AI in CMT

Wolfgang M. Pernice, PhD

Columbia University Medical Center, New York, USA; ECMTF

Key words: artificial intelligence, diagnostics, inherited neuromuscular diseases, CMT, machine learning, electrodiagnostics, deep-learning, MAVERICK, GENESIS, ClinVar, Musclesense, outcome measures, standardized datasets

ECRA's perspective on AI in CMT: where are we now?

Alongside related data sharing (D.3.1) and digital care (D.6.1) deliverables, the 2nd European CMT Specialist Conference (EUCMTSC) in Antwerp also sought to identify opportunities and current barriers to the adoption of deep-learning and artificial intelligence (AI) methods in CMT research. Across all stakeholders, including scientists, clinicians, and patients, considerable enthusiasm for the potential of these methods to address unmet needs - especially in diagnostics and clinical assessment of CMT - exists alongside caution, in light of both scientific risks, as well as open questions regarding the protection and fair use of patient data in the training and evaluation of models. Despite several promising examples of deep-learning methods enabling scientific advances in our field, the adoption of AI in CMT remains limited. Informed by the results of the 2025 EUCMTS conference, this report seeks to provide an overview on the state of AI in CMT, along with a preliminary set of strategic recommendations for the field.

Examples of AI in CMT research:

For the sake of clarity, we distinguish between classical *machine learning*, *deep learning*, and *AI*. While the former two describe a set of methods, *AI* generally describes a type of generative Large Language Models (LLMs) or Vision Language Models (VLMs) that most people are familiar with in the form of OpenAI's ChatGPT, Google's Gemini, or Anthropic's Claude line of consumer-oriented models. Importantly, under the above definitions, there currently is no example of a dedicated "AI model" in CMT. However, both classical machine learning and deep learning approaches have already shown promise in clinical diagnostics, genetic diagnostics, and in the development of outcome measures for CMT.

1. Clinical diagnostics – Machine Learning Assisted Neurophysiology: Nerve condition studies (NCS) and surface electromyography (EMGs) are a cornerstone of clinical diagnostics in CMT.[1] But the interpretation of electrophysiological data is laborious and depends heavily on the expertise and experience of the evaluator. Automating the interpretation of electrophysiological data in CMT could mitigate subjectivity in the analysis towards more objective and data driven precision electrodiagnostics. Both classical machine and deep learning

methods have shown promise, e.g. in the discrimination of healthy from neuropathic and myopathic patterns.[2,3] In CMT, Fagundes et al (PNS 2025) reported strong performance of classical Random Forests in CMT subtype classification (CMT1A, CMT1B, males with CMTX1 and CMT4C) based on electrophysiological features, and – as presented by Dr. Tomaselli (PL2-03) – for gradient boosting systems in distinguishing inflammatory from inherited neuropathies. Importantly, however, most studies are limited by relatively small NCS / EMG datasets (50-250 patients)[4]. This may explain the relative underperformance of deep learning methods such as Convolutional Neural Networks (CNNs) thus far: fully leveraging the power of deep learning approaches may well require 100-1000x more data. More importantly, current datasets typically originate from a single clinic / laboratory. The extent to which these early models generalize to data collected by others therefore remains fundamentally unknown. To realize (and properly assess) the promise of computational electrodiagnostics for CMT the collection of large, standardized, and multi-center training and benchmarking datasets should be a priority.

2. Genetic diagnostics – Variant Effect Prediction: Despite the discovery of now >130 genes that have by now been causally associated with CMT and related disorders, the yield of clinical genomic diagnostic pipelines often remains below 50%, especially for patients with axonal presentation.[5] Importantly, a large part of this diagnostic gap may not be due to a lack of identified candidate mutations, but due to a lack of methods to discriminate benign from pathogenic variants. This problem is by no means unique to CMT. In turn, numerous deep-learning approaches to predict the effect of genetic variants have been developed, most of which demonstrated marked improvements over non-deep-learning methods.[6–11]

Among this growing portfolio of methods, MAVERICK[7] stands out as the only model with a clear track record of utility in CMT and related disorders: as part of the GENESIS platform[12] (which hosts the largest collection of CMT exomes and genomes in the world), MAVERICK has enjoyed widespread use in the CMT genetic research community since 2023, and has contributed to the discovery of several new disease genes.[13–16] MAVERICK's adoption in the field may also be owed to the authors careful evaluation procedures that included not only a time-cutoff for variants included in training (only ClinVar variants added prior to 2020) vs. test sets (only ClinVar variants during 2020), but also stratification between genes with variants in the training set vs. novel genes that were not represented in the training set.

Still, as a supervised model trained on ClinVar annotations, it is likely that MAVERICK encodes biases related to the types of variants for which annotations are currently available. Self-supervised VEP models, such as EVE[8,10] or ESM1b[9], instead employ generative pre-text tasks (e.g. to predict the next amino acid in a protein sequence) to learn from the distribution of naturally occurring sequences, which should render them more useful in the discovery of novel pathogenic variants and disease genes in CMT. Importantly, even though self-supervised VEP models are thus trained without seeing any ClinVar annotations at all, self-supervised VEP models have yet been shown to outperform several other supervised VEP methods on ClinVar prediction tasks.[8–10] A direct comparison between MAVERICK and state-of-the-art self-supervised VEPs has not yet been published.

Regardless of model type, however, clear evidence of a modern VEP model making a decisive difference in the discovery of a new disease gene or variant in CMT has yet to emerge.[13–16] Indeed, all modern VEP models still exhibit high false-positive rates, such that the evidentiary burden in genetic variant prioritization for now largely remains on classical segregation based strategies within and across patient families. An important gap in the field remains

the lack of suitable CMT-specific benchmarking datasets, the likes of which have consistently proven a major accelerator of progress across all domains of deep-learning and AI. In turn, a well-defined, curated benchmark dataset could provide crucial direction and drive for progress in deep learning based VEP for CMT. How such a benchmark should be designed remains an important open question.

3. Outcome measures – deep-learning powered biomarkers and assessments: A major persistent barrier to clinical trials in CMT and related disorders is the lack of clinical outcome assessments (COAs) that are sufficiently sensitive to realistically support trials outside of the most common CMT subtypes.[17] Deep-learning methods have already delivered exciting results on this frontier as well.

Possibly the most mature example is the automation of Magnetic Resonance Imaging (MRI) based quantifications of the intramuscular fat fraction (IFF) in the lower limbs of patients with CMT. Over the last decade, lower-limb IFF has emerged a particularly responsive biomarker in CMT,[18] yet it historically required laborious manual segmentation of MRI slices by well-trained analysists. To address this challenge, Kanber, Morrow and colleagues developed Musclesense[19], a deep-learning based segmentation model trained on a relatively large and diverse collection of 571 MRI acquisitions from 166 individuals at both calf and thigh level. Direct comparisons of Musclesense vs. manual segmentation-based assessments of IFF in CMT1A showed very high correlations between both methods.[20] Importantly, both methods can be readily combined such that pre-generated segmentation masks can be checked and refined by human experts. This "human-in-the-loop" workflow dramatically enhances analytic throughput (from 15min to 30s on average per slice), while yet preserving the "human element" that may serve to qualify this procedure in the eyes of regulators that maintain skeptical of fully automated workflows in trials. An important open question is the extent to which Musclesense generalizes to data collected at different sites. Unsurprisingly, when tested against a small and arguable very challenging dataset of MRI images collected at a different center, with a different protocol, and for a different disease, Musclesense exhibited substantial performance degradation.[19] Fortunately, Musclesense remains in active development,[21] and coordinated multi-center studies to systematically evaluate IFF as an outcome in CMT based on standardized protocols and equipment are ongoing.[22]

Both machine and deep-learning approaches also have clear potential to advance capabilities across the rapidly growing field of digital health technologies (DHTs). As outlined in Dr. Schenone's keynote address on "Sensors, Robots and AI" (Plenary 2), the portfolio of DHTs currently under investigation in CMT ranges from pressure-sensing walkways for gait assessment (e.g. GAITRite)[23], to sensor-equipped gloves[24] and wearable inertial measurement units[25]. The Digital Assessment of Natural motion for Clinical Endpoint Research (DANCER) study (Plenary 3) presents a new international effort to validate advanced deep-learning powered computer vision technology for the assessment of gait and other movement patterns in CMT.

Importantly, most ongoing studies in this field collect highly complementary data, including but not limited to gold-standard clinical outcome assessments (COAs) to serve as points of comparison for new biomarkers and DHTs alike, across numerous centers and CMT patient population. Improving upon coordination and data sharing between and across these efforts presents itself as a major opportunity to synergistically advance the development and validation of useful "AI-powered" outcome measures for CMT.

Outlook and recommendations:

The examples above provide clear testament to the potential that modern machine and deep learning approaches hold to accelerate progress across the CMT biomedical research ecosystem. They also reveal a shared set of challenges, strategic solutions to which could substantially strengthen the "AI-readiness" of our field.

The need for rigorous standards of evaluation:

All machine learning approaches carry the risk of overfitting. At worst, such overfitting may entail the effective memorization of training data. More commonly, this risk manifests in the form of biases inherent to the training distribution that models readily encode. Importantly, these biases are generally undetectable unless models are evaluated on independent test datasets[26], such as data collected at different sites (see Musclesense), or by setting date-cutoffs (see MAVERICK). Importantly, the question of how to construct testing regimes that reliably predict real-world performance of deep learning models remains a topic of active research in AI generally. As a baseline however, a clear distinctions between, and methodological details on how training, validation, and testing sets were constructed, are fundamental standards that should be universally adopted in the CMT research as machine and deep learning methods begin to play an increasingly prominent role in our field.

Open benchmarks:

As illustrated by ImageNet as the original catalyst of the ongoing AI-revolution[27], well-designed open-source training and benchmark datasets present a critical component of any mature AI research field. Yet, standardized CMT-specific benchmarks, be it for computational electrodiagnostics, deep-learning powered VEP, or the development of new digital outcomes, are lacking. Establishing suitable benchmarks thus remains an important frontier for our field.

Synergistic data sharing:

The development of powerful clinical and scientific AI-tools for CMT will inevitably require large, standardized and multimodal datasets. The proactive assembly of such datasets, according to a clear set of equitable standards and policies, could dramatically accelerate research progress, while preemptively enshrining fair use principles. ECRA has initiated concrete worksteps on this topic for which we refer the reader to deliverable D.3.1.

References:

1. Charcot–Marie–Tooth disease and related neuropathies | Nature Reviews Disease Primers. https://www.nature.com/articles/s41572-025-00679-2.
2. Taha, M. A. & Morren, J. A. The role of artificial intelligence in electrodiagnostic and neuromuscular medicine: Current state and future directions. *Muscle & Nerve* **69**, 260–272 (2024).
3. Coppotelli, M., Albert, D. & Barrena, J. G. Machine Learning Models for EMG-Based Diagnosis of Neuromuscular Disorders.
4. Messekher, A., Mekaoui, S., Kedir-Talha, M., Kediha, M. I. & Mostefaoui, F. Electromyographic data recordings during isometric contractions of the biceps brachii and deltoid collected from patients and healthy subjects. **1**, (2023).
5. Record, C. J. *et al*. Whole genome sequencing increases the diagnostic rate in Charcot-Marie-Tooth disease. *Brain* **147**, 3144–3156 (2024).
6. Cheng, J. *et al*. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).

7. Danzi, M. C. *et al*. Deep structured learning for variant prioritization in Mendelian diseases. *Nat Commun* **14**, 4167 (2023).
8. Frazer, J. *et al*. Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).
9. Brandes, N., Goldman, G., Wang, C. H., Ye, C. J. & Ntranos, V. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat Genet* **55**, 1512–1522 (2023).
10. Orenbuch, R. *et al*. Proteome-wide model for human disease genetics. *Nat Genet* **57**, 3165–3174 (2025).
11. Gao, H. *et al*. The landscape of tolerated genetic variation in humans and primates. *Science* **380**, eabn8153 (2023).
12. Danzi, M. C. *et al*. The GENESIS database and tools: A decade of discovery in Mendelian genomics. *Exp Neurol* **382**, 114978 (2024).
13. Rebelo, A. P. *et al*. Biallelic variants in COQ7 cause distal hereditary motor neuropathy with upper motor neuron signs. *Brain* **146**, 4191–4199 (2023).
14. Borja, N. A. *et al*. KIF21A-associated peripheral neuropathy defined by impaired binding with TUBB3. *J Med Genet* **62**, 117–122 (2025).
15. Beijer, D. *et al*. A recurrent missense variant in ITPR3 causes demyelinating Charcot-Marie-Tooth with variable severity. *Brain* **148**, 227–237 (2024).
16. Dohrn, M. F. *et al*. Recurrent ATP1A1 variant Gly903Arg causes developmental delay, intellectual disability, and autism. *Ann Clin Transl Neurol* **11**, 1075–1079 (2024).
17. Reilly, M. M. *et al*. Trials for Slowly Progressive Neurogenetic Diseases Need Surrogate Endpoints. *Ann Neurol* **93**, 906–910 (2023).
18. Lower limb muscle MRI fat fraction is a responsive outcome measure in CMT X1, 1B and 2A - PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC10963306/#acn351979-bib-0032.
19. Kanber, B. *et al*. Musclesense: a Trained, Artificial Neural Network for the Anatomical Segmentation of Lower Limb Magnetic Resonance Images in Neuromuscular Diseases. *Neuroinform* **19**, 379–383 (2021).
20. O'Donnell, L. F. *et al*. Quantitative MRI outcome measures in CMT1A using automated lower limb muscle segmentation. *J Neurol Neurosurg Psychiatry* **95**, 500–503 (2024).
21. Release MuscleSense v2.1.0 · bariskanber/musclesense. https://github.com/bariskanber/musclesense/releases/tag/v2.1.0.
22. University College, London. *Muscle MRI in Charcot Mary Tooth Disease: A Prospective Cohort Study*. https://clinicaltrials.gov/study/NCT03550300 (2018).
23. Schenone, C. *et al*. Is gaitrite system sensitive in discriminating gait pattern of subjects affected by Charcot Marie tooth? A pilot study. *Neurol Sci* **46**, 6805–6813 (2025).
24. Alberti, M. A. *et al*. Innovative quantitative testing of hand function in Charcot-Marie-Tooth neuropathy. *Journal of the Peripheral Nervous System* **20**, 410–414 (2015).
25. Dinesh, K. *et al*. Disease-specific wearable sensor algorithms for profiling activity, gait, and balance in individuals with Charcot-Marie-Tooth disease type 1A. *J Peripher Nerv Syst* **28**, 368–381 (2023).
26. Pernice, W. M. *et al*. Out of Distribution Generalization via Interventional Style Transfer in Single-Cell Microscopy. Preprint at https://doi.org/10.48550/arXiv.2306.11890 (2023).
27. Chow, R. ImageNet: A Pioneering Vision for Computers. *History of Data Science* https://www.historyofdatascience.com/imagenet-a-pioneering-vision-for-computers/ (2021).