# Deep Learning Architectures for Time Series forecasting not based on Transformers

compiled by D.Gueorguiev 6/8/2024

## Introductory Notes

### Mixing Architectures as an alternative to Transformers

The first such architecture *MLP-Mixer* is presented in [1]. MLP-Mixer relies entirely on MLP networks which are applied repeatedly across spatial locations or feature channels. MLP-Mixer does not use convolutions or self-attention. MLP-Mixer relies only on basic multiplication routines, changes to data layout (reshapes and transpositions), and scalar nonlinearities.

Figure 1 depicts the architecture of a MLP-Mixer. MLP-Mixer accepts as an input a sequence of linearly projected image patches (also referred as tokens) shaped as a “**patches** **channels**” table as an input and maintains this dimensionality. Mixer makes use of two types of MLP layers: *channel-mixing MLPs* and *token-mixing MLPs*. The channel-mixing MLPs allow communication between different channels; they operate on each token independently and take individual rows of the table as inputs. These two types of layers are interleaved to enable interaction of both input dimensions.

A diagram of a process

Description automatically generated

Figure 1: MLP-Mixer consists of per-patch linear embeddings, Mixer layers, and a classifier head. Mixer layers contain one token-mixing MLP and one channel-mixing MLP, each consisting of two fully-connected layers and a GELU nonlinearity. Other components include skip-connections, dropout and layer norm on the channels.

## References

[1] [MLP-Mixer: An all-MLP Architecture for Vision, Ilya Tolstikhin et al, Google, 2021](https://github.com/dimitarpg13/deep_learning_for_time_series_forecasting/blob/main/literature/articles/MLP-Mixer-An_all-MLP_Architecture_for_Vision_Tolstikhin_Google_2021.pdf)

[2] [A decoder-only foundation model for time-series forecasting, A. Das et al, Google, 2024](https://github.com/dimitarpg13/deep_learning_for_time_series_forecasting/blob/main/literature/articles/A_decoder-only_foundation_model_for_time-series_forecasting_Das_Google_April_2024_Preprint.pdf)

[3] [Tiny Time Mixers (TTM): A Powerful Zero-Shot Forecasting Model by IBM, Nikos Kafritsas, 2024, Towards Data Science](https://towardsdatascience.com/tiny-time-mixers-ttm-a-powerful-zero-shot-forecasting-model-by-ibm-576b0e0af583)

[4] [Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series, V. Ekambaram et al, IBM, 2024](https://github.com/dimitarpg13/deep_learning_for_time_series_forecasting/blob/main/literature/articles/Tiny_Time_Mixers-Fast_Pre-trained_Models_for_Enhanced_Zero_Few-Shot_Forecasting_of_Multivariate_Time_Series_Ekambaram_IBM_2024.pdf)

[5] [TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting, V. Ekambaram et al, IBM, 2023](https://github.com/dimitarpg13/deep_learning_for_time_series_forecasting/blob/main/literature/articles/TSMixer-Lightweight_MLP-Mixer_Model_for_Multivariate_Time_Series_Forecasting_Ekambaram_IBM_2023.pdf)

[6] [TimesFM: Google's Foundation Model For Time-Series Forecasting, Nikos Kafritsas, 2023, AI Horizon Forecast](https://aihorizonforecast.substack.com/p/timesfm-googles-foundation-model)

[7] [MOIRAI: Salesforce's Foundation Transformer For Time-Series Forecasting, Nikos Kafritsas, 2023, AI Horizon Forecast](https://aihorizonforecast.substack.com/p/moirai-salesforces-foundation-transformer)

[8] [PatchTST for Time Series Forecasting: Original Results and My Single-Channel Experiments, Lalf Klein, 2023, Medium](https://medium.com/@lalf_klein/patchtst-for-time-series-forecasting-original-results-and-new-single-channel-experiments-f375699f7b91)

[9] [A Time Series is Worth 64 Words: Long-term Forecasting with Transformers, Y. Nie et al, IBM, Princeton U., 2023](https://github.com/dimitarpg13/deep_learning_for_time_series_forecasting/blob/main/literature/articles/A_Time_Series_is_Worth_64_Words-Long-term_Forecasting_with_Transformers_Nie_IBM_2023.pdf)