frontiers

Identification of Mild Cognitive Impairment from Speech in Swedish using Deep Sequential Neural Networks

Charalambos Themistocleous [1,2,*], Marie Eckerström [3] and Dimitrios Kokkinakis [1]

[1]The Swedish Language Bank, Department of Swedish, University of Gothenburg, Sweden
[2]Department of Neurology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[3]Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, University of Gothenburg, Sweden

Correspondence*:
Charalambos Themistocleous
charalambos.themistocleous@gu.se

ABSTRACT

While people with mild cognitive impairment (MCI) portray noticeably incipient memory difficulty in remembering events and situations along with problems in decision making, planning, and finding their way in familiar environments, detailed neuropsychological assessments also indicate deficits in language performance. To this day, there is no cure for dementia but early-stage treatment can delay the progression of MCI; thus, the development of valid tools for identifying early cognitive changes is of great importance. In this study, we provide an automated machine learning method, using Deep Neural Network Architectures, that aims to identify MCI. Speech materials were obtained using a reading task during evaluation sessions, as part of the Gothenburg MCI research study. Measures of vowel duration, vowel formants ($F1$ to $F5$), and fundamental frequency were calculated from speech signals. To learn the acoustic characteristics associated with MCI vs. healthy controls, we have trained and evaluated ten Deep Neural Network Architectures and measured how accurately they can diagnose participants that are unknown to the model. We evaluated the models using two evaluation tasks: a 5-fold crossvalidation and by splitting the data into 90% training and 10% evaluation set. The findings suggest first, that the acoustic features provide significant information for the identification of MCI; second, the best Deep Neural Network Architectures can classify MCI and healthy controls with high classification accuracy ($M = 83\%$); and third, the model has the potential to offer higher accuracy than 84% if trained with more data (cf., $SD \approx 15\%$). The Deep Neural Network Architecture proposed here constitutes a method that contributes to the early diagnosis of cognitive decline, quantify the progression of the condition, and enable suitable therapeutics.

Keywords: speech production, vowels, prosody, neural network, machine learning, dementia, MCI

1 INTRODUCTION

Individuals with mild cognitive impairment (MCI) portray a noticeable memory difficulty in remembering events and situations along with problems in decision making, planning, interpreting instructions, and orientation (40, 58, 14, 41, 56). These cognitive problems become frequent and more severe compared to the cognitive decline in normal aging (see also 11, 32). As the MCI progresses, MCI individuals face a higher risk of developing Alzheimer's Disease (AD).

In search of less strenuous and non-invasive techniques for assessing MCI, currently, there has been substantial interest on the role of speech and language and its potentials as markers of MCI. Language impairment in AD is well established (e.g., 6, 42, 57) and can be evaluated by using assessments, such as naming tests (8), discourse (9, 13, 33), verbal fluency tests (e.g., 23), complexity measures, such as phonemes per word, phone entropy, verbal fluency, and word recall (42, 57, 13, 20, 55, 26, 45, 6, 5, 3, 2). Findings with respect to syntax and phonology have been inconsistent though (for a discussion on the role of syntax in MCI, see 25). Also, many studies explored the interactions of language and other predictors from imaging, biomarkers etc., in dementia (4, 15, 44, 30, 27, 28, 16). The fact that language impairment occurs early and commonly in the progression of AD, motivated many researchers to identify markers of language impairment in MCI. For example, Manouilidou et al. (35) showed that while MCI individuals preserve morphological rule knowledge, they face processing difficulties of pseudo-words (for a discussion and review of current studies, see 47, 36)). As there is only a handful of studies on the acoustic properties of MCI speech (e.g., 43, 17), more research on speech acoustics is required to gain a better understanding of how MCI speech differs from that of healthy controls.

The development of automated machine learning models that can learn the characteristics of MCI and provide an early and accurate identification of MCI is of utmost importance for two main reasons: First, an early identification can enable multidomain life style interventions and/or pharmacological treatments at the MCI stage, or even earlier, which can potentially delay or might even prevent the development of AD and other types of dementia (29, 56). Second, the early identification, will provide time to patients and their families to make decisions about their care, family issues, and legal concerns (56).

The aim of this study is to provide an automated method that can identify MCI individuals and distinguish them from healthy controls using acoustic information. Specifically, in this study, we provide an automated machine learning method using Deep Neural Network Architectures that identifies individuals with MCI from healthy controls. We demonstrate its performance by using data from Swedish. Specifically, 55 Swedish participants, 30 healthy controls and 25 MCI, were instructed by a clinician to read a short passage, consisting of 144 words, as part of their evaluation. Reading tasks are being employed extensively in research because they provide rich linguistic data without straining the participants (19). Also, they have the advantage that they are restrictive with respect to the segmental environment of vowels and consonants, which is the same for all participants. Next, the speech material was transcribed and segmented into vowels and consonants. From the segmented material, we measured vowel $F1 - F5$ formant frequencies, $F0$, and duration. Vowel formants are a range of vowel frequency peaks in the sound spectrum. Formant frequencies are the primary acoustic correlates for the production of vowels. $F1$ and $F2$ usually suffice for the identification of vowels in most languages but higher order formant frequencies can provide information about the social—such as the age, gender, and dialect—and physiological properties of speakers (50, 52, 51). In Swedish, $F3$ also contributes to the distinction of rounded and unrounded

	N		Age	
	F	M	F	M
HC	19	11	68(7.6)	69 (5.7)
MCI	13	12	72 (5.1)	70 (5.6)

Table 1. Age and gender of healthy controls (HC) and participants with Mild Cognitive Impairment (MCI).

vowels (18). $F0$ is the acoustic correlate of intonation. Speakers vary the $F0$ of their utterances to produce various melodic patterns, such as when emphasizing parts of the utterance, asking questions, giving commands, etc. $F0$ (e.g., mean $F0$, $F0$ minimum and maximum) is found to be lower in individuals with depression (37, 12). In addition to frequency measurements, we measured vowel duration.

For the classification task, we have evaluated several Deep Neural Network Architectures based on Multilayer Perceptrons (MLP). MLPs are a type of sequential, Feed-Forward Neural Network, which when trained on a dataset, can learn a non-linear function approximator for the classification of MCI and healthy participant:

$$f(\cdot) : R^m \rightarrow R^o \tag{1}$$

where m is the number of dimensions for input and o is the number of dimensions for output. Given a set of vowel features $X = x_1, x_2, ..., x_m$ and a target y; namely, an array of values determining the condition of the participant (healthy controls vs. MCI), the neural network can learn the classification function. The advantage of this type of network for our data is that it can learn non-linear structures.

2 METHODOLOGY

In this section, we describe the development of the dataset and the structure of the predictors.

2.1 Speech materials

Participants for this study were recruited from the Gothenburg MCI study, which is a large clinically based longitudinal study on mild cognitive impairment (56). This study aims to increase the nosological knowledge that will enable rational trials in AD and other types of dementia. It also includes longitudinal in-depth phenotyping of patients with different forms and degrees of cognitive impairment using neuropsychological, neuroimaging, and neurochemical tools (56). Speech recordings were conducted as part of the additional assessment tests that conduced for the purposes of the Riksbankens Jubileumsfond – The Swedish Foundation for Humanities & Social Sciences "Linguistic and extra-linguistic parameters for early detection of cognitive impairment" research grant (NHS 14-1761:1).

2.2 Participants

The recordings were conducted in an isolated environment at the University of Gothenburg. 30 healthy controls and 25 MCI—between 55 and 79 years old ($M = 69, SD = 6.4$) participated in the study. The two groups did not differ with respect to age ($t(52.72) = -1.8178, p = n.s.$) and gender ($W = 1567.5, p = n.s.$), as is evident by the non-significant results from a t *test* and

an independent 2-group Mann-Whitney U test respectively. Participants were selected based on specific inclusion and exclusion criteria: i. participants should not have suffered from dyslexia and other reading difficulties; ii. they should not have suffered from major depression, ongoing substance abuse, poor vision that cannot be corrected with glasses or contact lenses; iii. they should not have been diagnosed with other serious psychiatric, neurological or brain-related conditions, such as Parkinson's disease; iv. they had to be native Swedish speakers; v. they had to be able to read and understand information about the study; and vi. they had to be able to give written consent.

Healthy controls had a significantly higher Mini-Mental State Exam (MMSE) score. (The MMSE score is a scale of 0 to 30 and represents the cognitive status of an individual). Mean MMSE score for the MCI participants was 28.2, which is close to normal (21). Ethic approvals for the study were obtained by the local ethical committee review board (reference number: L091-99, 1999; T479-11, 2011); while the currently described study was approved by the local ethical committee decision 206-16, 2016.

2.3 Acoustic Features

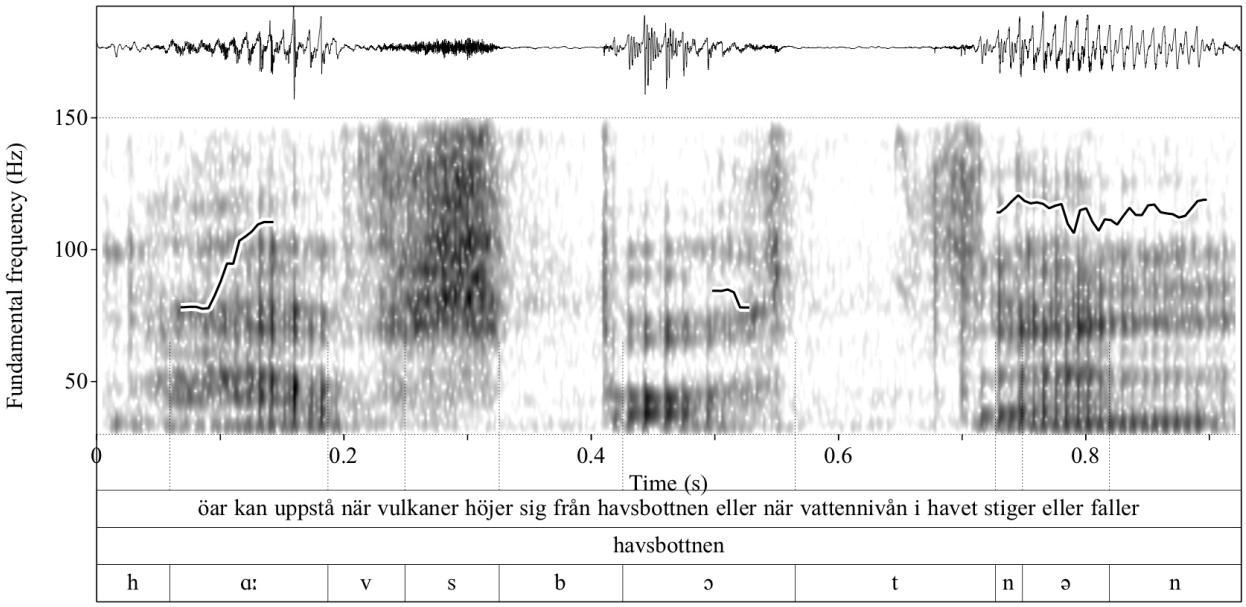

Figure 1. Waveform, spectrogram, and $F0$ contour—superimposed on the spectrogram—of an example utterance (upper tier). Shown in the plot is the segmentation of the word havsbottnen 'seabed' (middle tier); the individual sounds are shown in the lowest tier. Sound boundaries are indicated with thin vertical lines. The ordinate shows the $F0$ values whereas the abscissa the time in sec.

A. Segmentation: Each vowel was segmented in the acoustic signal; that is, we located the right and left boundary of vowels and consonants. A segmentation example is shown in Figure 1. Specifically, the figure shows the waveform (upper panel) and spectogram of the word havsbottnen 'seabed' taken here as an example from a larger sentence: öar kan uppstå när vulkaner höjer sig från havsbottnen eller när vattennivån i havet stiger eller faller "islands can occur when volcanoes

rise from the seabed or when water levels in the ocean rise or fall" (see the Appendix for the whole passage). There are also three different tiers with the transcriptions, the top tier defines the boundaries of sentences; the second tier in the middle shows the word boundaries; and the lower tier shows the segmental boundaries, namely the boundaries of consonants and vowels (see also the thin lines extending from the lower tier to the middle of the spectogram and demarcate vowels and consonants). For the segmentation, we have employed an automatic module for Swedish developed by the first author (54). As measurements and processes rely on accurate segmentation this step is crucial; therefore, all segmentation decisions were evaluated twice based on the following segmentation criteria: vowel onsets and offsets were demarcated by the beginning and end of the first two formant frequencies; the rise of the intensity contour at the beginning of the vowel and its fall at the end of the vowel served as additional criteria for vowel segmentation. Then, we measured the acoustic properties of using Praat (7). Overall, there were 4396 HC and 4273 MCI productions, which is a relatively balanced data set.

B. Acoustic Measurements: Vowel formants were measured at multiple positions. Traditionally vowel formants are measured using a single measurement at the middle of the vowel, which is supposedly the vowel target. Nevertheless, the shape of the formant contour can also convey information about participants' sociophonetic properties (see for a discussion 50). To this end, we conducted three measurements of formants at the 15%, 50%, 75% of vowels' duration. Vowel formants were calculated using standard Linear Predictive Coding (LPC-analysis) (38). We also measured vowel duration and fundamental frequency ($F0$) (49). The latter is the lowest frequency of speech; and it constitutes the main acoustic correlate of speech melody (a.k.a., intonation) (31). We calculated the minimum, maximum, and mean $F0$ for each vowel. $F0$ and formant frequencies were measured in Hertz.

C. Sociophonetic Features: In addition to the acoustic features, the model included as predictors information about participants' age and gender. Overall, the classification tasks included the following 24 acoustic and sociophonetic predictors:

1. Vowel Formants: We measured the first five formant frequencies of vowels (i.e., $F1$, $F2$, $F3$, $F4$, $F5$) at the 15%, 50%, and 75% of the vowels' total duration: i.e., $F1$ 15%, $F1$ 50%, $F1$ 75% ... $F5$ 15%, $F5$ 50%, and $F5$ 75%; We also provided the log-transformed values of F1, F2, F3.
2. Fundamental frequency ($F0$): We measured the $F0$ across the duration of the vowel and calculated the *mean* $F0$, *min* $F0$, and *max* $F0$.
3. Vowel duration: Vowel duration measured in seconds from vowel onset to vowel offset.
4. Gender: Participants' gender.
5. Age: Participants' age.

2.4 Models and experiments

In this section, we describe the neural network architectures employed in this work. Ten neural network architectures that differed in the total number of hidden layers from $h1 \dots h10$ were evaluated twice using validation split and cross-validation (the other parameters were the same across models). We present all ten models and not the best model only because i. we want to demonstrate the whole methodological process that led to the selection of the best model and stress out that the final model is the result of a dynamic process of model comparison; ii. different

Layer	Shape	Activation
Input Layer	Dense 300 (21 Input Dimensions)	ReLU
1 . . . 10 Hidden Layers	Dense 300	ReLU
Output Layer	1	Sigmoid

Table 2. Deep Neural Network Architectures from 1 to 10 Hidden Layers. All models employed stochastic gradient descent optimizer with 0.9 Nesterov momentum.

randomization of the data may provide different output; thus, a rigorous evaluation can demonstrate whether the output is consistent across models. For example, by demonstrating that the output is not random and that there is a pattern between the different models; and iii. the evaluation process is being part of the model and not external to the model as it can explain the final architecture of the model, such as the number of hidden layers in the model. An overview of the architectures is shown in Figure 2 and in Table 2. The neural architectures were implemented in Keras, a high-level neural networks API (10) running on top of TensorFlow (1) in Python 3.6.1. For the normalization and scaling of predictors, we employed modules from scikit-learn, which is a machine learning library implemented in Python (24, 39).

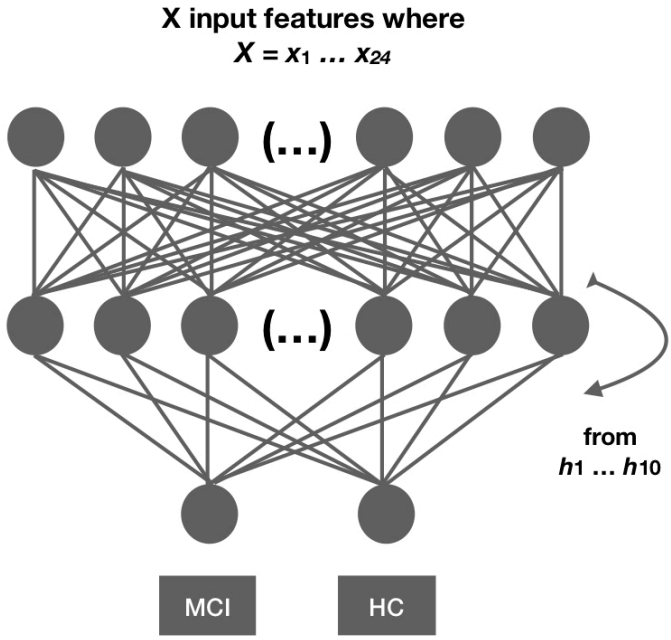

Figure 2. Network architecture. We developed 10 different networks with 21 predictors each. The networks differed in the number of hidden layers ranging from 1 . . . 10. Each network architecture was evaluated twice using cross-validation and evaluation split. Model comparison measures are reported for each evaluation separately.

2.4.1 Model design

1. Transformation. All predictors were centered and scaled, using standard scaling, which standardizes the features by removing the mean and by scaling to unit variance (for the scikit-learn implementation of a Standard Scaler, see 39). The mean and standard deviation are estimated on the training set. Then these estimated measures are used to transform the training and test sets separately. So, data in training and test sets are not transformed simultaneously.

	Condition positive	Condition negative
Predicted condition positive	True positive (TP)	False positive (FP)
Predicted condition negative	False negative (FN)	True negative (TN)

Table 3. Confusion matrix.

The reason for conducting different transformations is to avoid a bias from the test features when the mean and the standard deviation are estimated during standard scaling.

2. Layers. We tested ten different network architectures that differed in the number of hidden layers from $h1 \dots h10$; the input and output layers are excluded. The number of layers in the network can affect its accuracy. Most layers except from the output layer were trained with a ReLU activation function (34, 22). The last layer had a sigmoid activation.
3. Optimization. We employed a Nesterov stochastic gradient descent (SGD) optimization algorithm. The learning rate was set to 0.1 and the momentum was set to 0.9.
4. Epochs and Batch Size. (a) In cross-validation: network architectures were trained for 80 epochs with 35 as a batch size. (b) In 90%-10% validation split: networks were trained for 100 epochs with 35 as a batch size.

2.4.2 Model comparison and evaluation measures

During the training phase, the neural network learns the acoustic properties that characterize MCI and HC. During the evaluation phase, the network evaluates unknown data vectors from the test set; this time the corresponding label (i.e., MCI or HC) is not available to the model and makes a prediction whether these unknown data vectors correspond to MCI or HC productions. To estimate the performance of the neural network, we compare the predictions of the neural network with the classification made by clinicians using combined imaging and neurological, neuropsychological examination.

A confusion matrix represents the relationship between predicted values and actual values (see Table 3). The columns of Table 3 represent the actual condition (MCI or HC) and the rows represent the positive and negative predictions. A true positive (TP) indicates how many times the condition was MCI and the neural network actually predicted MCI; the false positive (FP) indicates when the condition was HC but the network predicted MCI; the false negative (FN) indicates when the condition was MCI and the network predicted HC; and lastly, the true negative indicates when the condition was HC and the neural network made the correct prediction, namely HC. The different neural network models were compared with each other based on the following evaluation measures: i. accuracy, ii. precision, iii. recall, iv. F1 score, and v. ROC/AUC.

1. Accuracy: The accuracy is the most commonly employed evaluation measure in classification studies. It refers to the number of correct predictions made by the model divided by the total number of all estimations: $Accuracy = (TP + TN)/(TP + TN + FP + FN)$. However, the accuracy is not always the best evaluation measure when the design is unbalanced and corrections are often required. To this end, the precision, recall, $F1$ *score*, and ROC/AUC curve provide more balanced estimates.
2. Precision The precision is the number of true positives divided by the sum of true positives and false positives, i.e., $Precision = TP/(TP + FP)$. So, when there are many FPs, the precision measure will be low.

3. Recall: Recall (a.k.a. sensitivity) is the number of true positives divided by the sum of true positives and false negatives, i.e., $Recall = TP/(TP + FN)$. This suggests that a low recall will indicate that there are many FNs.
4. $F1$ *score*: The $F1$ *score* is the weighted average of Precision and Recall: $F1$ *score* $= 2 \times ((Precision \times Recall)/(Precision + Recall))$. The $F1$ *score* captures the performance of the models better than the accuracy, especially when the design is unbalanced. A value of 1 indicates a perfect precision and recall, whereas a value of 0 designates the worst precision and recall. Because the $F1$ *score* can be less intuitive than the accuracy, most machine learning studies usually report the accuracy of the model.
5. ROC/AUC curve: The receiver operating characteristic (ROC) and the area under the curve (AUC) are two evaluation measures that display the performance of a model. The ROC is a curve that is created by plotting the true positive rate (i.e., the precision) against the false positive rate (i.e., 1-Recall). An optimal model has an ROC closer to 1 whereas a bad model has an ROC closer to 0.

2.4.3 Model evaluation

1. 5-fold group cross-validation. In a "5-fold group cross-validation", the data are randomized and split into five different folds and the network is trained five times. In each training setting, a different part of the available data is hold out as a test set. The "5 fold group crossvalidation" also ensures that there are no measurements from the same participants in the training and test sets as all data from a given participant will be either in the test set or in the training set but not in both sets. (In a simple "5-fold cross-validation" measurements from a given participant might be in both the training and test set after randomization which creates a bias, because the network will be trained on properties from given participants and then asked to provide predictions with respect to these participants.) To evaluate the cross-validation, we provide the mean and standard deviation of the accuracy we get from each evaluation. We also provide the ROC curve and the AUC scores that provide a corrected measure of the accuracy.
2. 90% – 10% Evaluation split. We also provide the findings from the validation split and discuss in detail validation measures, namely the accuracy of the model, the precision, recall, and $F1$ *score*. To this end, we split the data into two parts. The first part consists of the 90% of the data and functions as a training corpus whereas the second part, the remaining 10% functions as an evaluation set. Just like in the cross-validation, the speakers in the evaluation and test sets are different.

3 RESULTS

First, we present the results from the evaluation task and then, we present the results from the validation split.

3.1 5-fold group crossvalidation

We conducted a 5-fold group cross-validation. Within each fold the model is validated 80 times, which is the number of epochs of the model and the mean accuracy, mean validation accuracy, and the corresponding standard deviation are calculated. Table 4 provides the mean accuracy and the mean validation accuracy along with the corresponding standard deviation that results from the 5-fold crossvalidation. As seen by Table 4 models six to ten are consistent with respect to their

Table 4. Model $M1 \dots M10$ mean classification accuracy and mean validation accuracy and the corresponding SD from the 5-fold crossvalidation.

Model	Accuracy		Val. Accuracy	
	$Mean$	SD	$Mean$	SD
M1	98	3	75	12
M2	99	3	80	14
M3	99	2	81	15
M4	99	2	82	15
M5	99	2	82	14
M6	99	2	83	15
M7	99	2	83	16
M8	99	2	83	15
M9	99	2	83	16
M10	98	3	83	17

Table 5. 90%/10% validation split results. The table shows the accuracy, precision, recall, and $f1score$ for Model $1 \dots 10$.

Model	Accuracy	Precision	Recall	$F1$ $score$
M1	67	86	56	63
M2	68	92	56	66
M3	67	100	49	65
M4	68	63	62	62
M5	71	73	71	71
M6	68	73	72	72
M7	75	100	49	65
M8	65	100	49	65
M9	69	100	49	65
M10	66	95	51	64

classification accuracy. These models have six to ten hidden layers and all resulted in 83% mean cross-validated accuracy. Figure 3 displays the mean ROC curve and AUC of the 10 neural network models. The shaded area indicates the SD for the final model: M10. The results from the cross-validation clearly show that when trained using a Sequential Neural Network, speech features can be employed for the identification of MCI. To establish this finding, we provide a second evaluation by training the same networks on the 90% of the data and evaluating on the remaining 10%.

3.2 90%-10% evaluation split

Table 3 shows a comparison of the accuracy scores on the training set. The highest accuracy was provided by Model 7 that resulted in 75% classification accuracy and the second best model was Model 5 with 71% classification accuracy at the validation set.

4 DISCUSSION

The number of people that are developing dementia is increasing worldwide. Identifying MCI early is of utmost importance as it can enable a timely treatment that can delay its progression. A number of studies have shown that speech and language, which are ubiquitous in everyday communication, can provide early signs of MCI and other prodromal stages of Alzheimer's disease (e.g., 2). The aim of this study has been to provide a classification model for the quick and fast identification of MCI individuals, using data from speech productions.

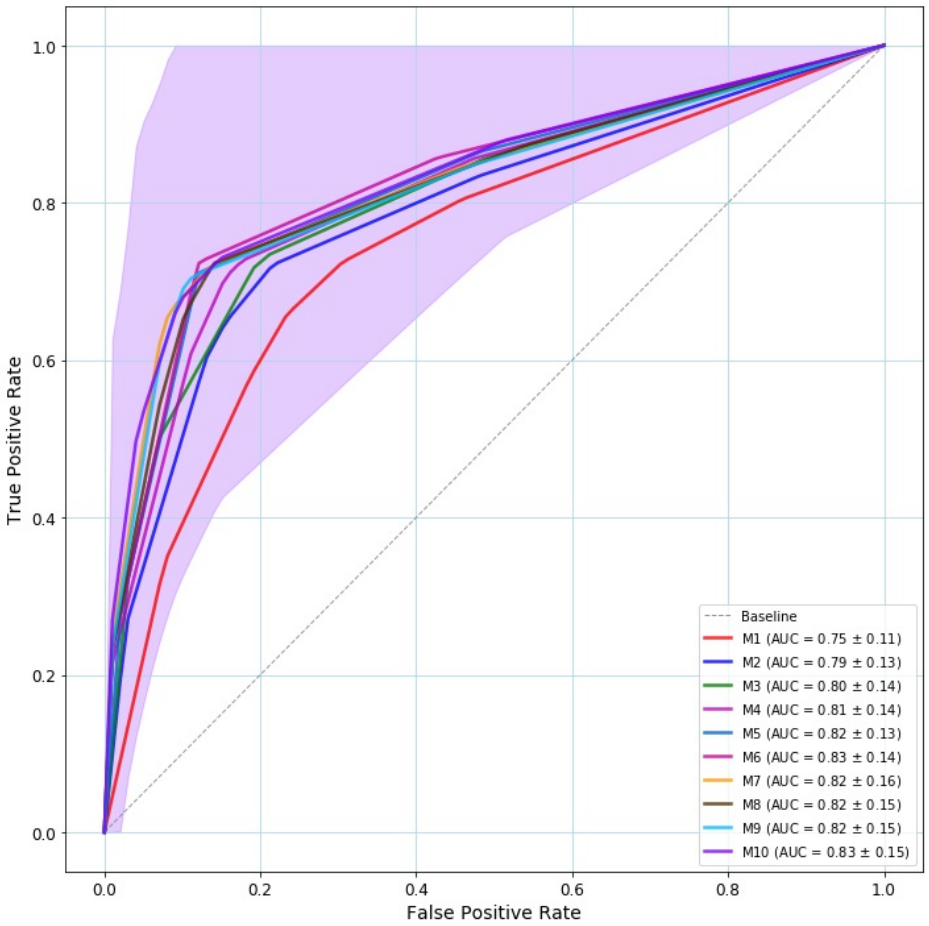

Figure 3. Mean ROC curve and AUC of the 5-fold crossvalidation. Model—$M1 \dots M10$— are represented by solid line with a different color. The baseline is represented by a dashed gray line. All models provided ROC curves that were over the baseline. The best model is the model whose ROC curve approaches the left upper corner. The shaded area indicates the M10's SD that is the outperforming model both in terms of ROC/AUC (83%) and validation accuracy (83%).

To this end, we have automatically transcribed, segmented, and acoustically analyzed Swedish vowel productions. The acoustic properties of vowels, namely their formants ($F1 - F5$), duration, fundamental frequency, age, and gender of participants were employed as predictors. Specifically, ten Deep Neural Networks Architectures were trained on the acoustic productions and evaluated on how well they can identify MCI and healthy individuals, by comparing model predictions (i.e., MCI or HC), with the evaluations conducted by clinicians using combined imaging and neuropsychological examination. We have trained ten models each with a different number of hidden layers. Models 6 to 10 resulted in 83% mean classification accuracy (see Table 4).

One important contribution of this study is that it provides a model that can identify MCI individuals automatically and with high accuracy, providing a quick and early assessment of MCI, by using only a simple acoustic recording, without other neuropsychological or neurophysiological information. Also, it demonstrates that speech acoustic properties play a central role in MCI identification and points to the necessity for more acoustic studies with respect to MCI. Nevertheless, 83% accuracy might still be low for clinical use, if it is going to be employed as the only assessment. Two aspects can account for these accuracy results. First, there is a significant symptom variability

among individuals with MCI, which has been stressed out by a number of papers including consensus papers for the diagnosis of MCI (e.g., 41, 58). Some of these symptoms are not related to speech, thus additional phonemic, moprhosytactic, etc., predictors might increase the accuracy. Also, by increasing the data and retraining the model, it is possible to improve model accuracy as it is evidenced by the fact that some of the crossvalidation folds resulted in considerably higher accuracy (cf., the SD is between 14% and 17%).

Moreover, this study presents the methodological process that can lead to the selection of the classification model of MCI vs. HC and the evaluation techniques that enable the selection of the final model from a set of ten different models. We have discussed two methods: i. validation split, and ii. crossvalidation. In the validation split, model 7 resulted in the highest accuracy, namely 75%. Nevertheless, the validation split is a weak evaluation method as it depends on the data selected as a training set and as a test set; different randomization of the data may provide a different output. It also depends on the split size (e.g., 75% - 25%, 80% - 20%, 90% - 10%). To avoid these confounds, we conducted a 5-fold crossvalidation, which performs multiple splits of the data, depending on the number of validation folds (cf. 48, 46). Most importantly, the significance of the proposed machine learning model formulation is not that it provides a specific model only but also because it offers a process for continuous evaluation and improvement of the model. Therefore, model evaluation and model comparison constitute indispensable parts of machine learning.

Future research is required i. to evaluate multivariable acoustic predictors, e.g., predictors from consonants and non-acoustic predictors, i.e., linguistic features, such as parts of speech, syntactic and semantic predictors, sociolinguistic predictors like the education of the speaker; ii. to establish whether these acoustic variables could be useful in predicting conversion from MCI to dementia; and iii. to create an automated differential diagnostic tools, which will enable the classification of unknown MCI individuals from conditions with similar symptoms (cf., 53). A system of this form, will require more data from a larger population, yet our current findings do provide a promising step towards this purpose.

In conclusion, this study has showed that a Deep Neural Network architecture can identify MCI speakers and can potentially enable the development of valid tools for identifying cognitive changes early and enable multidomain life style interventions and/or pharmacological treatments at the MCI stage, which can potentially delay or even prevent the development of AD and other types of dementia.

APPENDIX

- Text in Swedish: Ordet ö beskriver ett område som är helt avskuret från land och som är omgivet av vatten på alla sidor. Öar kan uppstå när vulkaner höjer sig från havsbottnen eller när vattennivån i havet stiger eller faller. Ett flertal öar uppstod mot slutet av den förra istiden. När isen smälte och vattnet rann ut i havet höjdes vattennivån så mycket att de låga landområdena översvämmades. Idag ser man bara de högsta topparna sticka upp över vattenytan som öar. Djur och växter som på något sätt lyckas ta sig till en avlägsen ö kan sedan vanligtvis inte komma därifrån igen. För att överleva är de därför tvungna att mycket snabbt anpassa sig till den nya omgivningen. De levande arter som finns på öar löper en ständig risk att bli utrotade. Detta kan inträffa när nya djur dyker upp eller när människor kommer dit och börjar störa dem.

- Translated text in English: The word island describes an area that is completely cut off from the land and is surrounded by water on all sides. Islands can occur when volcanoes rise from the seabed or when water levels in the ocean rise or fall. A number of islands occurred at the end of the last ice age. When the ice melted and the water ran out into the sea, the water level was raised so much that the low lands were flooded. Today, only the highest peaks can be seen across the water surface as islands. Animals and plants that somehow manage to get to a distant island usually do not leave the place again. Therefore, in order to survive, they are forced to adapt very quickly to the new environment. The living species on islands run a constant risk of being extinct. This can happen when new animals appear or when people get there and start to disturb them.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

CT conducted the acoustic analysis of the materials, designed and run the Deep Neural Networks architectures and wrote the first draft of the paper. DK supervised the data collection. Subsequently all authors worked on refining and revising the text. All authors approved the final version.

FUNDING

This research has been funded by Riksbankens Jubileumsfond – The Swedish Foundation for Humanities & Social Sciences, through the grant agreement no: NHS 14-1761:1.

ACKNOWLEDGMENTS

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

REFERENCES

1 .[Dataset] Abadi, M. et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems

2 .Ahmed, S., Haigh, A.-M. F., de Jager, C. A., and Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven alzheimer's disease. Brain 136, 3727–3737. doi:10.1093/brain/awt269

3 .Appell, J., Kertesz, A., and Fisman, M. (1982). A study of language functioning in Alzheimer patients. Brain and Language 17, 73 – 91. doi:doi.org/10.1016/0093-934X(82)90006-2

4 .Bayles, K. A. (1982). Language function in senile dementia. Brain and Language 16, 265–280. doi:https://doi.org/10.1016/0093-934X(82)90086-4

5 .Bayles, K. A., Kaszniak, A. W., and Tomoeda, C. K. (1987). Communication and Cognition in Normal Aging and Dementia (Boston, MA: College-Hill Press)

6 .Bayles, K. A., Tomoeda, C. K., and Trosset, M. W. (1992). Relation of linguistic communication abilities of Alzheimer's patients to stage of disease. Brain and language 42, 454–472

7 .[Dataset] Boersma, P. and Weenink, D. (2017). Praat: doing phonetics by computer (version 6.0.32)

8 .Bowles, N., Obler, L., and Albert, M. (1987). Naming errors in healthy aging and dementia of the Alzheimer type. Cortex; a journal devoted to the study of the nervous system and behavior 23, 519–24

9 .Caramelli, P., Mansur, L. L., and Nitrini, R. (1998). Language and communication disorders in dementia of the alzheimer type. In Handbook of neurolinguistics (Elsevier). 463–473

10 .[Dataset] Chollet, F. et al. (2015). Keras. https://github.com/keras-team/keras

11 .Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., and Gershon, S. (1986). Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—report of a national institute of mental health work group (Taylor & Francis)

12 .Cummins, N., Sethu, V., Epps, J., Schnieder, S., and Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. Speech Communication 75, 27–49

13 .de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H. F., and Minett, T. S. C. (2011). Microlinguistic aspects of the oral narrative in patients with alzheimer's disease. International Psychogeriatrics 23, 404–412

14 .Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer"s disease: revising the NINCDS-ADRDA criteria. Lancet Neurology 6. doi:10.1016/S1474-4422(07)70178-3

15 .Fleisher, A. S., Sowell, B. B., Taylor, C., Gamst, A. C., Petersen, R. C., and Thal, L. J. (2007). Alzheimer's disease cooperative study. clinical predictors of progression to alzheimer disease in amnestic mild cognitive impairment. Neurology 68. doi:10.1212/01.wnl.0000258542.58725.4c

16 .Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease 49, 407–422

17 .Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's Disease in narrative speech. Journal of Alzheimer's disease : JAD 49

18 .Fujimura, O. (1967). On the second spectral peak of front vowels: a perceptual study of the role of the second and third formants. Language and Speech 10, 181–193

19 .Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., and Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. Cerebral Cortex 20, 1799–1815. doi:10.1093/cercor/bhp245

20 .Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., and Tyler, L. K. (2013). Functional organization of the Neural Language System: Dorsal and ventral pathways are critical for syntax. Cerebral Cortex 23, 139–147. doi:10.1093/cercor/bhr386

21 .Grut, M., Fratiglioni, L., Viitanen, M., and Winblad, B. (1993). Accuracy of the Mini-Mental Status Examination as a screening test for dementia in a Swedish elderly population. Acta Neurologica Scandinavica 87, 312–317. doi:10.1111/j.1600-0404.1993.tb05514.x

22 .He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. CoRR abs/1502.01852

23 .Henry, J. D., Crawford, J. R., and Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. Neuropsychologia 42, 1212 – 1222. doi:https://doi.org/10.1016/j.neuropsychologia.2004.02.001

24 .[Dataset] Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python

25 .Kemper, S., LaBarge, E., Ferraro, R., Cheung, H., Cheung, H., and Storandt, M. (1993). On the preservation of syntax in alzheimer's disease: Evidence from written sentences. Archives of Neurology 50, 81–86. doi:10.1001/archneur.1993.00540010075021

26 .Kemper, S., Thompson, M., and Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. Psychology and aging 16, 600–614

27 .Khodabakhsh, A., Kuşxuoğlu, S., and Demiroğlu, C. (2014). Natural language features for detection of alzheimer's disease in conversational speech. In IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 581–584

28 .König, A., Satt, A., Sorin, A., et al. (2015). Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1, 112–124. doi:10.1016/j.dadm.2014.11.012

29 .Korytkowska, M. and Obler, L. K. (2016). Speech-language pathologists (slp) treatment methods and approaches for alzheimer's dementia. Perspectives of the ASHA Special Interest Groups 1, 122–128. doi:10.1044/persp1.SIG2.122

30 .Lehmann, C., Koenig, T., Jelic, V., Prichep, L., John, R. E., Wahlund, L. O., et al. (2007). Application and comparison of classification algorithms for recognition of Alzheimer's Disease in electrical brain activity (eeg). Journal of Neuroscience Methods 161. doi:10.1016/j.jneumeth.2006.10.023

31 .Leung, J. H., Purdy, S. C., Tippett, L. J., and Leão, S. H. S. (2017). Affective speech prosody perception and production in stroke patients with left-hemispheric damage and healthy controls. Brain And Language 166, 19–28. doi:10.1016/j.bandl.2016.12.001

32 .Levy, R. et al. (1994). Aging-associated cognitive decline. International Psychogeriatrics 6, 63–68

33 .Luz, S. and la Fuente, S. D. (2018). A method for analysis of patient speech in dialogue for dementia detection. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), ed. D. Kokkinakis (Paris, France: European Language Resources Association (ELRA)), 7–12

34 .Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30^{th} International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. 1–6

35 .Manouilidou, C., Dolenc, B., Marvin, T., and Pirtošek, Z. (2016). Processing complex pseudo-words in mild cognitive impairment: The interaction of preserved morphological rule knowledge with compromised cognitive ability. Clinical Linguistics & Phonetics 30, 49–67. doi:10.3109/02699206.2015.1102970. PMID: 26588013

36 .Mueller, K. D., Hermann, B., Mecollari, J., and Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and alzheimer's disease: A review of picture description tasks. Journal of Clinical and Experimental Neuropsychology 0, 1–23. doi:10.1080/13803395.2018.1446513. PMID: 29669461

37 .Nilsonne, Å., Sundberg, J., Ternström, S., and Askenfelt, A. (1988). Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. The Journal of the Acoustical Society of America 83, 716–728

38 .O'Shaughnessy, D. (1988). Linear predictive coding. IEEE Potentials 7, 29–32. doi:10.1109/45.1890

443 39 .Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011).
444 Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830
445 40 .Petersen, R., Smith, G., Waring, S., Ivnik, R., G. Tangalos, E., and Kokmen, E. (1999). Mild
446 cognitive impairment: Clinical characterization and outcome. Archives of neurology 56, 303–8
447 41 .Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., and Fratiglioni, L. (2014).
448 Mild cognitive impairment: a concept in evolution. Journal of Internal Medicine 275, 214–228
449 42 .Powell, J. A., Hale, M. A., and Bayer, A. J. (1995). Symptoms of communication breakdown
450 in dementia: carers' perceptions. European Journal of Disorders of Communication 30, 65–75
451 43 .Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., and Kaye, J. (2011). Spoken language
452 derived measures for detecting mild cognitive impairment. IEEE Transactions on Audio, Speech,
453 and Language Processing 19, 2081–2090
454 44 .Sarazin, M., Berr, C., De Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., et al. (2007).
455 Amnestic syndrome of the medial temporal type identifies prodromal ad - a longitudinal study.
456 Neurology 69. doi:10.1212/01.wnl.0000279336.36610.f7
457 45 .Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and
458 Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and alzheimer's
459 disease in late life: Findings from the nun study. Jama 275, 528–532
460 46 .Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for
461 classification tasks. Information Processing & Management 45, 427 – 437. doi:https://doi.org/
462 10.1016/j.ipm.2009.03.002
463 47 .Taler, V. and Phillips, N. A. (2008). Language performance in alzheimer's disease and
464 mild cognitive impairment: A comparative review. Journal of Clinical and Experimental
465 Neuropsychology 30, 501–556. doi:10.1080/13803390701550128. PMID: 18569251
466 48 .Tharwat, A. (2018). Classification assessment methods. Applied Computing and Informatics
467 doi:https://doi.org/10.1016/j.aci.2018.08.003
468 49 .Themistocleous, C. (2016). Seeking an Anchorage. Stability and Variability in Tonal Alignment
469 of Rising Prenuclear Pitch Accents in Cypriot Greek. Language and Speech 59, 433–461. doi:doi:
470 10.1177/0023830915614602
471 50 .Themistocleous, C. (2017). Dialect classification using vowel acoustic parameters. Speech
472 Communication 92, 13–22. doi:https://doi.org/10.1016/j.specom.2017.05.003
473 51 .Themistocleous, C. (2017). Effects of two linguistically proximal varieties on the spectral
474 and coarticulatory properties of fricatives: Evidence from Athenian Greek and Cypriot Greek.
475 Frontiers in Psychology 8, 1945. doi:10.3389/fpsyg.2017.01945
476 52 .Themistocleous, C. (2017). The nature of phonetic gradience across a dialect continuum:
477 Evidence from modern greek vowels. Phonetica 74, 157–172
478 53 .Themistocleous, C., Ficek, B., Webster, K. T., Wendt, H., Hillis, A. E., Den Ouden, D. B.,
479 et al. (2018). Acoustic markers of ppa variants using machine learning. Frontiers in Human
480 Neuroscience doi:10.3389/conf.fnhum.2018.228.00092
481 54 .Themistocleous, C. and Kokkinakis, D. (2018). THEMIS-SV: automatic classification of
482 language disorders from speech signals. In Proceedings of the 4th European Stroke Organisation
483 Conference (Gothenburg, Sweden)
484 55 .Thomas, C., Keselj, V., Cercone, N., Rockwood, K., and Asp, E. (2005). Automatic detection
485 and rating of dementia of alzheimer type through lexical analysis of spontaneous speech. In
486 IEEE International Conference Mechatronics and Automation, 2005. vol. 3, 1569–1574 Vol. 3.
487 doi:10.1109/ICMA.2005.1626789

56 .Wallin, A., Nordlund, A., Jonsson, M., Lind, K., Edman, Å., Göthlin, M., et al. (2016). The Gothenburg MCI study: Design and distribution of Alzheimerâ s disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. Journal of Cerebral Blood Flow & Metabolism 36, 114–131. doi:10.1038/jcbfm.2015.147. PMID: 26174331

57 .Weiner, M. F., Neubecker, K. E., Bret, M. E., and Hynan, L. S. (2008). Language in Alzheimer's Disease. The Journal of clinical psychiatry 69, 1223

58 .Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., et al. (2004). Mild cognitive impairment – beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. Journal of Internal Medicine 256, 240–246. doi:10.1111/j.1365-2796.2004.01380.x