



# Response to the letter to the editor on the article: a non-invasive, automated diagnosis of Menière's disease using radiomics and machine learning on conventional magnetic resonance imaging—a multicentric, case-controlled feasibility study

Marly F. J. A. van der Lubbe<sup>1</sup> · Akshayaa Vaidyanathan<sup>2,3</sup> · Marjolein de Wit<sup>1</sup> · Elske L. van den Burg<sup>1</sup> · Alida A. Postma<sup>4,5</sup> · Tjasse D. Bruintjes<sup>6,7</sup> · Monique A. L. Bilderbeek-Beckers<sup>8</sup> · Patrick F. M. Dammeijer<sup>9</sup> · Stephanie Vanden Bossche<sup>10,11</sup> · Vincent Van Rompaey<sup>12</sup> · Philippe Lambin<sup>2</sup> · Marc van Hoof<sup>1</sup> · Raymond van de Berg<sup>1,5</sup>

Received: 24 March 2022 / Accepted: 14 April 2022 / Published online: 5 August 2022  
© Italian Society of Medical Radiology 2022

The authors would like to thank Naganawa and colleagues for their interest [1]. Their primary concern is that our algorithm detects the type of MR sequence instead of Meniere's disease (MD) itself. This was hypothesized to be the result of a difference in the distribution of fast spin-echo-based and gradient-echo MR sequences. Indeed, extracted radiomic image features depend on the sequence type and acquisition parameters, and therefore, machine learning techniques are susceptible to such forms of bias [2, 3], especially when population inequalities are present (Fig. 1).

Our pragmatical trial was retrospective, and our sampling was based on data availability and reflected clinical practice. As concluded in our article, prospective studies need

to be done to fully verify our findings and to ensure that no covert bias explains the results. Other confounding factors then imaging parameters exist and could also be relevant, such as disease duration, the clinical setup to diagnose MD, or the choice for the control group. Such factors should be taken into account in the next clinical validation phase. Nevertheless, we aimed to prevent bias in our study design as much as possible. Amongst others, by gaining a large enough sample size and by sampling four centers that had a similar clinical setup in terms of diagnostic procedures for MD and asymmetric hearing loss. All images underwent pre-processing before features were extracted [1–3] to minimize the influence of heterogeneities in the multiparametric dataset.

Marc van Hoof, Raymond van de Berg have equally contributed to do this work.

✉ Marly F. J. A. van der Lubbe  
marly.lubbe@mumc.nl

<sup>1</sup> Department of Otolaryngology and Head and Neck Surgery, Maastricht University Medical Center +, Maastricht, The Netherlands

<sup>2</sup> The D-Lab, Department of Precision Medicine, GROW Research Institute for Oncology, Maastricht University, Maastricht, The Netherlands

<sup>3</sup> Research and Development, Oncoradiomics SA, Liege, Belgium

<sup>4</sup> Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, Maastricht, The Netherlands

<sup>5</sup> School for Mental Health and Sciences, Maastricht University, Maastricht, The Netherlands

<sup>6</sup> Department of Otorhinolaryngology, Gelre Hospital, Apeldoorn, The Netherlands

<sup>7</sup> Department of Otorhinolaryngology, Leiden University Medical Center, Leiden, The Netherlands

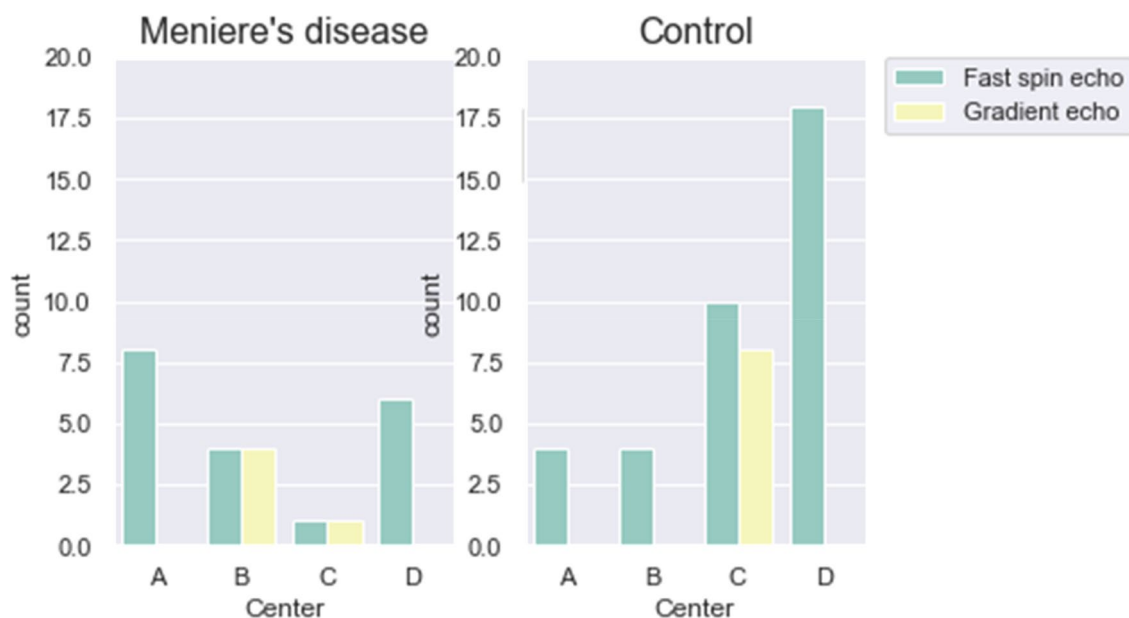
<sup>8</sup> Department of Radiology, Viecuri Medical Center, Venlo, The Netherlands

<sup>9</sup> Department of Otorhinolaryngology, Viecuri Medical Center, Venlo, The Netherlands

<sup>10</sup> Department of Radiology, Antwerp University Hospital, Antwerp, Belgium

<sup>11</sup> Department of Radiology, AZ St-Jan Brugge-Oostende, Bruges, Belgium

<sup>12</sup> Department of Otorhinolaryngology and Head & Neck Surgery, Faculty of Medicine and Health Sciences, Antwerp University Hospital, University of Antwerp, Antwerp, Belgium



**Fig. 1** The distribution of the MR sequences per center in the test cohort for the Meniere's disease and control group

A new post hoc analysis was performed to answer Naganawa's et al. questions regarding the distribution and accuracy of MR sequences. In total, 55 (21.2%) gradient-echo sequences and 205 (78.8%) fast spin-echo sequences were included in our study. Gradient-echo sequences were only included in centers B and C. These consisted of 19.3% ( $n = 11$ ) and 40.7% ( $n = 44$ ) of the total for those centers. The proportion of gradient-echo sequences in the MD and control group were 22.5% ( $n = 27$ ) and 20% ( $n = 28$ ). The proportion of gradient-echo sequences in the train and test group were 21.9% ( $n = 42$ ) and 19.1% ( $n = 13$ ). The proportion of gradient-echo sequences did not differ between patients and controls,  $X^2(1, N = 260) = 0.115, p = 0.743$ , nor between the training and test group,  $X^2(1, N = 260) = 0.093, p = 0.760$ . Within the training cohort, 74 (49%) fast spin-echo's existed in the MD group and 76 (51%) in the control group. The distribution in the test cohort was somewhat unequal, with 19 (35%) fast spin-echo's in the MD group and 36 (65%) in the control group. This difference did not reach statistical significance  $X^2(1, N = 66) = 0.003, p = 0.955.6$

Most importantly, we investigated if the accuracy of the diagnoses is above chance level, for the two different types of MR sequence. In the training set ( $n = 192$ ), the accuracy was 76% for fast spin-echo (prevalence MD 49%) and 60% for gradient-echo sequences (prevalence MD 52%). In the test set ( $n = 68$ ), the accuracy was 84% for fast spin-echo (prevalence MD 35%) and 77% for gradient-echo sequences (prevalence MD 38%). An exact binomial test was employed to determine if the accuracy was statistically significantly higher than the prior probability (prevalence). In the training set, this was the case for the fast spin-echo ( $p$

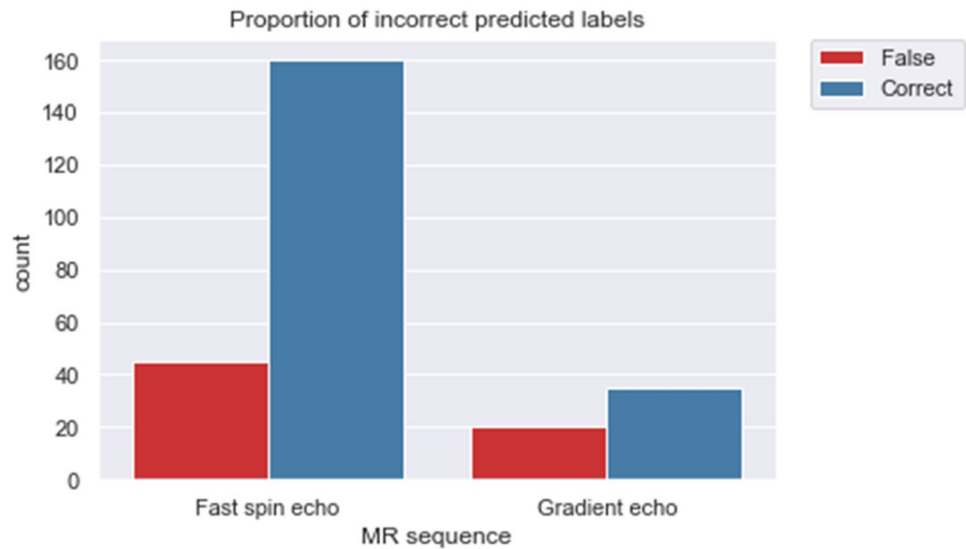
value =  $< 0.0001$ ), but not for the gradient-echo sequence ( $p$  value = 0.206). In the test set, this again was the case for the fast spin-echo ( $p$  value = 0.002), but not for the gradient-echo sequence ( $p$  value = 0.208). This marked finding could indicate that perhaps gradient-echo MRI is less suitable for inner ear radiomic evaluation or requires more training and/or more samples.

In conclusion, sampling based on data availability did not seem to result in an unbalanced distribution for patient, control, train, and test cohort. The accuracy of the radiomics algorithm with only fast spin-echo MR is similar (84%), as was presented in the original manuscript (82%), and is well above chance level ( $p = 0.002$ ). The MR sequence did matter, as the algorithm seemed to perform worse on gradient-echo MRI (Fig. 2) and was not significantly above chance level ( $p = 0.21$ ).

Although our study setup is not suitable to fully exclude the possibility, it is unlikely that the proposed classification model distinguishes between imaging types instead of MD vs. control. The results of the cross-validation analysis of our study [1] with various train-test iterations also support this hypothesis. Our study did not assess the effect of pre-processing; however, it might have prevented a large effect of distributional shift introduced by multiparametric images [2, 4].

Prospective and controlled studies with predefined image acquisition protocols are needed to further validate and develop the classification model, allowing for more detailed factor analyses. Another important goal for future study would be, as noted by Naganawa et al., to compare radiomics results (on conventional MRI) in patients who also

**Fig. 2** The number of the model's incorrect predicted labels for fast spin-echo and gradient-echo MRIs in the complete dataset (training and test)



received delayed contrast-enhanced MR (hydrops) imaging, considered the gold standard in our days.

**Author contributions** The first draft was written by the first author M.F.J.A. van der Lubbe and revised by M. van Hoof. All authors read and approved the final draft.

**Funding** The authors declare no funding or financial support was received for this manuscript.

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

**Human and animals participants** This response to the letter to the editor follows the ethical approval stated in the original manuscript which was performed in accordance with the guidelines outlined by Dutch and Belgium legislation. Subjects were enrolled and fully anonymized by the local investigators (Maastricht University Medical Center, University Hospital Antwerp, VieCuri Hospital Venlo, and Apeldoorn Dizziness Center) and were therefore not asked for their consent. According to the Medical Research Involving Human Subjects Act (WMO), ethical approval was not required due to the retrospective nature and anonymization of the data.

**Informed consent** This was not applicable to our response to the letter to the editor.

## References

1. van der Lubbe MF, Vaidyanathan A, de Wit M, van den Burg EL, Postma AA, Bruintjes TD, Bilderbeek-Beckers MA, Dammeyer PF, Bossche SV, Van Rompaey V, Lambin P (2021) A non-invasive, automated diagnosis of Menière's disease using Radiomics and Machine Learning on conventional MR imaging. A retrospective, multicentric diagnostic case-control study with training and independent validation cohort
2. Carré A, Klausner G, Edjlali M et al (2020) Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep.* <https://doi.org/10.1038/s41598-020-69298-z>
3. Kondrateva E, Pominova M, Popova E, et al (2021) Domain shift in computer vision models for MRI data analysis: an overview
4. Um H, Tixier F, Bermudez D et al (2019) Impact of image pre-processing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phys Med Biol.* <https://doi.org/10.1088/1361-6560/ab2f44>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.