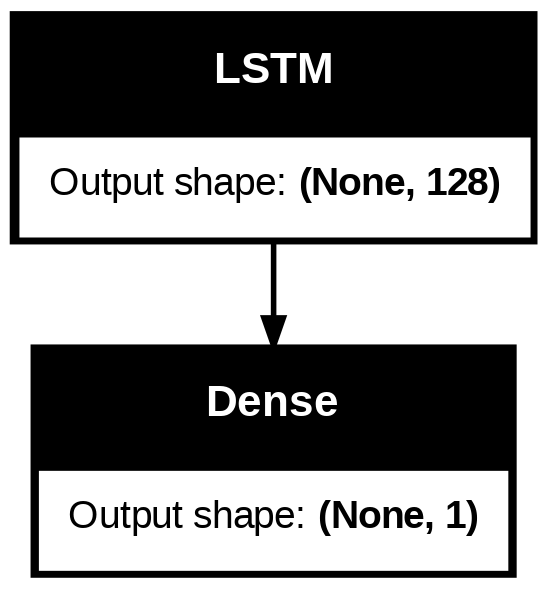
## 3.4 Artificial Intelligence Models

In this section, we used TensorFlow to build and train the machine learning model for the time series analysis of the data, the data preprocessing and evaluation metrics are also discussed.

**Data:** The data for the machine learning algorithm were obtained from the experimental investigations as indicated in figure 1. The datasets contain the amount of hydrogen produced at different light intensities and photoperiods over time. The first contains the quantity of hydrogen produced at light intensities of 100, 200, 300 µmolm^-2 s^-1, whereas the second dataset contains the quantity of hydrogen produced at photoperiods (day:night) 4:20, 08:16, 12:12, 18:06 and 24:00 hours. On each category, the data was preprocessed, normalized and reshaped into the input dimension that is suitable for the neural network. The data was split into train and test sets at 80% and 20% respectively. 20% of the train set was used as the validation during training of the model.

**Neural Network:** The neural network we used in this investigation is Long Short-Term Memory (LSTM). LSTM is a type of recurrent neural network (RNN) architecture that is designed to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequential data. We used a single layer neural network with 128 neurons at the input layer and 1 neuron at the output. We used Relu as the activation function, Adam with a learning rate of 0.01 as the optimizer and finally, we used Mean-Squared-Error (MSE) as the loss function. The network contains a total of 66689 trainable parameters. The summary of the model is presented in figure 1 below.



**Figure 1**

**Performance Evaluation:**  To measure the performance of the model, we used Mean-Squared-Error (MSE) and Mean-Absolute-Error (MAE). Whereas MSE is more sensitive to large errors and outliers, MAE is robust to outliers because it considers the absolute differences. When evaluating a neural network model, it is salient to use both MSE and MAE in order to get a comprehensive view of the model's performance, hence, our choice of both metrics. For reproducibility, the NumPy, Python random module and TensorFlow seeds where set to zero.

**4. Results**

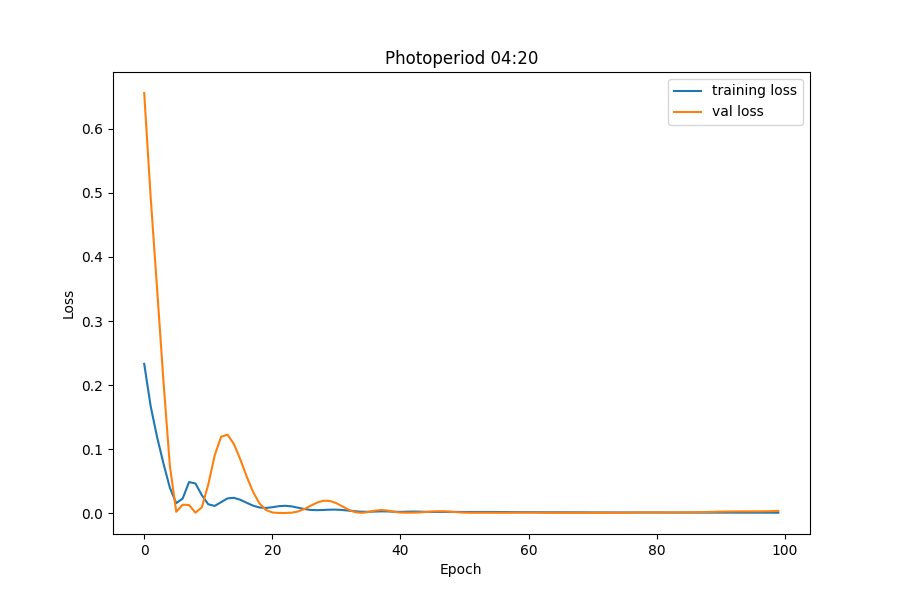
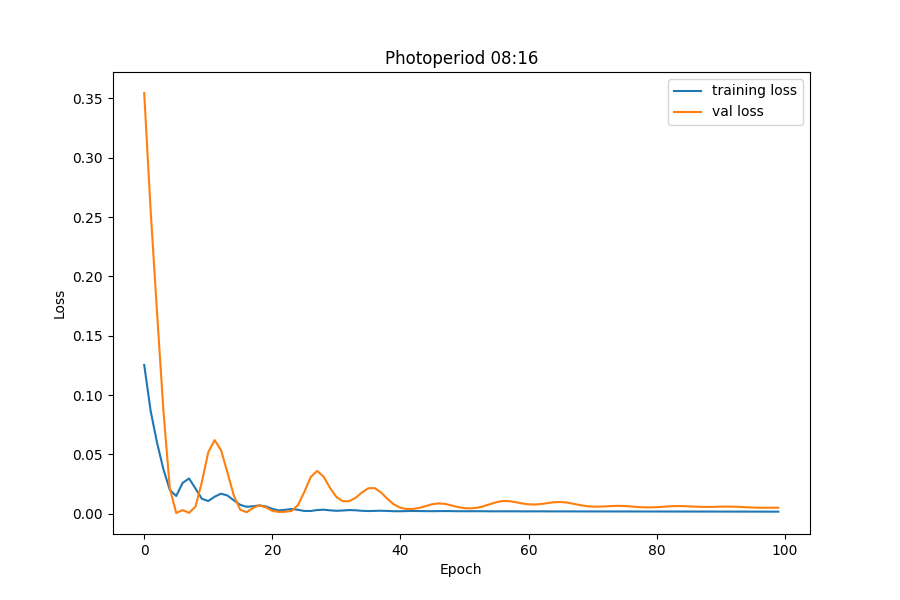
The results of this research using AI model are presented in tables 1 and 2, and figures 1 -5 below. Table 1 contains the average training losses. From this table, the MSE and MAE for both training and validation are significantly reduced. Figures 1-5 presents the actual plot of the losses on every epoch of training. Table 2 presents the performance of the model on unseen data which is the test data.

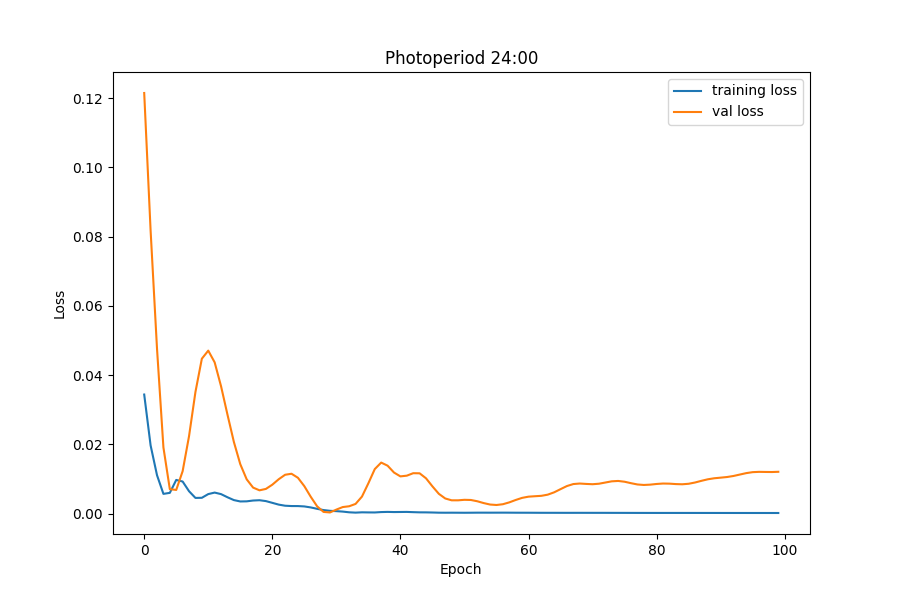
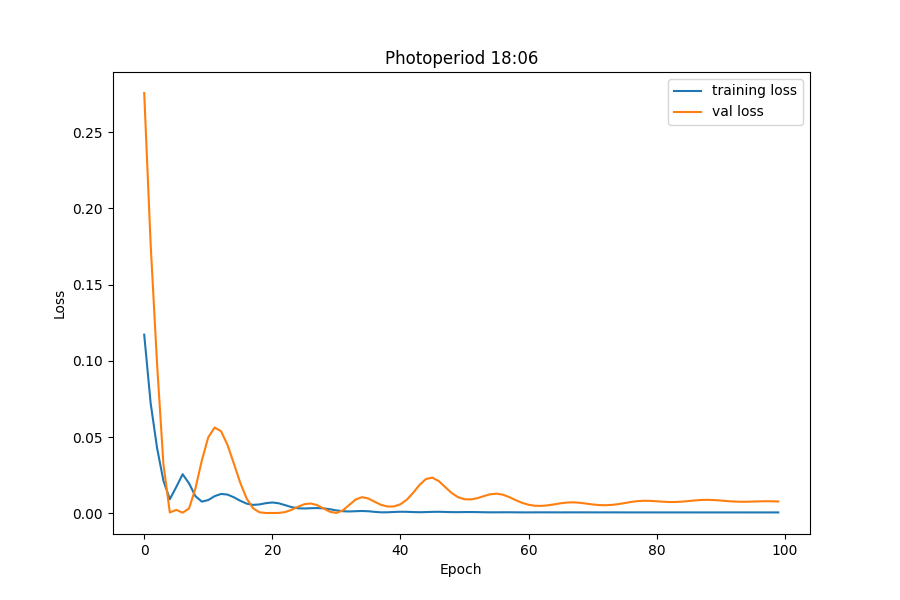
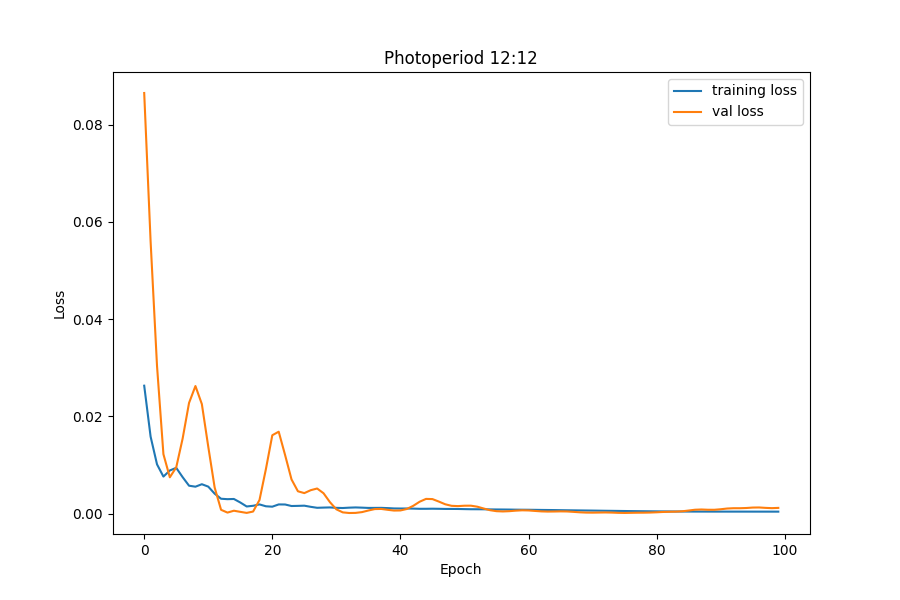
**Table 1: Mean losses during training**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Photoperiod** | **Train MSE** | **Train MAE** | **Val** **MSE** | **Val** **MAE** |
| 04:20 | 0.011532 | 0.063587 | 0.027181 | 0.088448 |
| 08:16 | 0.007136 | 0.054281 | 0.018686 | 0.103339 |
| 12:12 | 0.001954 | 0.026335 | 0.004559 | 0.045523 |
| 18:06 | 0.005247 | 0.042617 | 0.015442 | 0.099627 |
| 24:00 | 0.00192 | 0.025759 | 0.012365 | 0.096761 |

**Table 2: Performance metrics using test data**

|  |  |  |
| --- | --- | --- |
| **Photoperiod** | **Test MSE** | **Test MAE** |
| 04:20 | 0.002562 | 0.050258 |
| 08:16 | 0.00708 | 0.07569 |
| 12:12 | 0.08769 | 0.28837 |
| 18:06 | 0.060682 | 0.24606 |
| 24:00 | 0.032959 | 0.179118 |

****

****

**5. Discussion**

From the results above, the minimal values of MSE and MAE are indicating that our model is efficient. This shows that our aim of using AI to optimize hydrogen production with respect to time is feasible. With this, we can forecast the amount of hydrogen that can be generated over a given time and vice versa. We can also determine the photoperiod that is suitable for a given quantity of hydrogen. The model seems to have lower MSE and MAE at photoperiod 12:12, however, in general, we cannot rule out overfitting and underfitting which could be the consequence of fewer data points for training. Nonetheless, our limitation is that we do not have sufficient amount of data for this project.