

Addressing Bias, Statistical Validity, and Reproducibility in ML Models for Physiological Signal Processing.

Time	Topic		
	Title	Nature	Presenter
9:00-9:10	<ul style="list-style-type: none"> - Introduction to the use of ML in physiological signal processing, discussing the growth of research and projecting future trends. - Explaining the dependency of successful ML models on statistical concepts 	Talk, PPT	S.B.
9:10-9:25	<ul style="list-style-type: none"> - Commonly overlooked points in building ML models: a case study 	Hands-on article examination*	M.N.
9:25-9:30	<ul style="list-style-type: none"> - The scientific method and “search for truth”. - Workshop outline (successful ML model construction, statistical validity, and reproducibility.) 	Talk, PPT	S.B.
Machine Learning			
9:30-9:35	<ul style="list-style-type: none"> - Overview of end-to-end machine learning pipeline and where problems might arise. 	Talk, PPT	N.I.
9:35-10:05	<ul style="list-style-type: none"> - Considerations associated with data collection: A case study of EEG dataset. <ul style="list-style-type: none"> o Data collection: representativeness of sample and outliers. o Data preparation: mis-imputation, knowledge-leakage and not addressing bias. 	Hands-on dataset exploration**	M.N., N.I.
10:05-11:30	<ul style="list-style-type: none"> - Considerations associated with building ML models: A case study of EEG dataset. <ul style="list-style-type: none"> o Choosing a baseline. o Choosing classifiers: validity of assumptions and effect on performance. o Why validation is necessary and why testing is not sufficient. o Preferred methods of validation o Internal and external validity. o The tradeoff between overfitting and underfitting. o Model misspecification and choice of loss function o Performance reporting: minimum requirements and preferred metrics. 	Hands-on code examination^	M.N., N.I.
11:30-11:35	<ul style="list-style-type: none"> - Closing notes and distributing a one-page guideline handout based on this section. 	Talk	S.B.
Statistical Validity			
11:35-12	<ul style="list-style-type: none"> - Hypothesis-driven and data-driven Research: is this the correct categorization? - Statistical considerations <ul style="list-style-type: none"> o Power analysis o Correlation assumptions o P-hacking 	Talk, PPT	N. I.
Research Reproducibility			
12:00-12:10	<ul style="list-style-type: none"> - Replicable and reproducible research, a crisis? - Reproducibility requirements: adequate research design and adequate reporting 	Talk, PPT	M.N.
12:10-12:50	<ul style="list-style-type: none"> - Tips and tricks for programmatic reproducibility: A case study <ul style="list-style-type: none"> o Saving environmental dependencies o Saving and restoring states o Version control, model versioning and data versioning 	Hands-on code examination^	M.N., N.I.
12:50-13:00	<ul style="list-style-type: none"> - Closing notes and distributing a one-page guideline handout based on this section. 	Talk	S.B.