Predictive deep learning models for cognitive risk using accessible data

Kenji Karako*
Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan.

SUMMARY  The early detection of mild cognitive impairment (MCI) is crucial to preventing the progression of dementia. However, it necessitates that patients voluntarily undergo cognitive function tests, which may be too late if symptoms are only recognized once they become apparent. Recent advances in deep learning have improved model performance, leading to applied research in various predictive problems. Studies attempting to estimate dementia and the risk of MCI based on readily available data are being conducted, with the hope of facilitating the early detection of MCI. The data used for these predictions vary widely, including facial imagery, voice recordings, blood tests, and inertial information during walking. Deep learning models that make predictions based on these data sources have been proposed. This article summarizes recent research efforts to predict the risk of dementia using easily accessible data. As research progresses and more accurate predictions become feasible, simple tests could be incorporated into daily life to monitor one's personal health status and to facilitate an early intervention.

Keywords  dementia, deep learning, mild cognitive impairment, predictive model

1. Introduction
Globally, the population is aging, with the number of people age 65 and above reaching 727 million, representing 9.3% of the total population of 7.7 billion in 2020 (1). Japan has the world's highest rate of aging, with its elderly population accounting for 28.6% of its total population in 2020. Dementia, and especially Alzheimer's disease, is a significant challenge in such aging societies. The Cabinet Office predicts that by 2025, around 7 million elderly Japanese will have dementia, accounting for 20% of those age 65 and over (2). Globally, dementia cases are expected to rise to 152 million by 2050 (3). Early detection is crucial as many cases progress significantly before becoming apparent, particularly in the early stages of mild cognitive impairment (MCI), which often goes unnoticed due to its minimal impact on daily or social activities. Identifying MCI early is essential to preventing and halting the progression of dementia.

Over the past few years, various studies have been conducted to detect MCI early. A technology called deep learning has been particularly highlighted and utilized. Deep learning is one of the methods in the field known as machine learning. Essentially, machine learning techniques involve using an algorithm to discover features, rules, or patterns existing in the background of the data collected with regard to a certain event or task, and then using those features or rules to make inferences. Deep learning is an improved method of machine learning based on a technique called neural networks. A characteristic of deep learning is its ability to learn features, rules, or patterns from a large amount of input information, but one of the deep learning technologies, convolutional neural networks (CNNs) (4), can locally extract image information and convert it into data of a smaller size. For instance, in tasks where the goal is to discern whether an image contains a cat or a dog, a CNN learns to recognize essential patterns such as eyes, ears, and the mouth. This learning process involves repeatedly extracting relevant information, allowing the network to focus only on the data necessary for image recognition tasks. The CNN learns from a dataset designed for the specific task, including images of cats and dogs alongside the correct identification of each. Deep learning utilizes vast amounts of task-related data and correct answers to develop an algorithm capable of high-performance predictions and feature extraction. Over the past decade,
Deep learning has advanced significantly, demonstrating human-like or superior performance in areas such as image recognition, text generation, autonomous driving, facial recognition, and AI systems like ChatGPT.

Deep learning is increasingly used in medical research, including predicting dementia. Here, studies using deep learning from various perspectives to detect dementia early are described. Conventionally, dementia is assessed using the Mini-Mental State Examination (MMSE) to evaluate cognitive function (5). In addition, brain MRI scans and biomarker tests are used. However, markers like amyloid-beta require invasive procedures, making them impractical for widespread screening and early detection of dementia (6). This highlights the significant challenge of early detection, as opportunities for testing are limited unless patients proactively seek medical help. Moreover, administering the MMSE and performing an MRI scan are costly and time-consuming, making their use as screening tests impractical. Therefore, recent research has focused on developing more affordable and convenient methods of detecting dementia using deep learning. This approach differs from conventional testing methods by focusing on easily obtainable information, such as facial expressions, voice, basic blood tests, and gait data. The potential of these data types to detect dementia early will be detailed further. The key advantage of these sources is their ease of acquisition. If these prediction models evolve to offer a high level of accuracy, they could enable immediate on-site testing, known as point-of-care testing (PoCT), and these tests could be incorporated into daily life. Here, the potential to use deep learning-based methods of estimation for PoCT to detect dementia is summarized.

2. Estimation of MCI using facial images

Research has attempted to estimate dementia based on facial video. The field of image recognition, which has particularly advanced as a result of deep learning, encompasses object estimation, facial recognition, facial expression recognition, and detecting human figures in video. Predominantly developed through CNNs, models like AlexNet (7), ResNet (8), and VGG (9) have emerged to extract features from images, alongside object detection models such as Faster R-CNN (10) and YOLO (11) for real-time detection. These technologies are used in studies to estimate cognitive function based on facial videos (12). Prior studies reported younger-looking facial impressions in individuals without dementia (13), suggesting potential facial indicators of cognitive decline. This research focuses on estimating cognitive functions based on facial videos. For the study, videos ranging from 3 to 30 minutes in length were recorded of 34 elderly individuals age 65 and above, including 10 with MCI. Images were extracted from these videos at a rate of 5 frames per second, with 10 frames over 2 seconds forming one set for the model's input. A total of 3,822 data sets were created, with 3,058 sets used for training and 764 sets for evaluation, to solve a binary classification problem of distinguishing between MCI and health using deep learning. The study used ResNet, which is based on a CNN, to extract facial structure and motion information from facial videos (Figure 1). ResNet, a deep learning model linking over 50 layers of CNNs, was developed for image recognition tasks and is highly effective at extracting features from two-dimensional spatial information. The model to estimate MCI was created using two instances of ResNet: one as a model to extract spatial features from the face to estimate MCI and the other as a model to extract dynamic features based on facial dynamics to predict MCI. The spatial model randomly selects one image from a set of 10 frames over 2 seconds for input, focusing on static facial features. The dynamic model generates an optical flow from the same frame set, reflecting facial movements over 2 seconds, which ResNet then uses to extract features. Optical flow (14), represented by a three-dimensional vector for each

Figure 1. Deep learning structure to estimate MCI based on image and motion information. Facial videos are divided into still images and motion information using an optical flow. Deep learning models are created for each to extract spatial and dynamic features, which are then used to estimate MCI.
pixel, is analogous to the RGB structure of images, making ResNet suitable for extracting features from these data. The model ultimately estimates whether an individual has MCI based on two dynamically and spatially obtained features. The final model had a precision of 0.94, recall of 0.78, accuracy of 0.91, and an F1 score of 0.85. Despite the low recall and concerns over data on a small number of individuals and the balance between MCI and normal data, the ability to determine MCI at a certain level using deep learning represents a significant advance in early detection.

Estimation of cognitive functions based on video data, such as this, is also being performed in another study (15) and is an area of growing interest. If MCI can be estimated based on approximately two seconds of video data, this could allow for testing without a significant burden in everyday life or visits to hospitals and care facilities, enabling immediate examinations on-site.

3. Estimating Alzheimer's disease using speech information

Estimating dementia based on speech information is one of the most extensively studied tasks in the field of deep learning-based estimation of dementia (16). Alzheimer's disease, a type of dementia, initially manifests as language impairments. Focusing on this characteristic, the goal is to estimate the presence of Alzheimer's disease using speech data. Previous studies have reported that Alzheimer's patients tend to pause more frequently between words and speak more slowly than healthy individuals (17). Moreover, Alzheimer's patients are reported to have difficulties in finding appropriate words or expressions to match a sentence (18,19). Deep learning models are used to extract various vocal features from speech data. In order to estimate Alzheimer's disease, two primary features are extracted: features from continuous speech signals and features from speech converted to text to analyze the context and content of conversations. These features are then used for the final estimation task.

In order to extract features from speech signals, studies have used deep learning algorithms that are effective at continuous signal processing (20), such as long short-term memory (21) and recurrent neural networks (RNN) (22). These algorithms have the capability to internally retain a memory of past inputs, allowing the neural network to maintain information over a certain duration. This capacity enables the extraction of features needed to estimate Alzheimer's disease not just based on a single speech sample but also based on historical data. However, they have limitations in terms of storing information over extended periods, such as tens of minutes.

The second method involves converting speech into textual data and then extracting Alzheimer's disease characteristics from the context and content of the text. This approach estimates Alzheimer's based on the coherence and expressiveness of the text. A drawback is that features unique to speech might be missed. However, unlike with direct extraction of speech features, this approach allows for estimation based on lengthy dialogues that have been converted to text. Recent advances in deep learning for natural language processing, such as the use of the high-performance natural language model BERT (23), have led to proposed methods of estimating Alzheimer's using those technologies (24).

Data used to train and evaluate models come from tasks performed during studies. Primarily, tasks include semantic verbal fluency (25), where subjects list as many items as possible from a category like animals or vegetables within one minute (26,27), a natural speech task involving conversation without direct questions (28), and a picture description task where participants orally describe the content of a picture within a set time (29). Notably, the ADReSS database (30) offers open access to data from these tasks, including voice recordings, transcribed texts, and MMSE scores. Such databases are valuable for developing deep learning models to estimate Alzheimer's based on speech data.

4. Estimation of MCI using blood test information

One unique area of research to detect dementia early using deep learning involves blood test information (31). This research focuses on the relationship between systemic disorders like arteriosclerosis, which is the result of lifestyle diseases, and cognitive impairments, which include both MCI and severe dementia (32-34). It also considers other systemic disorders that might affect cognitive function, such as malnutrition (35), anemia (36), lipid metabolism (37), purine metabolism (38), and renal dysfunction (39). These can be detected via basic blood tests obtained during health check-ups. The research attempts to estimate MCI using blood test data, including 23 items like red and white blood cell counts, hemoglobin levels, hematocrit, albumin levels, and age, using a feedforward neural network, a basic form of deep learning, to predict MMSE scores. The input items obtained from the blood tests used are shown in Table 1. This neural network consists of a four-layer structure with intermediate layers as shown in Figure 2. Each intermediate layer has a neural network with 400 nodes, solving a regression problem that estimates the MMSE in the range of 0 to 30 based on 24 numerical items. Data used to train and evaluate the data were collected from 202 patients (average age: 73.48 ± 13.1 years). All patients received inpatient treatment including rehabilitation and pharmacotherapy for lifestyle-related diseases, with 142 patients having cerebrovascular diseases and 174 patients having at least one lifestyle-related disease. The feedforward neural network was trained and evaluated using the leave-one-out method,
which was applied to the blood test results and MMSE scores from the 202 patients. Actual MMSE scores and predicted MMSE scores were correlated ($r = 0.85, p < 0.001$). The mean absolute error was 2.02. Blood tests, primarily obtained during medical examinations and health check-ups, serve as the main data for this research. A cognitive function estimation model based on blood tests could effectively be utilized as a test to screen for dementia in medical facilities and during regular health check-ups. For instance, when elderly individuals undergo blood tests during a health examination or medical visit, their cognitive function can be estimated using deep learning in no time at all. If MCI or dementia is suspected, a medical facility could then encourage the individual to undergo a more detailed examination or visit an outpatient clinic. This estimation model could be an effective means for early detection of dementia, simply by undergoing a regular medical consultation or health check-up.

5. Estimation of MCI using inertial information during walking

Compared to the previously described models to estimate MCI, there is another approach that is more similar to everyday life, and it has the potential to be used for the early detection of MCI by estimating cognitive decline on the spot in everyday situations. Studies have estimated MCI using inertial sensor data collected by a wearable device when the wearer walks ($40,41$). In those studies, a small inertial sensor was affixed to the shin of 30 cognitively normal individuals and 30 individuals with MCI, and they were asked to perform a simple task of walking 20 meters, as well as a complex task of walking 20 meters while simultaneously performing cognitive tasks such as subtracting numbers or naming animals. Moreover, subjects were asked to always keep walking while performing the task. The device used for measurement was the Shimmer3 GSR+ Unit ($42$), which is equipped with a 3-axis accelerometer and a 3-axis gyroscope. Eight pieces of information, including three forms of acceleration, three angular velocities, and the total magnitude of the signal vectors of both the accelerometer and gyroscope sensors, were used to estimate MCI. A six-layer CNN and three types of RNN were used to estimate MCI, as shown in Figure 3. The eight signals input are time-series data, and the $8 \times T$ input information, segmented

![Figure 2. Structure of deep learning used to estimate the MMSE score based on blood test data. It consists of a forward propagation neural network with a four-layer structure.](https://www.biosciencetrends.com)
by a certain time T, is input into the CNN as an image. The features extracted by this process are then input into the RNN. Since the RNN has the characteristic of retaining past input information, the features extracted by the CNN from information before the most recent time T are retained, and features that incorporate time-series information are ultimately extracted. Finally, a binary classification of MCI is performed by a feedforward neural network. The leave-one-out method was used for model training and evaluation, achieving an accuracy of 73.33%, a sensitivity of 83.33%, and a specificity of 63.3%. The walking information used in this study was obtained by attaching a measuring device to the shin, which differs from walking data that can be easily obtained with commonly carried devices such as smartphones or smartwatches. Therefore, this method has not yet reached the point where it can be used for early detection in everyday life as it is. However, as research and data collection progress, this method could be effectively utilized as a method of detecting MCI early since it the sensor is easy to attach and measurement is performed simply by walking, potentially serving as a prompt before visiting a medical facility.

6. Conclusion

PoCT refers to methods that allow for immediate testing on the spot at the appropriate time. Conventional tests for dementia primarily involve brain imaging with MRI, peripheral biomarkers like amyloid-beta, and the MMSE, which are used for final diagnosis. These tests require a certain amount of time to conduct, and moreover, they are opportunities that will not arise unless individuals are aware of their symptoms and go to a hospital voluntarily. Due to the inconvenience of such tests, research has been conducted on methods that can estimate MCI using deep learning based on information that can be acquired more easily, without hassle, and without posing a burden. Deep learning has a high level of inferential performance and can learn from complex data, so data measured during events that indirectly reflect the impact of dementia could be used effectively, something that was difficult to achieve in the past. If high-precision estimation of MCI becomes possible based on information that can be obtained relatively easily, such as speech, facial expressions, blood, and gait, deep learning models will likely be incorporated into testing systems for PoCT. For example, estimation of MCI using facial recognition or blood tests could be combined with regular health checkups for the elderly, allowing for effortless estimation of MCI. Moreover, if estimation of MCI can be achieved using diverse data sources, this could lead to more accurate estimates. Furthermore, if estimation of MCI is widely adopted for PoCT, it could easily be configured into a smartphone or web app. Estimation could be performed at medical facilities and also in nursing homes and at home, so this test could be integrated into one’s daily life. Including simple tests using deep learning in daily life could allow for immediate detection of abnormalities, leading to the discovery of cognitive decline at an earlier stage compared to conventional methods. If tests are conducted daily and the data collected and used for research, this could lead to estimates of future changes in cognitive functions, such as one year or five years later, based on an analysis of daily data collected over time.

Funding: None.

Conflict of Interest: The author has no conflicts of interest to disclose.

References


Received January 12, 2024; Revised February 10, 2024; Accepted February 16, 2024.

*Address correspondence to:
Kenji Karako, Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, City of Kashiwa, Chiba 227-8568, Japan.
E-mail: tri.leafs@gmail.com

Released online in J-STAGE as advance publication February 20, 2024.