Question 1:

a. **Question a)**

1) This is a classification problem and not a prediction problem. The input is readouts from individual sensors which are a time series sequence.

2) An MLP network. The first layer of the MLP would correspond to the input and the output without depending on the system state encoding.

3) The network architecture for the MLP depends on a few things. The input layer would be equally as wide as the input coming from multiple sensors. The sigmoid activation function equal to number of classes. The hidden layers should be configured as a part of model selection however too much would lead to exploding or vanish gradients.

4) The MLP should be trained with back-prop and cross-entropy loss function. Use early stopping or L1 penalty regularisation to minimise it from overfitting.

5) Cross-validation-based estimation of the generalization error

6) Key parameters: Learning rate and the ehize of the hidden layer

7) Check false positives and false negatives. F-scre should be good on unseen data plus cross-validation estimate

8) Given the potential high dimensionality of the network input to exploit the time-delayed presentations, risk of overfitting. Lastly in the data, there is a risk of class imbalance.

b. **Question b)**

1) Anomaly binary classification detection problem. Seems to be heavy class imbalance as it seems that it saved only when the machine was not working.

2) MLP network. The input should be equally as wide as the input and the output is between 1 and 0 or binary values. The output layer should only have two.

3) This would be a feedforward network with hidden layers. Hidden layer size and amount of hidden layers should be determined as part of model selection

4) The training algorithm would be mini-batch backprop with early stopping and cross-entropy loss function

5) Cross-validation-based estimation of the generalization error

6) The key parameters of this network would be the size of the hidden layer and the learning rate

7) Performance evaluation can be done with f1, AUC and other metrics.

8) Class imbalance in the dataset would be a serious challenge as the model only learns from those datapoints

c. **Question C**

1. Generating data seems to be the task. Data augmentations technique.
2. Generative Variational autoencoder would be good here utilising CNN representations
3. The number of inputs and outputs is the same and corresponds to the size of available images. The hidden layers are to be chosen by model selection
4. Backpropagation with **reparameterization trick** and with regularisation. Mean square error as the loss function
5. Cross-validation-based estimation of the generalization error
6. Key parameters would be the size of the hidden layer, the learning rate and the underlying probabilistic model
7. Performance evaluation would be quality, realism and diversity of images
8. Creating hybrid scenarios depends on the capabilities of identifying desirable latent variables and the quality of the latent space

Question 2:

For a Hopfield network with bipolar {1, -1} nodes and the following weight matrix, **W**:

$$\mathbf{W} = \begin{bmatrix} 0 & -1 & -2 & 3 & 4 \\ -1 & 0 & -1 & -2 & 3 \\ -2 & -1 & 0 & -1 & -2 \\ 3 & -2 & -1 & 0 & -1 \\ 4 & 3 & -2 & -1 & 0 \end{bmatrix}$$

please find two patterns that are fixed-point attractors. Show your synchronous mode calculations to prove that the proposed patterns are fixed points, assuming that the node output is 1 if the net input is greater or equal to 0.

Question 3:

This question seems to be requesting two thigs. The first part is to propose a neural network approach to automated morphological labelling of the photos. The second part is to propose an end-to-end neural network-based system for categorising leaves based on their photos. Seems

Task 1) Automated morphological labelling of photos can be done using a CNN model. This is a classification problem but could be a regression problem if the morphological features are numerical instead of categorical. Learning would be done with Dropout with Adam optimizer and Dropout with residual connections depending on how deep the network becomes. Preprocessing is not needed but data augmentation for CNN generally improves training but is computationally expensive. Chalöemge would be that the dataset contains too much of one class but is not balanced out, photos of different sizes, different angles etcetera.

Task 2) Unsupervised task for grouping them rather than classifying. A SOM network could be used for this. The input would be morphological descriptors. The SOM grid could be 2-dim. The learning algorithm is soft competitive learning however you shrink the neighbourhood in in the output space. Key parameters are grid size, learning rate and schedule for neighbourhood adaption

Question 4: - restudy this

Reservoir is a recurrent neural network

To obtain better separations between states, reservoir should be a large network with sparse (1-20%) connectivity. Very often designed as a network with the small world connectivity pattern) and small spectral radius promoting stability

Training echo state networks is fast since it only concerned with training output weights which is a linear problem so we up with least mean square problem

Echo State networks, RNN and LSTM differs in learning/handling time dependencies. In RNN we adapt “non linear” network parameters through training while in ESN these nonlinear parameters involved in recurrent processing are fixed and only the linear read-out weights are adjusted with linear methods.

LSTM help in addressing the problem of vanishing gradients which RNN suffers from. This allows them to learn and handle longer sequence. In this regard, gating units in LSTM cells play a critical learn

Question 5)

1) .

Hopfield network with Hebbian learning

2)

To make the network “remember” patterns, the input corresponds to the memory pattern to be encoded - a combination of -1 and 1. With the Hebbian rule, the weights of the networks are adjusted to store the patterns.

3) .

The network capability is roughly 0.138N where N is the number of patterns. The capacity depends on how similar the patterns they store are to each other, network size and the variation of the learning rule. To measure capability, first, input your patterns then feed it random patterns until catastrophic forgetting happens. **To enhance memory capability for specific data - Refine the learning rate or by extracting and then using input data representations taht lead to lower correlations (cross-talk patterns)**

4) ..

Something called catastrophic forgetting happens. Once the upper boundary as bene reached, the network “forgets” the patterns it stored and the “forgetting” happens suddenly

5) .

Key challenges would be to
- Spurious attractors?
- Limited capacity heavily affected by the input data correlations (cross-talk)
- Limited data encoding approach (Bipolar or binary patterns)

Question 6)

a. Classification problem. MLP or RBF networks. The number of hidden layers and hidden nodes inside the layers are to be decided by model selection for the MLP. For RBF, the number of hidden nodes in the only hidden layer is to be decided by model selection. Input of the model = sinput size and output a sigmoidar.

b. Data representation -- A continuous non-linear function often in signal processing. Similar to a regression problem. Backprop for MLP, Competitive learning for RBF output layer and M

c. 70% training, 20% validation and 10 for testing.

d. Key parameters are: learning rate, hidden nodes, hidden layers. Model selection done with cross validation.

e. Loss function with cross-entropy loss. Measure system level performance with AUC. F1 score etcetera.

f. The node with the highest value is the output.

g. Class imbalance in the total dataset (not limited to the training set). Different microphones would give different data and cause problems. Can increase generalization because of different microphones.