Main Comments:

- The contributions of the work in the context of existing literature is not clear. Presently, the contributions stated are:

(1) a framework for traffic flow prediction, (2) an optimal input data selection algorithm, and (3) a CNN prediction model based on spatial-temporal correlations.

(1) & (2) are not contributions to knowledge. (3) can be a contribution provided it is established through literature. So, please rephrase and be specific about the contributions.

- The matrix presentation methods described for choosing the temporal and spatial input points seems heuristic /arbitrary. The ACF plot presented is for non-stationary data and 0.68 cannot be arbitrarily chosen as a cut-off! The spatial correlation method described is quite vague. How are the datasets are actually matched? Pearson's correlation? What about lagged spatial correlation??

- In figure one the prediction methods contain a few boxes which do not show any sequence details of the process. It is necessary to provide more information in significantly more details on how the boxes are linked and what happens inside a box. For eg. 'Determine network structure' Does not give information about which network, how to structure is determined, what elements of the structures are determined at every step of prediction or during the training stage. It is also unclear whether all these steps are necessary to be followed during the production process or just during the training of the network.

- How is the initial correlation analysis for matrix presentation is linked with STFSA? Does it provide/generate $R_{initial}$? Why two steps are necessary? Why not start with STFSA directly rather than using arbitrary correlation analysis?

- Fig 6 presents some confusing results. It shows that with longer length of data the prediction accuracy decreases. This is counterintuitive. The MAPE should show that prediction accuracy increases when input data length increases between 5-20mins and then it levels out for further data points. Again, if the input length in increased significantly to up to 24 hrs or longer, we should see another improvement in prediction accuracy. Longer length of data should not affect the prediction accuracy. For spatial sections, it was initially described only 8 neighboring detectors were used…however the results are presented for nearly up to 30 detectors. Please explain.
- Apart from the aforementioned issues, the rest of the results seem appropriate. But the STFSA which is claimed to be the main contribution of the work, does not seem to make a major difference to the MAPE error of the ANN or SVM accuracies published elsewhere. I understand that the model accuracy is dependent on the data characteristics to some extent, however STFSA framework is not making any significant difference.
- The input data size for ANN, SVR and CNN as it is needs to be presented and compared with ANN+STFSA, SVR+STFSA and CNN+STFSA. Computational time and efficiency needs to discussed.

Minor comments:

The paper should be organized better. STFSA should be presented before other methods. Also, no need to give so much details of CNN, ANN or SVR. Provide appropriate reference.

The phrase 'time correlation' should be changed to 'temporal correlation'.

Figure 1: There are some typos in this figure. The word person should be changed to Pearson.