1. What are the datasets the authors used for the experiment?

= For the experiment, the authors made use of a dataset of 141 subjects. The data originated from three prospective studies performed at the Paris Brain Institute between May 2009 and September 2017. Among those 141 subjects there were 44 healthy controls, 61 patients with relapsing remitting MS and 36 patients with progressive MS (13 patients with primary progressive MS and 23 patients with secondary progressive MS). I could not find the dataset so I think the authors have not made the dataset public.

2. What are the differences of Axial MLP with traditional MLP?

= A multilayer perceptron (MLP) is a type of artificial neural network (ANN) composed of multiple layers of perceptrons. Multilayer perceptrons are sometimes referred to as "vanilla" neural networks, especially when they have a single hidden layer. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

In case of Axial MLP, the authors have applied multiple MLPs axially like it is used in Axial Attention Transformers. This architecture facilitates information flow from different axial direction. In this experiment, the authors have used 6 MLPS in each of the L layers of their model. The results are passed through the activation function and summed. Then it is normalized across all but the batch dimension. The use of multiple layers of axial MLPs (L\*6 MLPs) is mainly what is different from a traditional MLP which is simpler in comparison.

3. What do you understand by the term CP (choroid plexus) segmentation?

= In brain MRI analysis, image segmentation is commonly used for measuring and visualizing the brain's anatomical structures, for analyzing brain changes, for delineating pathological regions, and for surgical planning and image-guided interventions. In case of CP (Choroid Plexus) segmentation, image segmentation is done on the patients’ brain MRI images to facilitate the study of the Choroid Plexus which will help in finding abnormalities. Choroid Plexus are structures of the brain ventricles which produce most of the cerebrospinal fluid (CSF). Several postmortem and in vivo studies have pointed towards their role in the inflammatory process in Multiple Sclerosis (MS). Automatic segmentation of CP from MRI thus has high value for studying their characteristics in large cohorts of patients.

4. What authors performed to automatically segment CP from non-contrast enhanced T1-weighted MRI?

= The only freely available tool for automatic CP segmentation is FreeSurfer which has poor accuracy. The authors proposed to automatically segment CP from non-contrast enhanced T1-weighted MRI using their newly introduced Axial-MLP model which is based on an assembly of Axial multi-layer perceptrons (MLPs). This was inspired by recent works which showed that the self-attention layers of Transformers can be replaced with MLPs. The authors compared their approach with a standard 3D U-Net, nnU-Net, FreeSurfer and FastSurfer. Their model outperforms FreeSurfer.

5. Which other nets are competitive with the Axial MLP net? Why are they considering it as a competition?

= The Axial MLP net was competitive with the state-of-the-art segmentation methods and achieved comparable results to those of 3D U-Nets and nnU-Net which are convolutional neural networks (CNN). Although the performance of the Axial-MLP as assessed by Dice coefficient was slightly lower than these two CNN models, the authors believe that, future improved MLP-based architectures have the potential to outperform convolutional models.

From the cross validation (CV) table, we see that the proposed Axial-MLP 8 and 3D U-Net 8 offered the best CV accuracy vs training time compromise. In fact, the 3d U-Nets were slightly better than Axial-MLP in terms of Dice coefficient, but their number of parameters was higher. On the other hand, nnU-Net performed slightly better on almost all metrics but the computation time for training was considerably higher. Even though nnU-Net required considerably more computational time, the improvement in performance improvement was not substantial compared to the 3D U-Net.

However, using MLPs to encode spatial information re-quires fixing dimension of inputs, which makes it difficultto be deployed on downstream computer vision tasks – suchas object detection and semantic segmentation – since theyusually require arbitrary resolutions of input sizes. Fur-thermore, single-stage design, following ViT, may constrain performances on object detection and semantic segmenta-tion since they make predictions based on feature pyramids.Large consecutive MLPs also bring heavy computation bur-den and more parameters, with high dimension of hiddenlayers.we propose ConvMLP:A Hierarchical Convolutional MLP backbone for visualrecognition, which is a combination of convolution lay-ers and MLP layers for image classification and can beseamlessly used for downstream tasks like object detectionand segmentation as shown in Figure 1. To remove con-straints on input dimension in other MLP-like frameworks,we first replace all spatial MLPs with channel MLPs forcross-channel connections and builds a pure-MLP baselinemodel. To make up spatial information interaction, we add light-weight convolution stage on top of the rest MLPstages and use convolution layers for down-sampling. Fur-thermore, to augment spatial connections in MLP stages,we add a simple 3 × 3 depth-wise convolution between thetwo channel MLPs in each MLP block, hence calling it aConv-MLP block. This co-design of convolution layers andMLP layers builds the prototype of ConvMLP model for im-age classification.