Efforts in Analyzing Risks and Factors Concerning Lifestyle-related Diseases and Long-term Care

Akihiro Chiba, Naoki Asanoma, Shozo Azuma, Akinori Fujino, and Syunsuke Aoki

Abstract

For a person to continue a healthy life, it is important for him/her to become aware of the risk of illness as soon as possible and take appropriate countermeasures. To help people live a healthy life, NTT is researching and developing technology for predicting disease risk by using machine learning and analyzing the risk factors from data on health-diagnosis results, genomic information, and data obtained from large-scale cohort surveys. Analysis of genomic data targeting lifestyle-related diseases and analysis of cohort-survey data targeting locomotive syndrome (which is attracting attention as a factor requiring long-term care) are introduced in this article.

Keywords: machine learning, genome data analysis, locomotive syndrome

1. Toward a healthy life that suits each person

We believe that most people have a desire to avoid illness and live a healthy life in old age. However, 19.7% of men and 10.8% of women are suspected of having type 2 diabetes, which is a lifestyle-related disease [1]. Moreover, the percentage of people aged 75 and over who have been certified as requiring long-term care is 23.3%, and that number is increasing yearly [2].

To prevent lifestyle-related diseases and the need for long-term care, it is necessary to recognize health risks as early as possible and consider what countermeasures to implement. NTT aims to contribute to both these steps by using information and communication technology (ICT). To make people aware of health risks, we have been working on (i) predicting the risk of severe disease in diabetic patients by using medical information [3] and (ii) predicting the risk of developing lifestyle-related diseases by using the results of health checkups [4]. We are currently researching and developing technologies for helping people become aware of health risks and take countermeasures.

To develop countermeasures, it is necessary to understand the relationship between the disease and risk factors such as a person's unique characteristics, past medical history, and current lifestyle. In this article, our efforts to analyze the relationship between a person's unique genomic data and lifestyle-related diseases as well as the relationship between cohort-survey[*1] data (including lifestyle habits) and locomotive syndrome for preventing the need for long-term care are introduced.

2. Factor analysis according to individual characteristics

We first introduce our efforts regarding analysis of genome data. NTT and the Institute of Medical Science of the University of Tokyo established the "Project Division of Genomic Medicine and Disease

*1 Cohort survey: A survey to track the incidence of disease in a certain population by conducting periodic health examinations, etc.

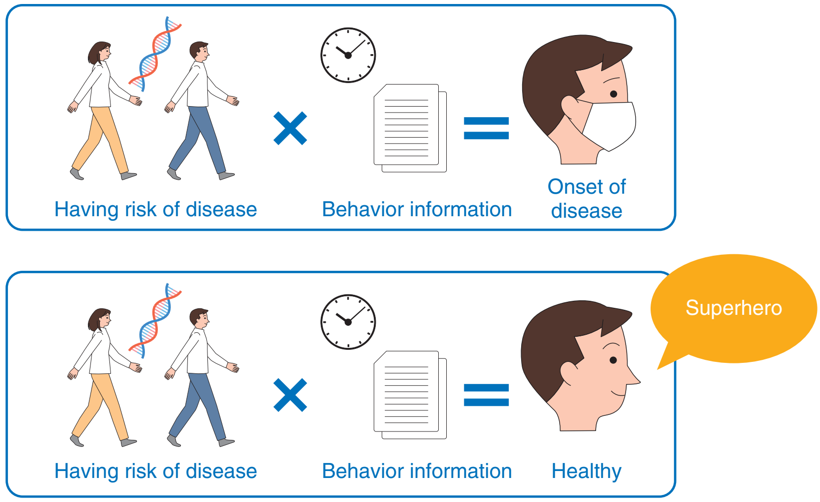


Fig. 1. "Superhero" model.

Prevention" in July 2019, which is a joint research project to elucidate disease-risk factors on the basis of genomic information and visualize desirable behaviors and lifestyles that prevent disease [5]. We are developing a database for collecting and analyzing (i) genomic information obtained from genetic tests as well as (ii) health-checkup results and time-series data about lifestyle habits of company employees who consented to participate in the research [6]. As a result of these developments, it will be possible to conduct research from two aspects; innate factors, such as individuals' genomes, and acquired factors, such as the environment in which people live, that affect their health status. Within the framework of this collaborative research, NTT decided to analyze the characteristics of "superheroes" in regard to maintaining health.

Superheroes in this context are people who, according to their genetic information, have a high risk of lifestyle-related diseases but maintain their health without developing them. This is the model proposed by NTT Life Science, which is a provider of genetic-testing services (**Fig. 1**). To find superheroes from the data, groups that are genetically at high risk or low risk of developing a disease are first divided on the basis of one or more single nucleotide polymorphisms (SNPs), i.e., differences in one base in the DNA base sequences of individuals that are related to the disease and already known from previous research. Other groups are divided on the basis of onset or non-onset of the disease. The general flow of data analysis is to analyze—by using a statistical method—whether there are any differences in data characteristics between combinations of these groups. If the physical condition and behavioral characteristics peculiar to superheroes obtained from the analysis of data on the basis of genetic-test results and health-checkup results are applied, it will be possible to identify factors for maintaining health and provide advice for improving health and link those factors to effective behavior change and disease-risk aversion.

By applying the risk-prediction technology for lifestyle-related diseases developed by NTT [4], it will be possible to predict the risk of disease onset and analyze factors according to the individual characteristics of a person by using genetic information and time-series data of health-checkup results. Lifestyle-related-disease risk-prediction technology makes it possible to predict the risk (probability of onset) until the onset of a disease by learning to rank that takes into account censored data. By combining the information from the risk-prediction model with the information on genetic disease risk, we believe that we can conduct a more detailed analysis of factors that contribute to health maintenance.

3. Factor analysis considering lifestyle, etc.

We now introduce our efforts in preventing the need for long-term care. Preventing such a need is an important social issue in Japan's "super-aged society" because it not only reduces social-healthcare costs but also improves the quality of life of the caregivers and the cared-for. To prevent long-term

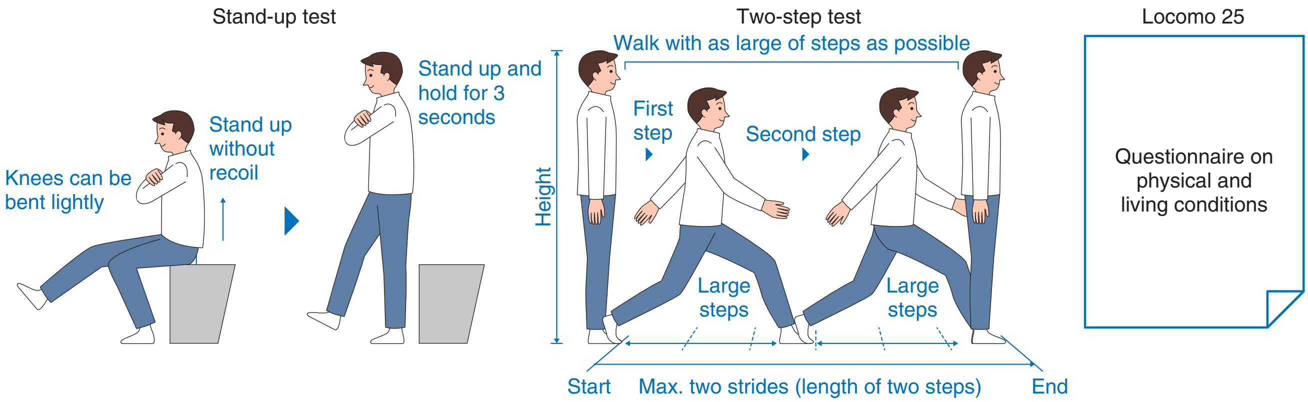


Fig. 2. Locomotive syndrome risk test.

care, we are working to elucidate the factors leading to long-term care and quantify risk by using the analysis technology and knowledge of medical and health data that we have accumulated thus far. To prevent the need for long-term care by using data analysis, data and medical knowledge concerning many elderly people are required. In April 2020, NTT established the Department of Preventive Medicine for Locomotive Organ Disorders at the 22nd Century Medical and Research Center of the University of Tokyo Hospital and began joint research.

In this joint research, we are investigating the causes of diseases related to the need for long-term care and intervention methods from data of about 4400 people collected from a cohort survey conducted over 15 years. According to a survey conducted by the Ministry of Health, Labour and Welfare, dementia is the most common cause of the need for support or care [1]. However, the proportion of fractures, falls, and joint diseases, which are related to the decline of motor and mobility functions, exceeds that of dementia, and the combination of both has become the largest proportion. In this joint research, we are focusing first on the decline in motor and mobility functions.

Locomotive syndrome is a concept that describes the decline in motor and mobility functions. Defined as "Locomotive syndrome means being restricted in one's ability to walk or lead a normal life owing to a dysfunction in one or more of the parts of the musculoskeletal system – muscles, bones, joints, cartilage, or the intervertebral discs" [7], it is an important concept regarding preventing the need for long-term care. Locomotive syndrome is classified using an index called the locomotive syndrome risk level from 1, least severe, to 3, most sever [7]. It is determined from the results of a stand-up test, a two-step test, and the "Locomo 25"*2 test (**Fig. 2**). For example, if one cannot stand up with one leg after sitting on a chair with a height of 40 cm, your locomotive syndrome risk level is deemed as 1. This result means their mobility function has begun to decline.

Although locomotive syndrome is an important concept regarding preventing the need for long-term care, awareness of it is currently around 43.8% [8]. In particular, younger generations are active and less likely to feel the risk of locomotive syndrome. However, we believe that identifying the risk of decline in mobility function from a young age and making early behavioral changes can reduce the risk of locomotive syndrome and ultimately reduce the risk of requiring long-term care. Therefore, to make it easier for many people to understand the risk of locomotive syndrome, we constructed a model for predicting future locomotive syndrome risk level by analyzing a large amount of data obtained from a cohort survey and implemented it in a research application (**Fig. 3**). By answering about 20 questions about age, weight, lifestyle-related topics via the app, users can predict their risk of having the locomotive syndrome risk level of 1 or higher in the next three years.

In a typical cohort survey, more than 1000 analytic elements are collected from health checkups and questionnaires completed by participants. In general, prediction accuracy will be higher if many factors are used in the analysis. Regarding our study, prediction accuracy will be higher if all the elements of the cohort survey are used. However, from the standpoint

*2 Locomo 25: A test in the form of 25 questions in regard to locomotive syndrome.

Fig. 3. Screen of app for predicting the locomotive syndrome risk level.

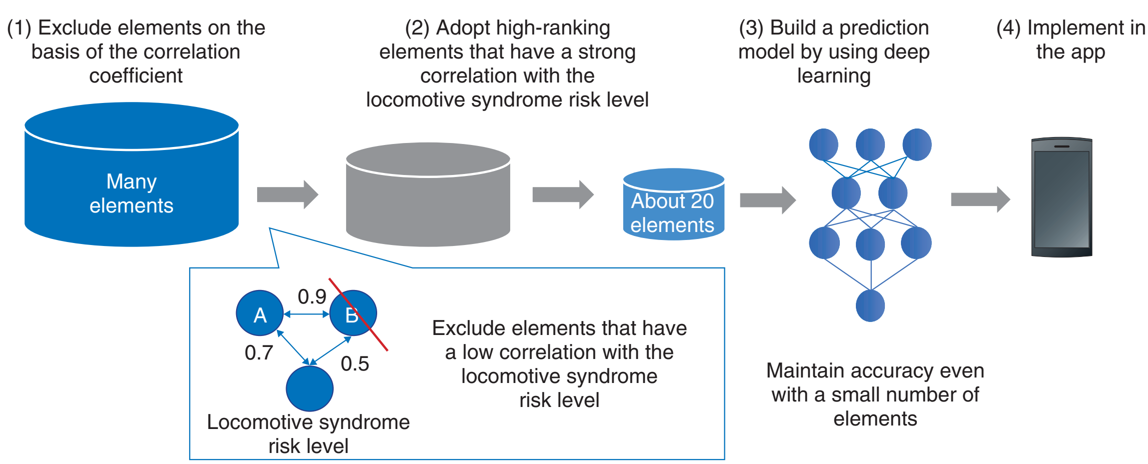


Fig. 4. Procedure from data analysis to application.

of the users of the application, we thought it would be better to obtain risk-prediction results by having them answer as few questions as possible. We therefore focused on the usability of the application and built a model that can predict the risk of locomotive syndrome with fewer elements while maintaining prediction accuracy (**Fig. 4**).

The necessary elements were optimized in accordance with the correlation among them. For example, when elements that were strongly correlated, such as diastolic blood pressure and systolic blood pressure, the element with the lower correlation with the locomotive syndrome risk level, which is the prediction target, was deleted. Therefore, the need for users to answer similar questions was eliminated, and the ability to efficiently enter the information required for prediction was improved. Next, about 20 elements that had a large correlation coefficient with the locomotive syndrome risk level were extracted as the elements that were input into the application. Using the correlation coefficient makes it possible to understand which behavior among the extracted elements contributes to the increase in the locomotive syndrome risk level and to take measures to prevent locomotive syndrome. Finally, a multi-layer neural network that predicts the locomotive syndrome risk level by using deep learning was trained, and a prediction model based on that network was built. It thus became possible to predict the locomotive syndrome risk level with a certain degree of accuracy even in the case of a small number of elements. Since the trained prediction model can be incorporated into the app and easily executed on a personal computer or smartphone, risk can be predicted even during a health checkup, for example.

The elements that were narrowed down included those related to exercise habits as well as questions related to pain. We were able to find elements that were not considered medically important thus far and confirm the value of analysis from this cohort survey data. In the future, we will verify the effectiveness of implementing health education via the app to prevent the need for long-term care.

4. Future of health support using ICT

In this article, taking lifestyle-related diseases and prevention of long-term care as examples, risk prediction and its factor analysis were introduced. We believe that predicting the future through machine learning and artificial intelligence will trigger a review of people's current behavior. We also believe that analyzing the factors that can change the future will help people change their behavior. We expect that from now onwards, various medical and health-related information—such as electronic medical-record data, healthcare data such as blood pressure measured daily, and biological information measured using wearable devices—will be collected. Each piece of data can only represent a part of a person's overall health status; however, by connecting those data, it becomes possible to understand the person from multiple perspectives. As a result, we believe that by fusing and analyzing large-scale data and medical knowledge, we can simulate complex events occurring in the body and reproduce them in digital space. In the future, we want to (i) contribute to developing the bio-digital twin, which will make it possible to simulate an entire person, by developing technologies related to risk prediction and (ii) make it possible to predict unknown events in the body and understand complex events.

References

[1] Ministry of Health, Labour and Welfare, "National Health and Nutrition Survey," 2019.

[2] Ministry of Health, Labour and Welfare, "Long-term Care Insurance Business Status Report (annual report)," 2018.

[3] H. Kurasawa, A. Fujino, and K. Hayashi, "Artificial Intelligence-based Health Management System: Unequally Spaced Medical Data Analysis," NTT Technical Review, Vol. 16, No. 8, pp. 24–28, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa4.html

[4] NTT press release, "Free Trials Utilizing Lifestyle Disease Risk Prediction Technology Performed with Insurance Companies," May 16, 2018.
https://group.ntt/en/newsrelease/2018/05/16/180516a.html

[5] NTT press release issued on June 27, 2019 (in Japanese).
https://www.ntt.co.jp/news2019/1906/190627b.html

[6] NTT press release issued on Sept. 28, 2020 (in Japanese).
https://www.ntt.co.jp/news2020/2009/200928a.html

[7] Locomotive Challenge! Promotion Council, "Locomotive Syndrome Pamphlet,"
https://locomo-joa.jp/assets/pdf/index_english.pdf

[8] Exercise Equipment for Health/Japan Association, "'Locomotive Syndrome' Awareness Survey," 2020 (in Japanese).
https://www.bjd-jp.org/wp/wp-content/uploads/2020/05/2020.pdf

Akihiro Chiba

Researcher, NTT Bio-Medical Informatics Research Center and NTT Smart Data Science Center.

He received a B.E. and M.E. in engineering from the University of Electro-Communications, Tokyo, in 2011 and 2013 and a Ph.D. in information science and technology from the University of Tokyo in 2020. He joined NTT Service Evolution Laboratories in 2013 and studied the analysis of vital sensing data such as heart rate. He is currently studying machine learning for biomedical data.

Akinori Fujino

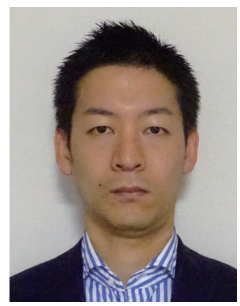
Senior Research Scientist, Supervisor, Learning and Intelligent Systems Research Group, NTT Communication Science Laboratories.

He received a B.E. and M.E. in precision engineering from Kyoto University in 1995 and 1997, and a Ph.D. in informatics from Kyoto University in 2009. He joined NTT in 1997. His current research interests include machine learning and knowledge discovery from complex data.

Naoki Asanoma

Research Engineer, NTT Bio-Medical Informatics Research Center, NTT Smart Data Science Center and NTT Medical Business Planning Office.

He received an M.S. in information and computer science from Waseda University, Tokyo, in 1997 and joined NTT the same year. He is currently engaged in research and development on analysis technology of factors that contribute to health maintenance.



Syunsuke Aoki

Manager, R&D Produce Group, NTT Research and Development Planning Department.

He graduated from the Osaka University Graduate School of Engineering in 2005 and joined NTT WEST the same year. After joining the company, he worked as a system engineer and experienced system development in government and public offices. He was then engaged in cloud service development at NTT Smart Connect and assumed his current position in 2019.

Shozo Azuma

Senior Research Engineer, Supervisor, NTT Bio-Medical Informatics Research Center and NTT Smart Data Science Center.

He received a B.E. in information science engineering from Osaka University in 1994 and an M.E. in information science engineering from Nara Institute of Science and Technology in 1996. He joined NTT Human Interface Laboratories in 1996 and has been engaged in research and development of natural language processing, geographical information systems, agent systems, and multimedia processing.