0.1 Comparison with other Pre-training techniques

The proposed bin-wise pre-training model was compared with the existing pre-training models: Stacked Autoencoders and Supervised Greedy layer-wise pre-training model. Stacked Autoencoder was configured with three stacks of encoder-decoder, whereas Supervised Greedy layer-wise pre-training model was shaped with six layers added sequentially.
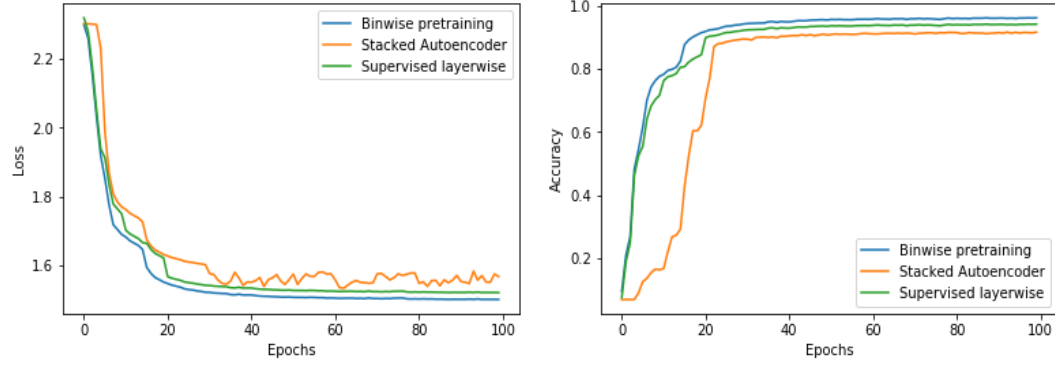

Fig 10(a) & Fig 10(b) shows the validation loss and Validation accuracy of Pretraining methods for MNIST dataset respectively
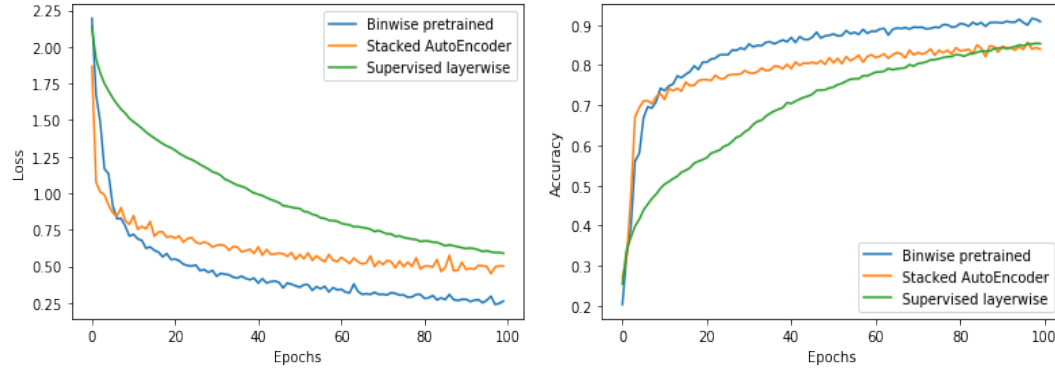

Fig 11(a) & Fig 11(b) depicts the validation loss and Validation accuracy of Pretraining methods for Fashion MNIST dataset respectively
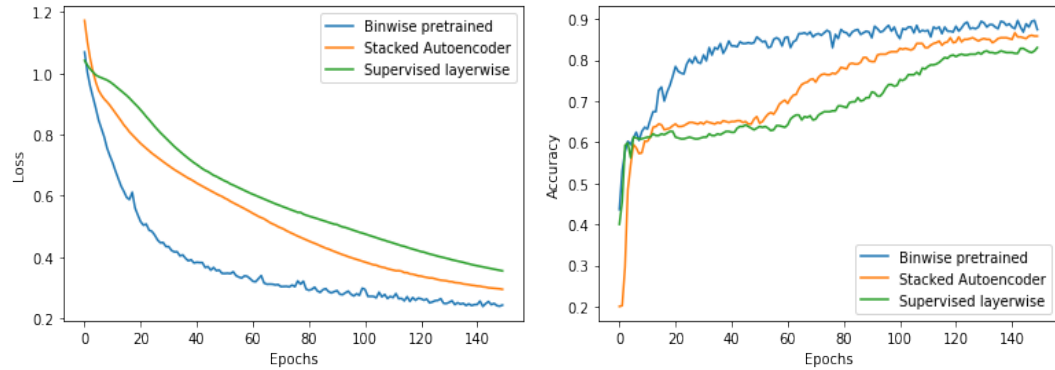

Fig 12(a)

1

& Fig 12(b) plots the validation loss and Validation accuracy of various Pretraining methods for University of Bern EEG dataset respectively

0.2   Regularization Capability

To evaluate the performance of the proposed work in Regularization perspective, the distribution of randomly initialized weights and weights after bin-wise pre-training were recorded and the change imbibed on them due to bin-wise pre-training were analysed. It was inferred that the weight boundaries were maintained within 0.04 and the variance of both the distribution remains the same throughout the pre-training process which confirms that the proposed model behaves as a good regularizer. Fig. 13. shows the comparison of distribution of randomly initialized weights and pre-trained weights. Fig. 14. depicts the pattern of weight updation during pre-training process and it was inferred that the new weights obtained had the desirable distribution, thus making the Deep Learning model simple with better Generalization behaviour along with less susceptible to overfitting and Vanishing or Exploding gradient problems.
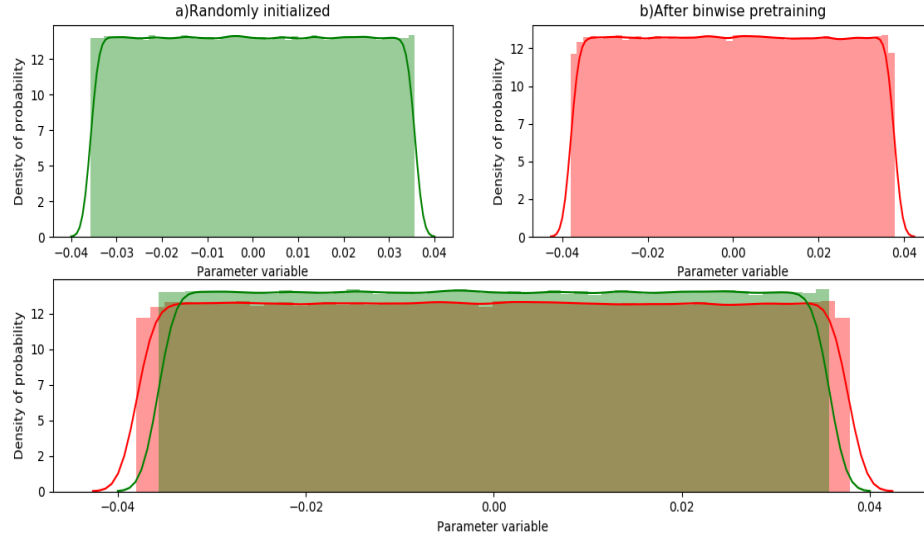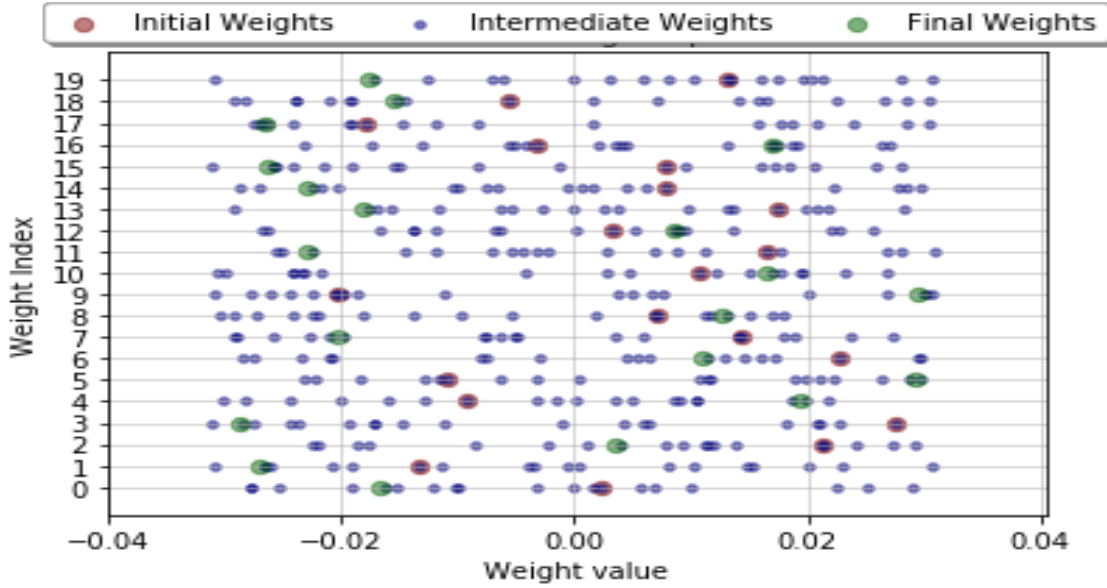
Fig. 14. Weight updation pattern of randomly selected twenty weights for twenty iterations during pre-training process.

0.3 Convergence Rate

Rate of convergence is an important metric that measures the ability of the model to converge in better minima with reasonable amount of time. From Fig. 7, 8 & 9 it has been inferred that the rate of convergence for all the existing approaches were mostly similar at the beginning and as the iteration proceeded the proposed unsupervised bin- wise pre-training model converges earlier than others. Selecting appropriate number of bins using Partial Information Decomposition alters the degree of change and directly influences the convergence rate. Fig. 15. shows the increase in Mutual Information between Neurons output ( ) and Input data ( ) as the iterations were proceeded.