Does the number of basins influence the modeling outcome using ML models?

A systematic methodology to answer the question that is being raised is presented. This method ensures that the underlying hydrology is conserved at the best to answer the question.

Step-1:

Draw an **artificial stream network**. The stream network shall have one limb or line per basin. In other words, if you have 531 basins in your dataset, your stream network will have 531 limbs or lines.

Step-2:

Use one of the existing techniques to order or name the stream developed in step 1. For example, if you are using Strahler's method to order your stream network, you will end up with a network like the one shown below. There are many methods to order or name a stream network. Therefore, you will have to do some research to determine the best method that is suited for the problem that you are solving.
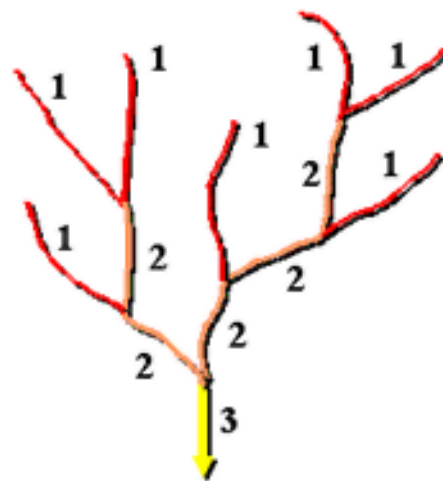
From ESRI's documentarian

Step-3:

Run your ML model, considering all the basins that you have in your dataset. Report the coefficient that is of your interest (e.g., NSE).

Step-4:

Run your ML model, considering the basins after removing the smallest stream number. If you are referring to the above figure, the smallest number will be 1. You can randomly remove the streams from the network.

Step-4-1:

>Assume that you have 50 basins with "1". You can run your ML model after removing those 50 basins. In other words, you will have 531–50=481 basins in your simulation. Report the coefficient that is of your interest (e.g., NSE).

Step-4-2:

>Assume that you have 50 basins with "1". You can run your ML model after removing one of those 50 basins. You can randomly remove the basin from the network. In other words, you will have 531-1=530 basins in your simulation. You repeat this procedure by removing them one by one. In other words, your final simulation will have 531–50=481 basins in your simulation. Report the coefficient that is of your interest (e.g., NSE). You will have an array of NSEs.

Step-5:

>Run your ML model, considering the basins after removing the next smallest stream number. If you are referring to the above figure, the next smallest number will be 2. You can randomly remove the streams from the network.

Step-5-1: Refer to Step-4-1

Step-5-2: Refer to Step-4-2

Step-6:

Continue your simulation work until you reach the largest stream number in your network. If you are referring to the above figure, the largest stream number will be 3.

Step-7:

Plot your NSEs and answer the question that has been raised.

In summary, a reverse algorithm needs to be developed to address the problem. Moreover, being conversant with spatial operations using products from ESRI or other vendors is required.