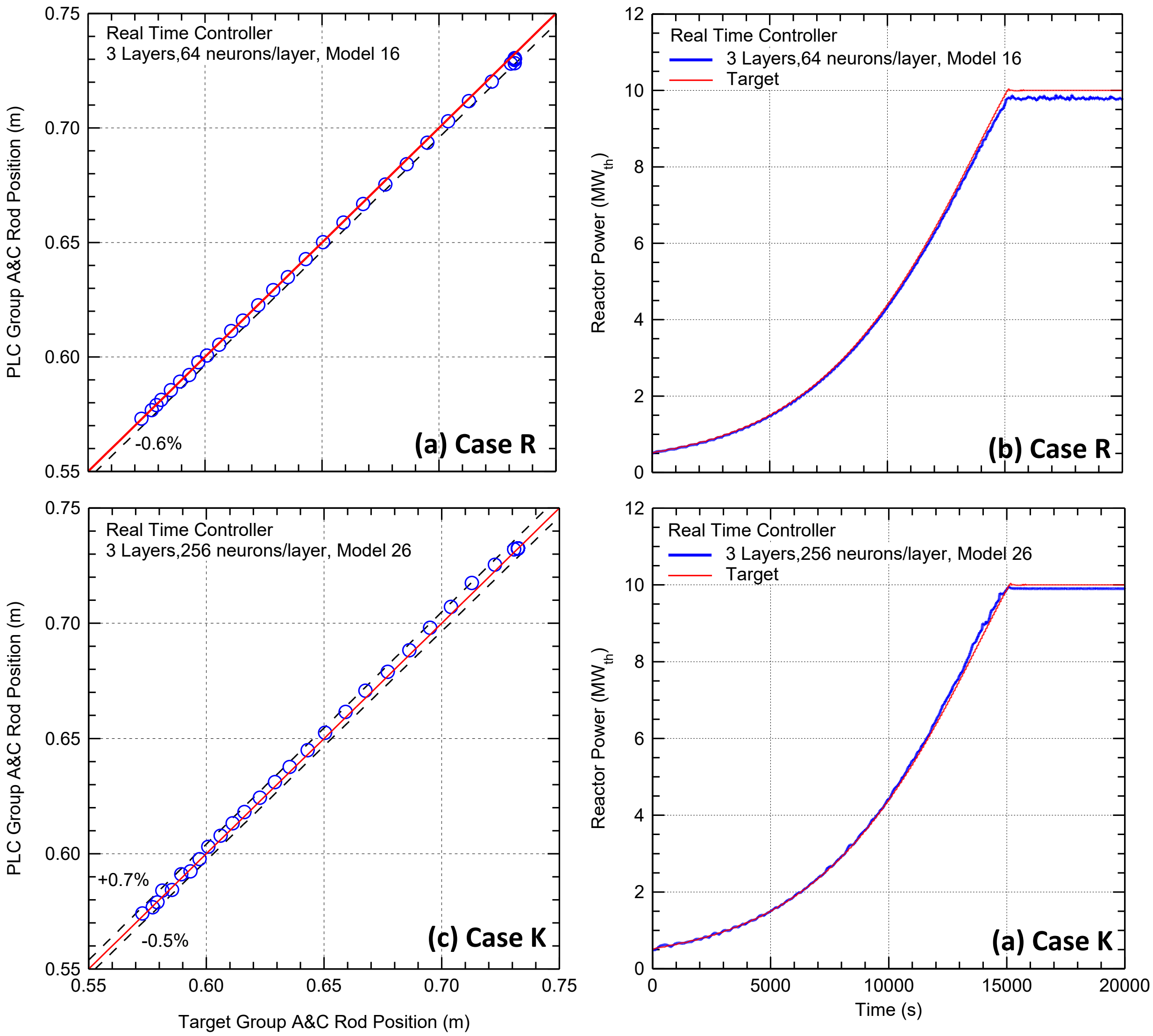


Fig. 9. Comparison of predicted position of Group A and C control rods and the thermal power of the VSLLIM microreactor for Model 16 of Case R with sixty-four neurons/layer and Model 26 of Case K with 256 neurons/layer to the target values.

6. Summary and Future Work

This work investigated training the controllers of the VSLLIM microreactor for remote operation during startup transients using three ML algorithms evaluated the performance for real-time control. These algorithms are: (a) the SL paradigm implemented into a Python program using the PyTorch ML library using a LSTM neural network, and the RL paradigm using (b) the SAC and (c) the A3C algorithms implemented in Python programs using the Tensorflow and Kares ML libraries. The trained neural networks using these three ML algorithms are integrated into a reactor control PLC program and coupled to the transient Simulink model of the VSLLIM microreactor using the LOBO NCS platform. This arrangement investigated the performance of the ML algorithms for real-time control of the VSLLIM reactor.

The dynamic, physics-based MATLAB Simulink model of the VSLLIM microreactor generated 797 transient data sets of startup transients of the reactor from an initial subcritical state to steady state full power operation at different power levels up to 10 MW_{th}. The SL algorithm trained the LSTMs using these data sets. The trained SL algorithm using LSTMs predicted the position of the Group A and C control rods in the core of the VSLLIM reactor with an accuracy up to >99.9%. However, this high accuracy did not translate to good real-time performance when the trained LSTM networks are incorporated into the reactor control PLC. The trained networks using the SL algorithm showed mediocre performance for real-time control of the VSLLIM microreactor, withdrawing the control rods too rapidly or too slowly. Owing to the absence of control feedback, the SL methods are unable to correct their predictions once the reactor thermal power diverged beyond the specified range of the training data.

The RL SAC and the A3C algorithms are trained while in direct control of the VSLLIM Simulink model. The exploration nature of the SAC algorithm caused the training cases to be highly inconsistent in their ability to train the models. The algorithm did succeed in training a total of thirteen different neural networks, four with 256 neurons/layer in the actor network, and nine with sixty-four neurons/layer in the actor network. The SAC trained ANNs showed superior real-time control performance when incorporated into the reactor control PLC compared to those trained using the SL algorithm. The A3C algorithm was unable to successfully train the LSTM networks to complete the startup scenario and displayed large shifts in the values of the network parameters during training. This caused the predictions of the neural network to converge to the extreme positions of fully withdrawn and fully inserted the control rods in the core of the VSLLIM reactor.

This research successfully demonstrated that the SAC ML algorithm can train artificial neural networks capable of controlling the startup of the VSLLIM microreactor. The PLC controller program with the trained networks can communicate with a remote operator station for monitoring the reactor's operation and sending commands to controller. This remote-control function has been demonstrated within an isolated Ethernet network at the UNM-ISNPS computational lab. Future work will expand the remote-control setup with signal and data encryption allowing the remote operator to securely send commands and receive monitoring data over an open internet connection. This capability will be further evaluated by having members of the Purdue team in Indiana remotely control the VSLLIM dynamic system model running on a computer at the UNM campus in Albuquerque, New Mexico.

Acknowledgements

DOE Nuclear Engineering University Program, through a subaward to the University of New Mexico from Purdue University, supported this research.

References

Abadi, M., 2016, "TensorFlow: A System for Large-Scale Machine Learning," in Proceedings 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16), November 2-4, 2016, Savannah, GA, USA.

Agarwal, V., Ballout, Y. A., Gehin, J. C., 2021, "Fission Battery Initiative: Research and Development Plan," Idaho National Laboratories technical report INL/EXT-21-61275, Idaho Falls, ID.

Bae, J., Kim, J.M., Lee, S.J., 2023, "Deep reinforcement learning for a multi-objective operation in a nuclear power plant," Nuclear Engineering and Technology, 55(9), 3277-3290. https://doi.org/10.1016/j.net.2023.06.009

Bernard, L.A., Lanning, D.D, Ray, A., 1984, "Digital Control of Power Transients in a Nuclear Reactor," IEEE Transactions on Nuclear Science, NS-31(1), 701-705.

Cetiner, S. M., et al., 2016, “Supervisory Control System for Multi-Modular Advanced Reactors,” Oak Ridge National Laboratory technical report ORNL/TM-2016/693, Oak Ridge, TN.

Chollet, F., 2015, Keras, https://keras.io

El-Genk, M.S., Palomino, L.M., 2019, “A Walk-Away Safe, Very Small, Long-LIfe, Modular (VSLLIM) Reactor for Portable and Stationary Power,” Annals of Nuclear Energy, 129, pp. 181-198.

El-Genk, M.S., Schriener, T.M., 2022, “A Cybersecurity Platform for Simulating Transient Responses of Emulated Programmable Logic Controllers in Instrumentation and Control Systems for a PWR Plant,” Journal of Cyber Security Technology, Vol. 6(1-2), 65-90. https://doi.org/10.1080/23742917.2022.2059323

El-Genk, M.S., Schriener, T.M., Palomino, L.M., 2021. “Passive and Walk-Away Safe Small and Microreactors for Electricity Generation and Production of Process Heat for Industrial Uses.” J. of Nuclear Engineering and Radiation Science, 7(3), 031302.

El-Genk, M.S., Schriener, T.M., Shaheen A., 2024, “Training and Testing Machine Learning Algorithms for VSLLIM Reactor Controller,” UBN-ISNPS, Albuquerque, NM, USA.

Goorley, T., 2014, “MCNP6.1.1-Beta Release Notes,” Technical Report LA-UR-14-24680. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996. “Reinforcement Learning,” Journal of Artifical Intelligence, 4, 237-285.

Tang, C., et al., 2022, “Deep Learning in Nuclear Industry: A Survey,” Big Data Mining and Analytics, 5(2), 140-160, DOI: 10.26599/BDMA.2021.9020027

Palomino, L., El-Genk, M.S., Schriener, T.M., 2019, “Post-operation Dose Rate Estimates for the Very-small, Long-life, Modular (VSLLIM) Reactor,” ASME J. Nuclear Engineering and Radiation Science, NERS-19-1045

Paszke, Adam, et al. 2017, "Automatic differentiation in pytorch," In Proceedings of 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, Jan 24, 2018.

Radaideh, M.I., et al., 2022. “Multistep Criticality Search and Power Shaping in Microreactors with Reinforcement Learning,” arXiv:24006.15931, 1-15

Schriener, T.M., El-Genk, M.S., 2022, “Simulated False Data Injection Attacks on Emulated and Hardware Programmable Logic Controllers of the Pressurizer in a Representative Pressurized Water Reactor Plant,” Journal of Cyber Security Technology, 6(4), 216-241.

Sutton, S.S, Barto, A.G., 2018. *Reinforcement Learning, second edition*, MIT Press, Cambridge, MA, USA.

The MathWorks, 2022, MATLAB version 2020b, www.matlab.com

Wang, P., Yan, X., Zhao, F., 2019. “Multi-objective optimization of control parameters for a pressurized water reactor pressurizer using a genetic algorithm,” Annals of Nuclear Energy, 124, 9-20.