Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study

Marcus D'Souza^{1*}, MD; Caspar E P Van Munster^{2*}, MD; Jonas F Dorn³, PhD; Alexis Dorier³, MSc; Christian P Kamm^{4,5}, MD; Saskia Steinheimer⁵, MD; Frank Dahlke³, MD, PhD; Bernard M J Uitdehaag², MD, PhD; Ludwig Kappos¹, MD; Matthew Johnson⁶, PhD

¹Neurologic Clinic and Policlinic, Departments of Medicine, Biomedicine and Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland

³Novartis Pharma AG, Basel, Switzerland

⁵Department of Neurology, Inselspital, University of Bern, Bern, Switzerland

⁶Microsoft Research, Cambridge, United Kingdom

*these authors contributed equally

Corresponding Author:

Marcus D'Souza, MD Neurologic Clinic and Policlinic Departments of Medicine, Biomedicine and Clinical Research University Hospital Basel and University of Basel Petersgraben 4 Basel, 4031 Switzerland Phone: 41 61 265 4151 Email: marcus.dsouza@usb.ch

Abstract

Background: In chronic neurological diseases, especially in multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients. Traditional scales are not sensitive enough to detect slight changes. Video recordings of patient performance are more accurate and increase the reliability of severity ratings. When these recordings are automated, quantitative disability assessments by machine learning algorithms can be created. Creation of these algorithms involves non–health care professionals, which is a challenge for maintaining data privacy. However, autoencoders can address this issue.

Objective: The aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

Methods: In this study, 20 pre-rated videos of patients performing the finger-to-nose test were recorded. An autoencoder created encoded frame vectors from the original videos and decoded the videos again. The original and decoded videos were shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether the 200 videos were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-Expanded Disability Status Scale definitions of limb ataxia. Furthermore, the neurologists tested whether ratings were equivalent between the original and decoded videos.

Results: In total, 172 of 200 (86.0%) videos were of sufficient quality to be ratable. The intrarater agreement between the original and decoded videos was 0.317 (Cohen weighted kappa). The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The interrater agreement between the original videos was 0.459 and that between the decoded videos was 0.302. The agreement was higher when no deficits or very severe deficits were present.

RenderX

²Department of Neurology, Multiple Sclerosis Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, Netherlands

⁴Neurocenter, Luzerner Kantonsspital, Luzern, Switzerland

Conclusions: The vast majority of videos (172/200, 86.0%) decoded by the autoencoder contained clinically relevant information and had fair intrarater agreement with the original videos. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non–health-care professionals are involved.

(J Med Internet Res 2020;22(5):e16669) doi: 10.2196/16669

KEYWORDS

autoencoder; video-rating; machine learning algorithms; deep neuronal network; Neurostatus-EDSS

Introduction

In chronic neurological diseases, especially multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients [1]. Traditional scales used to assess MS, such as the Expanded Disability Status Scale (EDSS), are not sensitive enough to detect slight changes in motor performance [2]. Video recordings of patient performance are more accurate and increase the reliability of severity ratings [3,4]. Moreover, when these recordings are automated, quantitative disability assessments by machine learning algorithms (MLA) can be created [5]. Machine learning algorithms are potentially more sensitive in detecting small changes between images; however, they require high-resolution images because of the high dimensionality of the data [6,7]. Creation of these algorithms usually involves non-health care professionals, which is a potential challenge for maintaining data privacy. Autoencoders can address this issue. They embed visual information into a lower-dimensional latent space that preserves information needed for algorithm development but is not visually interpretable by humans. [6]. An autoencoder consists of an encoder that creates encoded videos by creating a sequence of coded frame vectors and a paired decoder that transforms the coded frame vectors back into the original video. Videos encoded in this way can be shared with non-health care professionals, while the decoder can be used to verify if the essential information from the video has been captured. However, it is unknown whether the condensed data in the coded frame vectors contain clinically relevant data. Therefore, the aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

Methods

Study Design and Participants

This study was a subproject of the ASSESS MS study [5] and was approved by the local ethics committees. All participants gave their written informed consent prior to inclusion. In the ASSESS MS study, 9 standardized movements were recorded on video; these movements covered overall motor function, including upper extremity function, truncal stability, and mobility. A detailed description of the movements can be found elsewhere [8]. For this study, we used recordings of the finger-to-nose test. The execution of the finger-to-nose test was standardized using a detailed protocol: Each participant was instructed to close their eyes and abduct their arms to 90° at the

shoulder in full extension before touching their nose with the tip of their index finger. Both sides were tested. Original and decoded videos of 20 participants were shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether these 200 videos in total were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-EDSS definitions of limb ataxia [9] (subscore grade 0=no ataxia; grade 1=signs only; grade 2=tremor or clumsy movements easily seen, minor interference with function; grade 3=tremor or clumsy movements that interfere with function in all spheres; and grade 4=most functions are very difficult). The decoded videos were shown firstly, and after an interval of 2-3 weeks, the original videos were shown in the same order to minimize recall bias. The neurologists tested whether these videos were human-readable after decoding.

Autoencoder

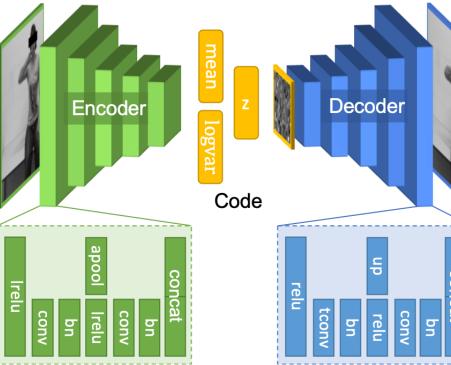
A variational autoencoder was trained on 2230 videos comprising the 9 standardized motor performances included in the ASSESS MS study. The autoencoder was structured so that the frames of each video were encoded into a lower-dimensional space and then decoded into their original form.

Figure 1 depicts the structure of the autoencoder [10]. An encoder network was presented with a single frame from the video without further context. The frame passed through 5 encoding blocks. In each block, the input was processed in a block inspired by a densely connected convolutional network [11], wherein a skip connection was provided between the input and output layers in addition to a convolutional layer/batch normalization sequence. Each block halved the resolution of the image and doubled the feature depth. This network predicted the mean and variance of a normal distribution, which was then sampled to produce a code. The code was presented to a second network that consisted of 5 decoding blocks. Each decoding block consisted of a skip connection (which performed a simple upsampling process) and a transposed convolutional block like that used in a deep convolutional generative adversarial network [12]. Each block doubled the resolution and halved the feature depth. The network was trained using a multi-scale structural similarity-based perceptual loss function [13] with Kullback-Leibler regularization as per Kingma and Welling [10]. The input images were 256×256 RGB-D images with a code length of 256. The training hyperparameters were as follows: the learning rate was 0.001, the convolutional kernel size was 5, and the number of initial filters was 8. The model was trained for 400 epochs.



D'Souza et al

Figure 1. Structure of the variational autoencoder.



The key property of interest to us was that when a frame is in its coded form, it is computationally prohibited to decipher it without access to the decoder [6]. An autoencoder as described above reduces the dimensionality of the input data (in our case, videos) by passing the data through an "information bottleneck" [14]. The resulting coded, or latent, space sufficiently describes the data in a way that allows an accurate partial reconstruction. The shared latent embedding is optimized to represent the salient information that is similar across frames of multiple videos (in our case: the movement), whereas dissimilar aspects (eg, background aspects, details of physical features) are less well conserved. Neural networks are a machine learning approach that is inspired by biological neuronal computation; these networks have demonstrated exceptional performance in complex image-related tasks in recent years [15-17]. Given this success, in this study, we used a neural net approach called a variational autoencoder [18]. A variational autoencoder has at its center a coded vector of vastly reduced dimensionality. This is because the decoder requires millions of floating point values to be set precisely before the coded vector can be successfully decoded into an image. At the same time, the coded vector contains all the information necessary to reconstruct that frame; interestingly, due to the variational constraints during training, the frame has semantically meaningful cosine distances to other visually similar frames. This property is very useful for machine learning tasks that operate upon these coded vectors because the coded frames can be used in place of the original video frames without the possibility that a human could use it to recognize the depicted participant.

Statistics

Intrarater agreement between the ratings of the original and the decoded videos was assessed using the Cohen weighted kappa with linear weights (ie, disagreements of 1, 2, and 3 were weighted by factors of 1, 2, and 3, respectively). A Cohen kappa of 0 corresponds to chance agreement; 0-0.2, to slight agreement; 0.21-0.4, fair agreement; 0.41-0.6, to moderate agreement; 0.61-0.8, to substantial agreement; and 0.81-1, to almost perfect agreement [19]. All analyses were performed in MATLAB (MathWorks, Inc).

Results

The characteristics of the study population and the participating neurologists are summarized in Table 1.

In total, 172/200 (86.0%) videos were of sufficient quality to be ratable. The Cohen weighted kappa indicating intra-rater agreement between the original and decoded videos was 0.317. The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The inter-rater agreements of the original and decoded videos were 0.459 and 0.302, respectively. As depicted in Figure 2, agreement was higher when no deficits (grade 0) or very severe deficits (grade 4) were present. Note that most videos that were not ratable were judged so by neurologists 2 and 5.

RenderX

D'Souza et al

 Table 1. Characteristics of the patients and neurologists who participated in the study.

Characteristic	Value
Patient characteristics (n=20)	·
Age (years), mean (95% CI)	44.4 (27-74)
Gender (female/male), n (%)	12 (63%)/7 (37%)
Disease duration (years), mean (95% CI)	13.2 (1-40)
Median EDSS ^a (range)	3.5 (0-6.5)
Type of MS ^b (RRMS ^c /SPMS ^d), n (%)	19 (95%)/1 (5%)
Neurologists (n=10)	
Gender (female/male), n (%)	5 (50%)/5 (50%)
Years of experience in neurology, mean (range)	8.8 (3 to >30)

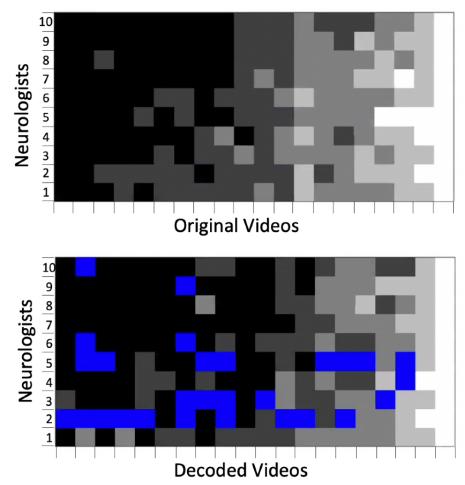
^aEDSS: Expanded Disability Status Scale.

^bMS: multiple sclerosis.

^cRRMS: relapsing remitting multiple sclerosis.

^dSPMS: secondary progressive multiple sclerosis.

Figure 2. Ratings by 10 neurologists of the original and decoded videos. The colored squares represent the different grades for limb ataxia of the finger-to-nose-test according to the Neurostatus-Expanded Disability Status Scale subscores: black=0, dark grey=1, grey=2, bright grey=3, and white=4. The blue squares represent videos that were judged as not ratable by the neurologists.





D'Souza et al

Discussion

Principal Findings

In this proof-of-concept study, 172/200 (86.0%) of the decoded videos were of sufficient quality to be ratable. We found fair intrarater agreement between the original and decoded videos. The agreement was better for minor and severe deficits in motor function.

Data security and privacy are increasingly requested by health care professionals for data capture, analysis, and storage [20]. At the same time, the use of machine learning algorithms and deep neuronal network techniques as subdomains of artificial intelligence is increasingly infiltrating all areas of health care [21,22]. The use of new technologies and electronic tools for capture and automated analysis of clinical data generally requires the involvement of non-health care professionals, which creates challenges regarding data privacy. To our knowledge, this is the first study to use an autoencoder to allow the analysis of patient videos while preserving data privacy.

Patients with MS may present with slight changes in motor performances over their disease course. Clinical assessment of these changes is notoriously difficult. Video analysis of motor performances allows automated analyses and quantification of disability by using machine learning algorithm–based analysis

Acknowledgments

This study was supported by Novartis.

Conflicts of Interest

decoded found fair ed videos.
ts in motor the time this is due to accordance of the video quality or the test-retest reliability of the finger-to-nose test. To our knowledge, no data are available regarding this psychometric property of the finger-to-nose test.
Limitations
A limitation of this proof-of-concept study is the class imbalance of the patient videos according to the four grades of limb ataxia

systems such as those used in the ASSESS MS study; however,

it requires a huge data set [5]. Since the creation of machine learning algorithms usually involves non-medical collaborators,

encoding of these videos is essential. The intra-rater agreement

Conclusions

In this proof-of-concept study, we have shown that the vast majority (172/200, 86.0%) of videos decoded by an autoencoder contained clinically relevant information regarding upper extremity motor performance represented by the finger-to-nose test and had fair intrarater agreement. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non-health care professionals are involved.

for the finger-to-nose test [9,21]. Further iterations of the deep

neural network are necessary to increase the intrarater reliability.

MD received travel support from Bayer AG, Biogen, Teva Pharmaceuticals, and Sanofi Genzyme and research support from University Hospital Basel. CM received travel support from Novartis Pharma AG, Sanofi Genzyme, Teva Pharmaceuticals, and Merck Serono; honoraria for lecturing and consulting from Novartis Pharma AG, Biogen-Idec, and Merck Serono; and compensation for serving on a scientific advisory board from Biogen-Idec, Roche, Merck Serono, and Sanofi Genzyme. JD is an employee of Novartis Pharma AG. AD is an employee of Novartis Pharma AG. CK has received honoraria for lectures and research support from Biogen-Idec, Novartis Pharma AG, Almirall, Bayer Schweiz AG, Teva Pharmaceuticals, Eli Lilly, Merck Serono, Sanofi Genzyme, and the Swiss Multiple Sclerosis Society. SS has received travel support from Bayer, Merck, and Novartis and has received honoraria for consulting from Bayer, Merck, Roche, and Teva. FD is an employee of Novartis Pharma AG. BU has received consultation fees from Biogen-Idec, Novartis Pharma AG, EMD Serono, Teva Pharmaceuticals, Sanofi Genzyme, and Roche. The Multiple Sclerosis Center Amsterdam has received financial support for research from Biogen-Idec, Merck Serono, Novartis Pharma AG, and Teva Pharmaceuticals. In the last 3 years, LK's institution (University Hospital Basel) received consultancy, steering committee, and advisory board fees from Actelion, Alkermes, Almirall, Bayer, Biogen, Celgene, df-mp, EXCEMED, GeNeuro SA, Genzyme, Merck, Minoryx, Mitsubishi Pharma, Novartis, Roche, Sanofi-Aventis, Santhera, Teva, and Vianex and as well as royalties for Neurostatus products. These fees were used exclusively for research support in the Department of Neurology. For educational activities of the Department, the institution received honoraria from Allergan, Almirall, Bayer, Biogen, EXCEMED, Genzyme, Merck, Novartis, Pfizer, Sanofi-Aventis, Teva, and UCB. MJ is an employee of Microsoft Research.

References

- Cohen JA, Reingold SC, Polman CH, Wolinsky JS. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. Lancet Neurol 2012 May;11(5):467-476. [doi: <u>10.1016/s1474-4422(12)70059-5</u>] [Medline: <u>22516081</u>]
- van Munster CEP, Uitdehaag BMJ. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs 2017 Feb 9;31(3):217-236. [doi: <u>10.1007/s40263-017-0412-5</u>] [Medline: <u>28185158</u>]
- Burggraaff J, Dorn J, D'Souza M, Morrison C, Kamm CP, Kontschieder P, et al. Video-Based Pairwise Comparison: Enabling the Development of Automated Rating of Motor Dysfunction in Multiple Sclerosis. Arch Phys Med Rehabil 2020 Feb;101(2):234-241. [doi: 10.1016/j.apmr.2019.07.016] [Medline: 31473205]

- D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. Mult Scler J Exp Transl Clin 2018 Aug 09;4(3):205521731879239. [doi: 10.1177/2055217318792399] [Medline: 30116550]
- Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors 2015 Jun 24;2(1):e11. [doi: 10.2196/humanfactors.4129] [Medline: 27025782]
- Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci Biobehav Rev 2017 Mar;74:58-75. [doi: 10.1016/j.neubiorev.2017.01.002] [Medline: 28087243]
- Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large scale multi sample study. Hum Brain Mapp 2018 Oct 11;40(3):944-954. [doi: <u>10.1002/hbm.24423</u>] [Medline: <u>30311316</u>]
- 8. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Mult Scler 2018 Aug 31;25(12):1673-1681. [doi: 10.1177/1352458518796690] [Medline: 30168739]
- 9. Kappos, L. https://www.neurostatus.net/. 2011. Neurostatus Scoring Definitions; Version 04/10.2 URL: <u>https://www.neurostatus.net/</u>
- Kingma DP, Welling M. An Introduction to Variational Autoencoders. Found Trends Mach Learn 2019;12(4):307-392. [doi: <u>10.1561/2200000056</u>]
- Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. Convolutional Networks with Dense Connectivity. IEEE Trans Pattern Anal Mach Intell 2019 May 23:1-1. [doi: <u>10.1109/tpami.2019.2918284</u>] [Medline: <u>31135351</u>]
- 12. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2016 Presented at: International Conference on Learning Representations; 2016 May 2-4; San Juan, Puerto Rico p. 1-16 URL: https://arxiv.org/abs/1511.06434
- Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. 2004 Presented at: Thirty-Seventh Asilomar Conference on Signals, Systems and Computers; 2003 Nov 9-12; Pacific Grove, CA, USA p. 1402-1402. [doi: 10.1109/acssc.2003.1292216]
- 14. Liou C, Huang J, Yang W. Modeling word perception using the Elman network. Neurocomputing 2008 Oct;71(16-18):3150-3157. [doi: 10.1016/j.neucom.2008.04.030]
- 15. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: IEEE. USA: IEEE; 2016 Dec 12 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA; 27-30 June 2016; as Vegas, NV, USA p. 770-778 URL: <u>http://ieeexplore.ieee.org/document/7780459/</u> [doi: <u>10.1109/cvpr.2016.90</u>]
- Cao Z, Simon T, Wei S, Sheikh Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In: IEEE. Honolulu, HI: IEEE; 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; 21-26 July 2017; Honolulu, HI, USA p. 7291-7299 URL: <u>http://ieeexplore.ieee.org/document/8099626/</u> [doi: <u>10.1109/cvpr.2017.143</u>]
- 17. Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. 2018 Presented at: International Conference on Learning Representations; 2018 April 30-May 3; Vancouver, BC, Canada.
- Kingma D, Welling M. Auto-Encoding Variational Bayes. 2014 Presented at: International Conference on Learning Representations; 2014 April 14-16; Banff, Canada p. 1-14 URL: <u>https://arxiv.org/abs/1312.6114</u>
- Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159-174. [doi: <u>10.2307/2529310</u>] [Medline: <u>843571</u>]
- 20. Beinke JH, Fitte C, Teuteberg F. Towards a Stakeholder-Oriented Blockchain-Based Architecture for Electronic Health Records: Design Science Research Study. J Med Internet Res 2019 Oct 7;21(10):e13585. [doi: <u>10.2196/13585</u>] [Medline: <u>31593548</u>]
- 21. Rowe M. An Introduction to Machine Learning for Clinicians. Acad Med 2019;94(10):1433-1436. [doi: 10.1097/acm.00000000002792] [Medline: 31094727]
- 22. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286. [doi: 10.2196/12286] [Medline: 30950797]

Abbreviations

EDSS: Expanded Disability Status Scale MS: multiple sclerosis RRMS: relapsing remitting multiple sclerosis SPMS: secondary progressive multiple sclerosis



Edited by G Eysenbach; submitted 13.10.19; peer-reviewed by S Allin, A Aminbeidokhti; comments to author 06.01.20; revised version received 19.02.20; accepted 19.03.20; published 08.05.20

<u>Please cite as:</u> D'Souza M, Van Munster CEP, Dorn JF, Dorier A, Kamm CP, Steinheimer S, Dahlke F, Uitdehaag BMJ, Kappos L, Johnson M Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study J Med Internet Res 2020;22(5):e16669 URL: <u>https://www.jmir.org/2020/5/e16669</u> doi: <u>10.2196/16669</u> PMID: <u>32191621</u>

©Marcus D'Souza, Caspar E P Van Munster, Jonas F Dorn, Alexis Dorier, Christian P Kamm, Saskia Steinheimer, Frank Dahlke, Bernard M J Uitdehaag, Ludwig Kappos, Matthew Johnson. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.05.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

