"This is a copy of a presentation from the 40th Congress of the European Committee for Treatment and Research of Multiple Sclerosis, which was in Denmark; the references to "Merck" or "Merck KGaA" within refer to (1) Merck KGaA, Darmstadt, Germany; (2) an affiliate of Merck KGaA, Darmstadt, Germany; or (3) one of the businesses of Merck KGaA, Darmstadt, Germany, which operate as EMD Serono in the healthcare, MilliporeSigma in the life science and EMD Electronics in the electronics business in the U.S. and Canada.

There are two different, unaffiliated companies that use the name "Merck". Merck KGaA, Darmstadt, Germany, which is providing this content, uses the firm name "Merck KGaA, Darmstadt, Germany" and the business names EMD Serono in the healthcare, MilliporeSigma in the life science and EMD Electronics in the electronics business in the U.S. and Canada. The other company, Merck & Co., Inc. holds the rights in the trademark "Merck" in the U.S. and Canada. Merck & Co., Inc. is not affiliated with or related to Merck KGaA, Darmstadt, Germany, which owns the "Merck" trademark in all other countries of the world."

FOR REACTIVE MEDICAL USE ONLY

Deep Learning methods can predict with high accuracy cladribine tablets treated patients who improve clinical and cognitive disability in 2 years

P204

Marco Battaglini [1,2], Francesco Sforazzini [2], Ludovico Luchetti [2], Francesco Cacciante [2], Giordano Gentile [2], Andrzej Smyk [3], Ali-Frederic Ben-Amor [4], Nicola De Stefano [1,2]

[1] University of Siena, Department of Medicine, Surgery and Neuroscience, Siena, Italy
[2] Siena Imaging, Siena, Italy, [3] Merck Healthcare KGaA, Darmstadt, Germany
[4] Knowlepsy, Marseille Innovation, Technopole de Chateau-Gombert, Marseille, France

INTRODUCTIONS

Highly effective treatments such as cladribine tablets (CladT) can have beneficial effects on disability in patients with relapsing-remitting (RR) multiple sclerosis (MS). Proposed artificial intelligence (AI) methods to predict clinical disability in MS patients, although promising, often suffered by limited generalizability due to the heterogeneity of treatment and MRI protocols.

OBJECTIVES

The purpose of this work is to test the ability of Deep Learning (DL) architectures to predict clinical improvements at two years in RR-MS patients treated with CladT (3.5mg/kg) and compare their performance with prediction obtained using traditional statistical methods.

METHODS

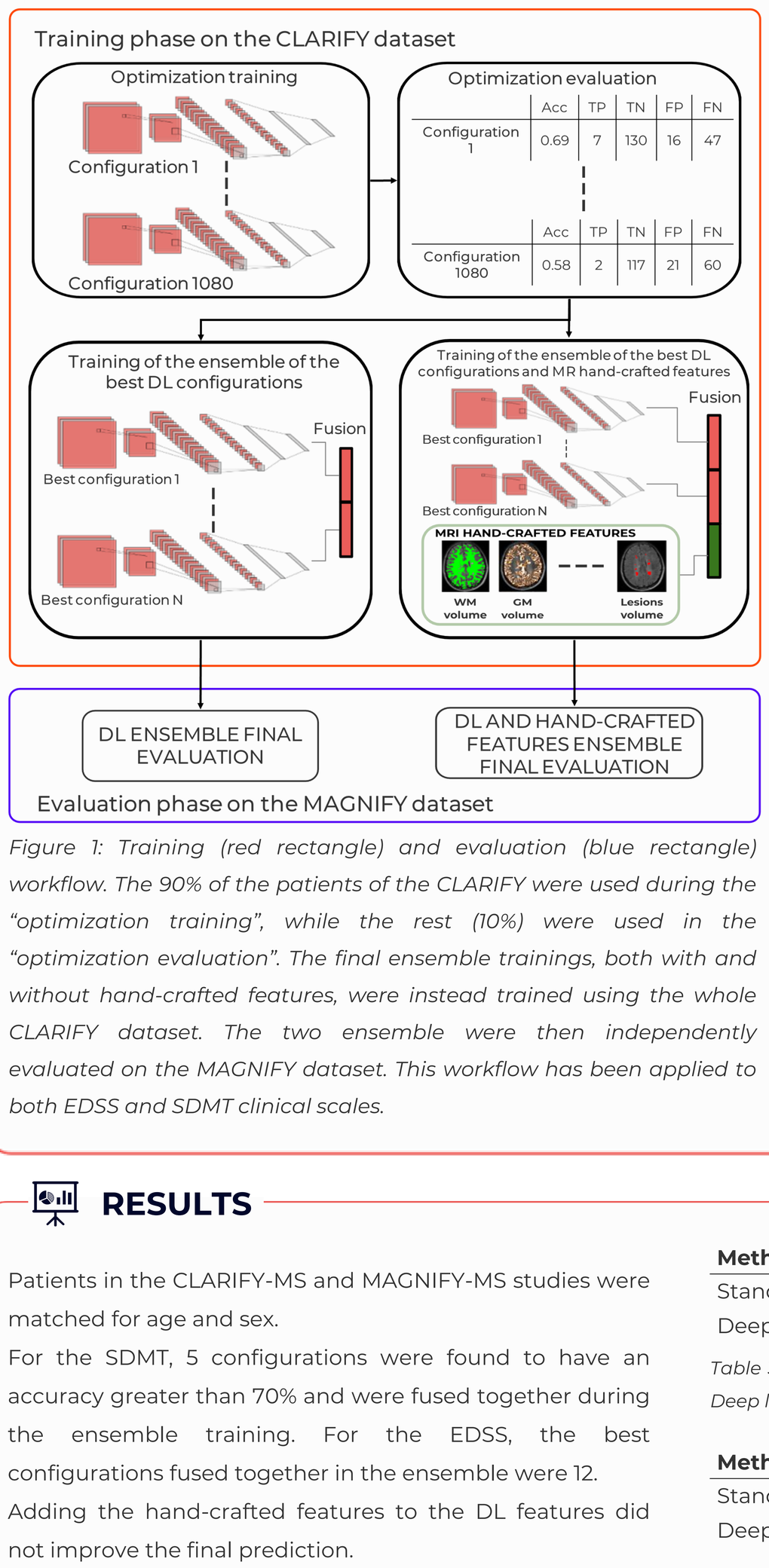

Figure 1: Training (red rectangle) and evaluation (blue rectangle) workflow. The 90% of the patients of the CLARIFY were used during the "optimization training", while the rest (10%) were used in the "optimization evaluation". The final ensemble trainings, both with and without hand-crafted features, were instead trained using the whole CLARIFY dataset. The two ensemble were then independently evaluated on the MAGNIFY dataset. This workflow has been applied to both EDSS and SDMT clinical scales.

The MRI datasets of RR-MS patients enrolled in the phase IV CLARIFY-MS (NCT03369665) and MAGNIFY-MS (NCT03364036) studies, served as training and test groups, respectively. The two trials had the same MRI protocol acquisition.

Over 2 years of follow-up, each patient was labeled as:

- physically improved if EDSS decreased by 0.5, 1 or 1.5 based on the initial EDSS assessment (EDSS<=1, between 1-5 and higher than 5, respectively),
- cognitively improved if SDMT increased by 20% or more.

Clinical scale	CLARIFY - total	CLARIFY - AGE	CLARIFY - improving	MAGNIFY - total	MAGNIFY - AGE	MAGNIFY - improving
SDMT	306 (94 males)	37±10.17	47 (15%)	228 (77)	39±9.79	41 (18%)
EDSS	315 (95)	37±10.49	48 (15%)	230 (77)	39±9.79	34 (15%)

Table 1: Summary of the total number of patients and those improving over a 2-year period (with percentage), for both EDSS and SDMT, in the two cohorts used in this study.

The hand-crafted features for stepwise logistic regressions were volumes of 7 grey-matter regions, lesion volumes, and measures of white matter integrity assessed on 7 regions by T1W/T2W map. The whole DL workflow has been depicted in Figure 1.

Parameter	Values
Network structure	ResNet 10, ResNet 18, ResNet 50
Initial learning rate	0.0001, 0.00001
Optimizer	SGD, ADAM
SGD - weight decay	0.0003, 0.00003
Scheduler	stepLR, polynomial
stepLR - step	5, 10
stepLR - gamma	0.2, 0.5
Input channels	[FLAIR], [T1w, FLAIR], [T2w, FLAIR], [T1w, T2w, FLAIR]
Dropout	0.3, 0.5, 0.7

Table 2: All the combinations between the values of the parameters above (1080) were trained and evaluated during the optimization of the DL algorithm.

Accuracy, true and false positives (TP and FP), and true and false negatives (TN and TF) for the logistic regression and DL were calculated and compared with the contingency table comparison t-test (significance $p<0.05$).

RESULTS

Patients in the CLARIFY-MS and MAGNIFY-MS studies were matched for age and sex.

For the SDMT, 5 configurations were found to have an accuracy greater than 70% and were fused together during the ensemble training. For the EDSS, the best configurations fused together in the ensemble were 12.

Adding the hand-crafted features to the DL features did not improve the final prediction.

Method	Accuracy	True positive	True negative	False positive	False negative
Standard statistics	56%	19	110	77	22
Deep learning	76%	8	166	21	33

Table 3: Results for both the logistic regression and the deep learning methods for the SDMT cognitive scale. Deep learning performance was significantly better than standard logistic regression ($p < 0.0001$).

Method	Accuracy	True positive	True negative	False positive	False negative
Standard statistics	70%	12	150	46	22
Deep learning	75%	9	163	33	25

Table 4: Results for both the logistic regression and the deep learning methods for the EDSS physical scale. Deep learning performance was comparable to the standard logistic regression ($p > 0.1$).

CONCLUSIONS

Deep learning demonstrated to be a **significantly better predictor** of **cognitive improvement** in **cladribine tablets-treated patients**, with respect to standard statistics.

It also showed a **higher accuracy** in **clinical improvement prediction**, although it did not achieve statistical significance.

Altogether these results suggest that **AI-based methods** can provide **useful** insights **for MS patient management**.

DISCLOSURES

This study was sponsored by Merck (CrossRef Funder ID: 10.13039/100009945).

M. Battaglini: is co-founder of Siena Imaging s.r.l.
Andrzej Smyk: employee of Merck Healthcare KGaA, Darmstadt, Germany
Ali-Frederic Ben-Amor: former employee of Ares Trading SA, Eysins, Switzerland, an affiliate of Merck KGaA.
N. De Stefano has received honoraria from Biogen-Idec, Bristol Myers Squibb, Celgene, Genzyme, Immunic, Merck, Novartis, Roche and Teva for consulting services, speaking, and travel support. He serves on advisory boards for Merck, Novartis, Biogen-Idec, Roche, and Genzyme, Immunic. He has received research grant support from the Italian MS Society and is co-founder of Siena Imaging s.r.l.
F. Sforazzini, L. Luchetti, F. Cacciante, G. Gentile,: have nothing to disclose.