Continuous, Interpretable, Minimalistic Machine Learning (CIM-ML): Semi-causal Modelling of Electro-Optical BIG DATA

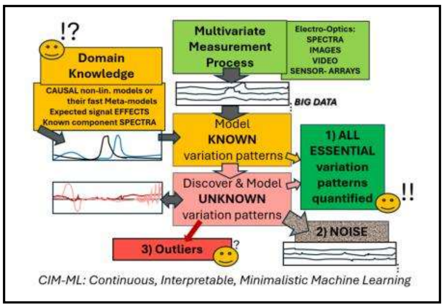
Harald Martens[1,2]

[1] Norwegian U. of Science and Technology NTNU, Dept, Engineering Cybernetics,
O. S. Bragstads plass 2D, N-7034 Trondheim (Harald Martens - NTNU)
[2] Idletechs AS, Teglbrennerveien 7c, N-7016 Trondheim, Norway / Idletechs Sarl, Rue Marconi 19, CH-1920 Martigny, Switzerland
The CIM-ML project in Idletechs is partly financed through the Swiss STREAMSTEP project. E-mail: harald.martens@idletechs.com.

Modelling needed: Compared to classical **univariate** measurements, **multivariate** measurements (whole spectra, images and videos, sensor arrays) give new opportunities. Modern electro-optical measuring instruments are very information rich. They have increasing spectral, spatial and temporal resolution, have smaller size and lower cost, and find increasing use in many discplines. But the raw spectral- or imaging-data of intact materials often have selectivity problems. So they need mathematical modelling, statistical validation and human interpretation in order to give reliable and relevant information.

Which data modelling is most suited? Traditional mechanistic mathematical models are important but sometimes over-simplified, yet computationally cumbersome. ANN/CNN AI solutions are flexible and generic, but often difficult to train, to validate and to understand. How to combine the best - and avoids the worst – of nonlinear mechanistic mathematical modelling and black box ANN/CNN deep learning?

The **Continuous, Interpretable, Minimalistic Machine Learning** (CIM-ML) from Idletechs AS is a new, generic ML methodology and software, for simpler, safer, greener, more understandable and cost-effective problem solving and system control in technical and scientific settings [1, 2**]**. CIM-ML **works** because measured data from the **material** domain of real-world, physical objects, systems or landscapes - complex as they can be - are far simpler than data from the **immaterial** domain of e.g. natural language.



The CIM-ML approach is particularly well suited for electro-optical BIG DATA, where some – but not all- of the causal relationships in the data are already known. The ML is initialialized from domain knowledge. Simple self-modelling then splits the overwhelming stream of multivariate inputs into three separate output streams:

<u>1) Relevant **ESSENCE:**</u> All systematic variation patterns, expected or unexpected, found in the stream of input data. These quantified variations in e.g. light scattering, light absorption and shadows are combined statistically and used for classification, quantification, anomaly detection, and human process interpretation and -control.

<u>2) Irrelevant **NOISE:**</u> More or less random measurement errors, summarized statistically and used for later CIM-ML optimization.

<u>3) **OUTLIERS:**</u> Instrument issues, operator errors, anomalies etc, potentially signifying dangerous problems.

As new variation patterns arise and are discovered, the continuous CIM-ML process is **self-extending** and **self-correcting**. The modelling results are efficiently compressed, which gives efficient data transmission, storage and usage. Its local, open-ended linear and nonlinear structure allows the CIM-ML to combine prior knowledge and new data. Its continuous, semi-causal deep learning can describe a wide range of linear and nonlinear relationship types in the e.g. high-speed diffuse spectroscopy data [3]. And the CIM-ML solutions are simple and low-dimensional, and thus easy to validate statistically and interpret graphically, even in their compressed state. So, this **machine learning** also stimulates **human learning**.

The spectral, mathematical and statistical methodology behind the CIM-ML has been cited more than 28000 times (HARALD MARTENS - Google Scholar). CIM-ML implementations have now been industry hardened in many real-world applications. Although still under development, CIM-ML software already offers cost-efficient and humane AI for r*eal-world science and technology*, in particular for electro-optical measurement applications.

[1] Martens, H. (2015) Quantitative Big Data: where chemometrics can contribute.
Journal of Chemometrics, **Volume 30, Issue 9** pp 563–581; DOI: 10.1002/cem.2740J.
[2] Martens, H. (2025) : A greener, safer and more understandable AI for natural science and technology.
Journal of Chemometrics **Volume 39, Issue 2** Perspective paper; DOI: 10.1002/cem.3643.

[3] Martens, H. (2025). The Lure of Curvature. In: Bec, K., Huck, C. (eds) Proceedings of the 21st International Conference on Near Infrared Spectroscopy. ICNIR 2023. Springer, Cham. https://doi.org/10.1007/978-3-031-84794-3_5.