The Role of Big Data in Aging and Older People’s Health Research: A Systematic Review and Ecological Framework

Xinyue Zhang 1,2,*, Xiaolu Gao 1,2,+, Danxian Wu 1,2,†, Zening Xu 3 and Hongjie Wang 1,2,‡

1 Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; zhangxy.18b@igsnrr.ac.cn (X.Z.); wudanxian15@mails.ucas.ac.cn (D.W.); wangbj.16s@igsnrr.ac.cn (H.W.)
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 China Institute for Urban Governance, Shanghai JiaoTong University, Shanghai 200030, China;
xuzn.13b@igsnrr.ac.cn
* Correspondence: gaoxl@igsnrr.ac.cn
† Current address: Real Estate and Urban Construction Development Research Center, Shenzhen 518031, China.

Abstract: Big data has been prominent in studying aging and older people’s health. It has promoted modeling and analyses in biological and geriatric research (like cellular senescence), developed health management platforms, and supported decision-making in public healthcare and social security. However, current studies are still limited within a single subject, rather than flourished as interdisciplinary research in the context of big data. The research perspectives have not changed, nor has big data brought itself out of the role as a modeling tool. When embedding big data as a data product, analysis tool, and resolution service into different spatial, temporal, and organizational scales of aging processes, it would present as a connection, integration, and interaction simultaneously in conducting interdisciplinary research. Therefore, this paper attempts to propose an ecological framework for big data based on aging and older people’s health research. Following the scoping process of PRISMA, 35 studies were reviewed to validate our ecological framework. Although restricted by issues like digital divides and privacy security, we encourage researchers to capture various elements and their interactions in the human-environment system from a macro and dynamic perspective rather than simply pursuing accuracy.

Keywords: aging; big data; older people’s health; ecological framework; systematic review

1. Introduction

In both developed and developing countries, the healthcare and social security systems are under stress because of the climbing pension, and healthcare expense of the older age group [1–4]. As physical function declines, older adults struggle with a higher risk of age-related diseases (ARDs) and more difficulties to live their daily life with descending mobility and ascending learning barriers with updating techniques [5]. Aging issues inherently involve older people’s health, from the cell, organ, or individual to the family and society [6].

Recently, big data has presented outstanding contributions to the research on aging and older adult health, especially when playing the roles of data collection and analytic tools. In terms of the micro-level, it has promoted statistical analysis in biological and geriatric research in identifying, classifying, and predicting biomarkers related to cellular senescence and ARDs by means of data integration, data sharing (including open biomedical databases and electronic health records (EHRs)), and large-scale distributed computing [7]. At the meso-level, it has supported a preliminary interdisciplinary interest, which combines computer science and medical science, to establish health management systems and platforms, and home-based assisted living environments to provide pre-diagnosis...
and treatment of ARDs [8,9]. As for macro-level studies, it has upgraded the architecture of tele-medicine from e-health (e-health) and mobile health (m-health) to smart health; in other words, tele-medicine has participated in public health and social security since it involved not only patients and physicians but also the network of family and citizens as smart health is a part of smart homes and smart cities [10].

However, current studies are still limited within a single subject, rather than flourished as interdisciplinary research in the context of big data. As healthcare has been considered critical to cope with the aging society and sustainable development, many studies have delivered much efforts to explore the potential of big data in elevating effectiveness and efficiency, which generally reflect sustainability in resources and finances [11–13], although some researchers noted that lifestyle [14], and the interaction between humans and environment with social, cultural, political, and economic elements should also correspond to sustainability. Studies are expected to present more inspirations to understand the coupling of human-environment systems in the process of aging. This is the gap between data and knowledge outcomes [15]. The role of big data as an analytical tool has been emphasized over its role as a data product and resolution service. It has been a step away from the expected changes in perspective pointed out by Mayer-Schnberger and Cukier [16]. Combining the studies mentioned above, when examining big data-based research in the dynamic, multi-scale human-environment system, it is obvious that big data, as a data product, is the link to integrate different dimensions and disciplines through data-sharing and integrating. As an analysis tool, it is the reflection of interactions among the physical, environmental, economic, social, cultural, and political elements through the cell, organ, individual, and society via modeling. As a resolution service, big data presents a complete embodiment of integration and interaction inside the human-environment system.

With the opinion of big data as the key of integration and interaction in interdisciplinary research, this paper attempts to propose an ecological research framework for big data based on aging and older people’s health research and validates the framework following the process and requirements of a systematic review; 35 documents have been screened. We firstly demonstrate the concept of aging, old age, older adults, and big data in detail to describe how the “5V” characteristics of big data match aging research in its multiple scales and dimensions, described later in the ecological framework. Through a literature review, the existing studies have mainly exerted the characteristics of large “volume” and high “velocity” and have partially made use of “variety”, though limited by the digital divide and privacy security. Most of all, it is the low “veracity” and low “value” that leave us unsettled, as they are much harder to overcome, for the reason that they require insights and inspirations to extract information from data and to share knowledge.

2. Aging, Old Age and Big Data: Concepts, Context and Framework
2.1. Concepts and Context
2.1.1. Aging

Aging issues are mostly health-related. With respect to aging, economists and sociologists are concerned about populations aging, which is generally noted as the result of the growing life expectancy and falling birth rate. It is widely believed that aging challenges are mainly related to the conflict between the structure of the aging population and the structure of the national economy [2,17], especially considering the expense of healthcare and pension. Biologists, physicians, and psychologists concentrate on the functional deterioration of cells and organs as well as mental disorders. However, only individuals age [18], rather than the population, even though the accumulated aging individuals have aroused many social concerns with regard to public healthcare, housing, building environment, transportation, pension policy, and well-being as the economic burden of the productive population. Therefore, health is the core of aging.

As for health in aging, it is necessary to trace the biological origin. There are three leading theories, namely, mutation accumulation, antagonistic pleiotropy, and disposable soma, explaining why species age in an evolutionary way. Antagonistic pleiotropy supposes that
natural selection loses its influence over genes processing senescence, which means that genes might become deleterious (actively or passively) in a later stage, even though they were programmed to benefit individual fertility [19,20]. Another conjecture suggests that it is because of the protecting environment of developing society that aging and ARDs have become the major contributors to modern mortality of population, while natural mortality usually results from extrinsic hazards like infection, predation, starvation, or cold before individuals get old [21]. Later in the 1950s, biological aging research made advances in cellular senescence focusing on how the aging process performs in the way of chemical secretion, metabolic reactions, signalling pathways, and biomarkers [22]. There has been a period of free radical theory of aging (FRTA) speculating that the cumulative damage in determining life span results from the ubiquitous involvement of endogenous oxygen radicals in metabolic reactions [23–25]. Furthermore, it has come up with inflamm-aging, suggesting senescence as a disease syndrome [26].

There are, likewise, three leading theories in social gerontology about aging, disenagement theory, activity theory, and continuity theory [18]. The disengagement theory describes aging as a time when older persons engage in mutual separation with society, for instance, retiring from work. This disengagement is regarded as a natural and normal tendency in the biological rhythm of life. However, retirement in contemporary society could be dynamic in time and style. Nowadays, older people probably experience a shifting engagement from working relationships to private and lively relationships in their later life. The activity theory of aging argues that the more active people are, the more likely they are to be satisfied with life. The continuity theory of aging notes that people are inclined to maintain the same habits and lifestyles as much as possible when growing older [18]. Together, these two theories assume that people do not really get older if they continue their earlier lifestyle and values. However, it is not appropriate to overlook physical function decline and aging when considering the social context.

Briefly speaking, aging is a significant process and dynamic progress. Biologically, aging is the process of progressive deterioration, rather than a fixed period, while demographically it is the process of accumulating an older age group. Moreover, it is socioeconomically and historically constructed, processed, and progressed during interpretations [18]. Many studies suppose a better allocation, higher efficiency, and greater productivity of public services as a solution to the aging society; however, it has been troubled during this one and half centuries since the first aging society occurred in Europe. As observed and mentioned, aging and health involve different scales and different elements within the human-environment system which requires researchers to provide perspectives over the system to sort the muddles out dynamic interactions. Progressive problems need to be settled in progress, and studies need to be dynamic-thinking to seize that progress.

2.1.2. Old Age

To understand the dynamics of aging is a prerequisite for understanding the uncertainty of old age, which comes from historically constructed views of old age. From the increasingly mentioned perspective of life course, old age indicates the later stage of life, which can delay healthcare as the average life expectancy climbs. It was inferred that the average life expectancy of Chinese people during the 19 century was less than 40, while this number has now reached 77, in 2019. The aged population of the past might now correspond with young people enjoying their prime of life. In the agricultural society, it is more likely that damage, from physical changes or the outside environment, causes the turning point into old age, which shapes the stereotype of old age as unhealthy and unpleasantly struggling. After the retirement introduced into the industrial society, withdrawing from work and social relations has taken the role as the beginning of old age. In contemporary information society, it is not surprising that a retired couple enjoys a period of energetic and healthy life as “post-middle age” before experiencing the general “later life” with a significant decline in physical functioning or suffering from ARDs. Note that retirement age is a historically and socially constructed symbol rather than a biologically or even
chronologically stable determination of old age. It seems that we might not assert which is later until the end, and perhaps we should determine old age and later life according to the human life span.

2.1.3. Big Data

Big data has gone through a rapid increase over the past decade. It is generally recognized as “a large amount of unstructured data continuously generated that could not be collected, stored, managed, and processed within a tolerable time by general computer systems or traditional technologies” [27]. This is characterized by the “5V”, namely, “(large) volume, (high) variety, (high) velocity, (low) veracity, and (low) value (density)” [28,29], which makes big data prominent in aging and health research for the reason that it largely enriches the data category and information within the same time to conduct fieldwork. Traditional data used for studying elderly adult health include laboratory data, clinical records, census, and survey data (including cross-sectional data, longitudinal data, and panel data). This data is limited to sample size, spatial extent, and time span by implementation costs. Particularly, it is contradictory to acquire detailed data from a large area covering a long time, which is more available through big data instead of traditional investigation. Available data include, for example, the Health and Retirement Survey (HRS, U.S.), the Japanese Study of Aging and Retirement (JSTAR), and the China Health and Retirement Longitudinal Study (CHARLS). These studies usually publish their data months after the start-up of the investigation. Furthermore, the data is hard to exceed quantities of tens of thousands, while big data provides millions of data points that are continuously generated.

In addition to the large volume and high updating velocity of data, it is also attractive that big data technically translates information into the data; information that was previously processed by the brain like posture, expression, and emotion, into machine-readable binary data. It is also remarkable that commercial engagements, especially social media and internet enterprises, have contributed to maintaining the available data of user preferences and behaviours like flavor, location, trajectory, and health status from smart phone and application recordings. Then, it is researchers who are responsible for interpreting data into a human readable information. Considering its low value density and low veracity, methods like machine learning, network analysis, and knowledge graphing are used in data mining, which reflects the data-intensive paradigm [30]. Boyd et al. note big data as a thinking and computing turn [31], though analytic methods often stand out over data, criticisms have called for knowledge engagement when data mining, in case of repeating validation and verification without revealing anything new.

2.2. Ecological Framework of Big Data in Aging and Older People’s Health Research

As repeatedly mentioned, big data is not merely an analysis tool but plays different roles as statistical support, architecture support, mapping support, and decision-making support in regard to research in different dimensions and scales. From this perspective, we propose an ecological framework of big data for studying aging and older people’s health to explain how big data should work as interaction and integration among disciplines inside the human-environment system and in what perspective researchers should extract information from big data analyses to contribute to the understanding of aging, old age, and the aging human-environment rather than promoting accuracy without inspiring insights (see Figure 1). Although the current roles of big data do not escape from the scope of analytic tools, we insist a comprehensive view of human-environment system significant to better interdisciplinary research, when realizing that all elements are under dynamic processes and interactions with each other. This is why big data is distinguished as the reflection of this process and interaction.

This framework is structured by the axis of the space scale, time scale, and organization hierarchy that refers to the dynamics dimension and process of aging and old age. The cube reflects the human-environment system, which comes from the geographic perspective.
After all, humans are the core of aging and health, and our physiological function, psychological cognition, and behavior closely interact with our living environments—both the natural, social, and built environments. The red circle represents big data in Figure 1, illustrating its role in integration and interaction through space, time, and organization (discipline).

The spatial and temporal scale depict the spatiality and temporality of aging. Since this is a process of individual and demographical change, it obviously time-sensitive and space-sensitive that the aging process goes forward or backward across different areas over time. As we emphasize, aging, old age, and older people are time-dependent and progressive, particularly in a time of society development in economy, politics, and technology, in short, the civilization rhythm. Meanwhile, society is also space-dependent, since it changes with the expansion of space. Society spatially emerges in neighbourhoods, villages, towns, cities, states, countries, and the world. Aging probably hits one society rather than another and puzzles one in a higher/lower scale rather than another lower/higer one.

Organizational hierarchy is somewhat similar to the space scale in respect to society, and reflects the hierarchy of the study object from biological aging and individual aging to social aging, as shown in Figure 1, corresponding to the hierarchy in Section 4. Yet, it also denotes corporations, authorities, and other institutions related to big data collection, storage, management, processing, sharing, and maintenance. In this sense, it describes an interdisciplinary characteristic to research aging and older people’s health, considering that organizations are under and focusing on different disciplines to resolve aging issues, such as internet enterprises and computer science, authorities relevant to health, transportation, housing, employment, finance, and social security, and social organizations concerned with the environment, ecology, and social service. As acknowledged, big data has increased progress in discipline integration and intersection.

Figure 1. Framework. ① denotes the subjects of aging in different spatial extents like cell, organ, individual, and population from small towns to worldwide. ② denotes different processes of aging through time from cellular senescence in months to evolutionary aging in millions of years. ③ denotes the analysis of tempo-spatial patterns of aging; ④ denotes interactions and coordinations of elements in different dimensions and scales inside the human-environment system, which refers to the role and position of big data. ⑤, ⑥, and ⑦ indicate current study domains as the results of the aging function \( f \), mapping from the human-environment system to engaged disciplines with \( f(x,y,z) \).
3. Materials and Methods

3.1. Search Strategy

Following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [32], a literature scoping process was conducted by searching in two information retrieval platform, Scopus and Web of Science (WoS) (see Figure 2). Scopus, the largest abstracts and citations database of peer-reviewed literature, provides EMBASE and MEDLINE, while WoS contains a core collection with SCI/SSCI, MEDLINE, and SciELO as well as a non-English database, including Chinese Science Citation Database, KCI-Korean Journal Database, and Russian Science Citation Index. This paper did not expected to concentrate on specific processes and parameters of clinical experiments in geriatrics or life sciences, but it was designed to grasp the mainline and examine our ecological framework of big data applications in aging and older people’s health research. The searching strategy was to identify studies related to individual aging, population aging, and health using big data methods, noticing that topics of health involve lots of subjects such as geriatrics, healthcare devices and equipments, housing environment, and public services, facilities and policies, and well-being. Meanwhile, the records were limited to English publications since 2008, the year when big data was highlighted by Nature. Please refers to Supplementary Table S1 for a summary of the search terms and query strings used in this review.

Figure 2. Process of systematic review following PRISMA.

3.2. Study Selection

3.2.1. Inclusion Criteria

Studies were included providing that the (1) subject relates to older people’s (or older population’s) health; (2) participants (if there were) were aged at least 50; (3) materials
correspond to big data, e.g., open source big database, or at least large volumes (e.g., ten thousand) with sophisticated database management; and (4) methods associated with big data analyses, especially methods of machine learning.

3.2.2. Exclusion Criteria

Documents would be excluded if they are (1) editorials, conference abstracts, reports, or reviews; (2) irrelevant topics like material aging, less relevant topics like pension policy, long-term care policies regarding aging as background information without data analysis, and disease-related studies in which patients scarcely aged older; (3) eligible documents but ranked as low quality after screening, specifically, employing big data as a background to design conceptual smart healthcare frameworks, prototypes, or platforms without practical operations and older participants, focusing on algorithm modification without promoting better understanding of the old age and older people, part of the participants aged under 50; (4) lack of full paper. The amount of records in each inclusion or exclusion process is shown in Figure 2.

3.2.3. Quality Assessment

A quality assessment was also conducted in the aspect of standpoint, material, methodology, manuscript organization, and citations to select proper literature corresponding to the exclusion criteria (see Table 1). Any article classified as low quality in more than two items was excluded; in fact, there were few studies with only one item of low quality, unless it belongs to the kind of focusing on the performance of methods.

Standpoint was used to distinguish studies that were potentially pushing the field of aging and older people’s health forward rather than providing statistics which are waiting for further inspiring explanations. We used material and methodologies related to determining whether big data was adopted in an eligible manner instead of merely big volume. Considering the general requirements of conferences and proceedings, the pages and references were used to filter less qualified work.

Table 1. Quality assessment of initial included records.

<table>
<thead>
<tr>
<th>Quality Item</th>
<th>Quality Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Standpoint</td>
<td>(1) focusing on interpreting health-related factors to cope with later life and aging society</td>
</tr>
<tr>
<td>Material</td>
<td>(1) open source big database</td>
</tr>
<tr>
<td>Methodology</td>
<td>(1) big data platform like MapReduce, Hadoop</td>
</tr>
<tr>
<td>Manuscript organization</td>
<td>/</td>
</tr>
<tr>
<td>Citation</td>
<td>(1) 25 citations over 5 years;</td>
</tr>
</tbody>
</table>
3.3. Outcome

3.3.1. Overview of Search Result

In total, 1956 papers were exported from databases, in which 1657 were screened through the abstract, after removing duplicates. There were 1366 documents discarded, such as studies about material chemistry, geology, energy engineering, pediatrics, and strategic agendas, etc. Finally, 35 were qualified following a full-text assessment.

Among the search results, before filtering by publish year, 2011 marks the beginning of an explosive increase of big data applications in aging and older people’s health research. Meanwhile, 2017 marks a forward-turning point from conceptual strategies to empirical research on ARDs, healthcare devices or platforms, and data analyses supporting decision-making in public health. In respect to leading countries, United States (8), Korea (8), and China (7) account for the largest portion of research, while Spain (2), Japan (2), Denmark (1), Algeria (1), Romania (1), Portugal (1), Israel (1), Singapore (1), New Zealand (1), and India (1) share the remaining. Apart from the affiliations of the lead author and the corresponding author, cooperators are also from UK (2), Canada (2), Belgium (1), the Netherland (1), France (1), Russian Federation (1), Malaysia (1), and Australia (1).

3.3.2. Research Topics

As for the topics of research interest, nearly 70% of the inclusions (24/35) concentrate on gerontology, namely, biomarkers (3) [33–35], frailty index (1) [36], epigenetics of aging (1) [37], aging of brain functional connectivity networks (1) [38], ARDs (16) that contain dementia (6) [39–44], stroke (2) [45,46], Parkinson’s Disease (1) [47], fracture (1) [48], hypertension (1) [49], mild cognitive impairment (1) [50], depression (1) [51], pressure ulcers (1) [52], polypharmacy side-effects (1) [53], and mortality related to sarcopenia and frailty (1) [54]. The rest are devoted to technical support and decision support like cloud-based healthcare platforms focusing on dementia (1) [55], memory recall training (1) [56], fraud detection in medicare (2) [57,58], prediction of readmission risk (2) [59,60], well-being (2) [61,62], population portraits (1) [63], built environment and health outcomes (1) [64], patterns of living activities (1) [65], trajectories (1) [66], and geospatial patterns of points of interest (1) [67].

3.3.3. Algorithms and Models

The methods were wide ranging, but machine learning stands out the most. Neural networks (11) took the top frequency and it consisted of convolutional neural networks (6), deep neural networks (1), long short-term memory networks (1), multilayer perceptrons (1), deep unified networks (1), and naïve bayesian networks (1). Secondly, it’s the Support Vector Machine (8), and the following random forest (6), boosting (6), including Extreme Gradient Boosting (XGBoosting, 1), gradient boosting (1), AdaBoost (2), decision tree (4), bayesian methods (2), bagging (2), association rules (2), k-means clustering (1), and classification and regression tree (1). Furthermore, semantic analyses (2) works as well, and three studies used the open big data framework MapReduce (1) to manage and store the study data. Moreover, researchers also designed their own novel measurements and modifications based on generalized linear models (4), including logistic regression (3), swarm intelligence clustering (1), and Probabilistic Latent Spatial Semantic Modeling (1).

4. The Role of Big Data in Studying Aging and Older People’s Health

All of the included articles (and some of the excluded ones) were reviewed to sort out the main advances in aging and older people’s health research from using big data, and to analyze how researchers adopt big data to execute analyses from the perspective of our ecological framework.
4.1. Statistical Support in Biology and Medicine

4.1.1. Identifying Biomarkers in Cellular Senescence

Big data is not only a data product from experiments helping establish open database like ProteomicsDB and Swiss-Prot database [68]), but also provides statistical support in inference, stochastic modeling, and cluster testing in studying aging phenotypes and biomarkers [22] about identification, clustering, classification, correlation tests, and signal pathway simulation.

The reviewed papers in cellular senescence refer to biological aging research through temporal scales in the framework, such as, to identify oncogene-induced senescence regulators using machine learning-based high-throughput functional screening and interrogating, to classify staining chemicals [37]. Benmounah et al. corroborated the slight variation of DNA-methylation during aging by a distributed big data clustering solution based on the MapReduce framework, incorporating particle swarm optimization, ant colony optimization, and artificial bee colony analysis [55]. Sagers et al. focused on the variation and combination of laboratory biomarkers in predicting chronological age (CA; i.e., time-dependent age) and biological age (BA) by means of machine learning modified linear regression [34], which revealed the linear and non-linear relationships across age groups, genders, and race groups. Biomarker is one promising subject to provide clinical anti-aging therapies with a quantitative foundation. Mamoshina et al. modeled a hematological aging clock with multilayer feed-forward neural network trained by data from different race groups, considering the mixed influence of race, geographic, environment, and behavior in order to measure and predict CA and BA. The analysis results showed that low-glucose diets and metformin are positive as anti-aging interventions [33]. Besides, researchers carried out a whole-body MRI-based analysis of biopsychosocial parameters, cardiovascular indexes, metabolomics, lipidomics, and microbiome variables in association with anthropometrics, demographics, and socioeconomic data to quantify the subclinical disease burden occurring with aging in all organ systems [35], or experiments using fMRI to mine the functional connectivity changes during brain aging [38].

4.1.2. Diagnosis, Therapy and Prevention of ARDs

To date, efforts in archiving and integrating clinical and research data have witnessed the development of open databases, such as the Biomedical Translational Research Information System (BTRIS), The Cancer Genome Atlas (TCGA), The Cancer Imaging Archive (TCIA), the immunology database and analysis portal (ImmPort) system, and the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [69]. Meanwhile, public health databases are also continuously advancing with the promotion of electronic health records (EHR). Thus, data mining has become popular in studying ARDs.

Disease studies correspond to both biological (medical) and individual aging research through the temporal scale in Figure 1. Early research focuses on detection and diagnosis. Specifically, Kim et al. constructed a dementia diagnosis and prediction model using support vector machine with sociodemographic data, biological examination indices, and personal medical and disease history from The Korean National Health Insurance Service Senior Cohort Database (KNHIS-SC) [42]. Coughlan et al. developed a game-based spatial navigation assessment combining genetic parameters and demographic parameters to distinguish at-risk individuals of AD for further personalized treatment [70]. Later, it was extended to more subjects like identification of disease risk factor [40] and risk prediction of multimorbidity and comorbidity [71] using markers (biochemicals and anthropometrics such as blood pressure, blood sugar, heart rate, waist-to-hip ratio, BMI, etc.) and demographic data including age, gender, race, economic status, diet, and self-reported medical history; hospital readmission for financial concerns [59,60], personalized drug prescriptions and polypharmacy management [53], fragility index assessment (FI) [36], clinical subtype classification [43], mortality prediction [54,72], death causes [73], and spatial distribution [74].
In short, a general advancing roadmap in studying cellular senescence and ARDs is presented for detection, diagnosis, classification, reasoning, monitoring, treatment management, and medicine development with big data methods varying from numeric data mining, semantic text analysis, network theories, image processing, knowledge graphing, and sensor-based pattern recognition.

4.2. Technical Support in Health Care System

4.2.1. Smart Devices in Personal Health Management

Personal health management has been advanced since the promotion of portable electronic health (e-health) devices, then later, mobile health (m-health) devices after the development of information and communication technologies, especially the widespread use of smartphones [75,76]. Considering the possibility of the centenarian cohort born after the millennium [17], it is undoubtable that personal health management takes a significant position, as older adults grow fragile, to accomplish conventional daily activities.

Together with assisted living environments, smart device research represents the main big data applications for individuals aging through both temporal and spatial realms, Figure 1. Initially, personal health management began with portable devices to monitor basic physiological indicators, such as body temperature, blood pressure, blood sugar, and heart rate, then the monitoring developed from breakpoint recording to continuous data collecting [77]. Later driven by big data, health services have advanced, combining simple recordings with early disease diagnosis [78] using motion data (e.g., walking, waving, sitting, etc.) [79], voice signals [45], and video (facial expression) with machine learning-based models for real-time monitoring and analyzing [80], posture recognition and warning [81], daily activity patterns analyzing [82], health interventions providing treatment advice and auxiliary measures [83], and personalized treatment and rehabilitation measures [84–86]. Thus, it provides life assistance for older adults when physical function declines, improving the quality of life (QoL) through earlier intervention. The devices have evolved from large equipment and portable terminals (such as smartphones or wearable sensor like wristbands) [87,88] to embedded furniture with integrated systems, discussed in the next section [89,90].

Among these living assistances, fall detection and the monitoring of older drivers’ health have drawn the most attention. In terms of data types, it can be divided into sensors and image data. The types of fall detection vary from inertial sensors for gait recognition [91] like accelerometers [92], and vibration sensors [93], to RFID-based systems to avoid obstacles [94]. Accordingly, these devices usually include laboratory-assembled sensors [95], modified commercial systems (like smartphones [96] and Microsoft Kinect [97,98]), or hybrid environments based on cloud computing [99]. As for older driver’s health, the issues are briefly divided into car-based health monitoring (in case of sudden stroke [100,101]), cognitive assessment for driving ability via in-car driving behavior as drivers have to handle the trip complexity [102,103] and classification in regards to driving behaviours [104].

4.2.2. Ambient Assisted Living Environment and Smart Home

Ambient Assisted Living (AAL) systems imply the further integration and extension of “non-invasive” smart devices and equipments [105]. Systems are first home-based, then probably extended outwards to be a part of smart cities to help fragile older citizens regain independence and life control and social participation. In this environment, independent devices are embedded as nodes, and big data technologies serve as linkages to support data storage, management, transmission, and computation, together with the Internet of Things (IoTs) to establish the environment network, which shows a preliminary anticipation of interactive individuals to social aging across different temporal and spatial scales in our framework.

Based on contextual behaviour analysis [106] and emotion awareness [80], using videos [107], voices [108], and thermal data [109], AAL offers not only health management integrated from independent equipment, but also indoor security monitoring [107], like decision-making assistance in medicine [106], as well as healthcare strategies [110],
caregiving assessments [111,112], power system monitoring, and emergency warnings for the elderly living alone [113]. Experimental projects have been conducted to expand AAL from indoor spaces to outdoor public spaces, that is, projects for smart cities, such as e2Wall [114], STIMULATE, AGNES, and City4Age [115].

However, studies are inclined to the model design and modification for higher accuracy by methods such as adjusting algorithms in relation to different sensors and data type, or algorithm designed for data storage, transmission, and real-time analysis. Similarly, lots of researches have focused on designing architecture, prototypes or frameworks of smart homes based on big data [116,117], and analyzing performance via available big data platforms like Spark and Hadoop [118,119]. However, studies are limited to the cost of real deployment and operation, as most are laboratory examinations by researchers or older volunteers. There is scarcely a contribution to the concept of aging processes and old age. All these situations lead to their exclusion in screening.

4.3. Mapping Support in Human-Environment Interactions

Within the concept of active aging, older adults are suggested to be active participants in urban life. As shared and open source geographic big data accumulates, it is considered beneficial to recognize and classify mobility patterns [120]. According to our ecological framework, using geographic big data to analyze older people in daily activities correspond to the range from individual to population across different temporal and spatial scales, and help to understand the aging process individually through their activity changes.

Individual behaviours can reflect the inner function changes during the aging process. It is worth studying the purpose, frequency, strength, pattern [67], and characteristics [65] of the movement, as well as the space and place during these interactions, especially focusing on changes as elderly gradually lose independence and mobility [121]. With GPS location available, life-space mobility (proposed and developed by Cantor [122,123]) has become favourable to reflect the health status of the elderly in respect to cognitive health and independence [124]. Apart from discovering the function loss, digital map data, remote sensing data, and street view data facilitate analyses for environmental exposure impacts on human health, in view of human-environment interactions, reflected by daily movement (for instance, assessing the quality of neighbourhood air, water, green spaces [125], food, and medical service accessibility [126–128]. Different environments shape patterns on elderly activities, physical and mental health status, and response to satisfaction [64]. Particularly, older people are more likely to be affected as they have a harder time fitting into the environment, while historically accumulated influences occur. Furthermore, data mining with trajectories help identify geographic boundaries of activity patterns in case of getting lost, especially in older people suffering from cognitive impairment [66].

Furthermore, big data can be used to distinguish the spatio-temporal characteristics of older people on public transportation, which relates to appraised transportation policies [129]. Moreover, there is a thriving prospect for research on health-oriented migration in later life, provided that cross-province follow-up investigations are enabled by mobile GPS data. Though some of these studies currently lose sight of the older age group, their efforts set up a favourable example in applying big data to aging and older people’s health research.

4.4. Decision-Making Support in Public Services

According to the research mentioned above, big data provides full decision-making support from data, system architecture, analysis, and mapping methods into integrated solutions, which are essential to smart city and smart health resources to improve the efficiency of the healthcare system, to alleviate financial burden, promote regional health equity, and realize individual well-being.

For example, it is considered necessary to integrate EHRs, laboratory data, and social media data into evidence-based medicine research [130,131], as the fact that government responses have been encouraging during COVID-19 with location tracking and data shar-
ing. The global burden of disease (GBD) project is another practice using big data to support decision-making [132], which reflects positive international cooperation and efforts to human well-being. Studies attempt to improve efficiency, and some have also conducted risk control, like detecting medical fraud to optimize the settlement of claims and avoid unnecessary expenses [57,58]. Some studies have used big data to prevent over-treatment [133] and evaluate factors in relation to successful aging through semantic analysis of social media data [62]. It also occurs in AAL that decision-making support recommends a practical healthcare scheme; however, in public service, the participation of big data moves from individual aging to social aging by aggregating demographic data for hospital monitoring [134] and epidemic control. Meanwhile, with more detailed user data, the assessment of healthcare resources can better improve the consistency between demands and provisions. It is supposed that big data is a prosperous solution in reducing economic burden and medical pressure in an aging society.

5. Discussion

Initially, big data helps discover trends and provide solid validation of previous studies. In terms of empirical research, big data offers facts to observe and methods to analyze in a more cost-efficient way. With machine learning boosting, prediction analysis draws most of the interest, but keeps the interpretation of predictions as a black box. Furthermore, the expected interdisciplinary research remains unmet as non-computer scholars substitute their classic statistical analyses with big data methods. Analysts likely remain new to aging and gerontology studies while gerontologists likely remain confused by neural networks. Given that the promotion in connotation, process, and mechanism of aging and old age falls behind the model and data analysis, the lack of producing knowledge and information [15] confront big data with criticisms. As Kwan argued, there are considerable influences that computerized algorithms might bring to research results for knowledge production [135], and it is rather crucial to develop theories on patterns and paths of physiological and psychological aging and the aging society with big data and algorithms revealing trends and exceptions.

This is closely related to the sparsity problem of big data that the low value density and low veracity of big data require extraordinary efforts of researchers to come up with elaborate designs for data mining. Thanks to the large volume and variety of big data, it encourages the possibility of interesting results while grasping the hidden details from null data, increases the difficulty—especially of understanding the dynamic concepts of aging and designing algorithms. Besides, many papers has been devoted to discovering trends via large amounts of data, which might show no advances in comparing trend analysis with traditional data. It is to capture and interpret the exceptions that distinguishes big data as great science progress. Considering that algorithms are designed and manipulated as “discursively framed by previous findings, theories, and training” [15], it is indispensable that knowledge participation occur as an interdisciplinary cooperation by algorithm engineers and gerontologists to communicate in their domains and disclose the chain of “data-information-knowledge”.

However, there are still two obstacles in the way. The first and most basic is the digital divide, referring to the different levels of participation in the internet between age groups due to a lack of devices and capabilities [136]. This impedes the generation of big data and balance of producing big data, although the EHRs and bus card data make up for it somewhat. For example, poor elderly citizens can hardly afford the cost of broadband, computers, or mobile phones, while some have difficulties in learning and using online applications. Thus, those who are disengaged with the internet fail to generate big data. From a historical perspective of technological development, the digital divide is not a unique phenomenon of the information society, but a specific presentation of the “technology divide”. Different age and education groups have grown under different technological backgrounds, but live in the same technological moment; therefore, the technology divide inevitably occurs because of the differences in owning and using the
latest technology, especially when aging people encounter cognitive and learning decline. Again, this phenomenon implies the interactions of individual to society within the human-environment system through time.

The second challenge is regarding privacy security, which has been on the top of controversies and concerns since the big data era began, particularly considering that health data consists of massive ethical privacy issues [137,138]. If it turns against research users for privacy security reasons, big data immediately loses inspirational significance for data mining. Moreover, the privacy issue impedes the implementation of data sharing, which impressively facilitates the move into big data. Traditionally, the release of health surveys and research is strictly examined to ensure informed consent and ethical reviews. However, it is hard to execute a thorough screening in the internet industry to ensure data security. If privacy leakage occurs, possible problems include financial scams, identity theft, cyber-violence [139], health-related gene discrimination, and public safety disasters [140]. Although regulatory authorities are working on data auditing and protection, there are challenges and difficulties. These include the extent organizations collect user data, the extent it is appropriate for organizations to profile users and train their recommendation models without offending privacy security, and the loss of autonomy given that machine learning creates an information cocoon.

6. Concluding Remark

Aging is both a global challenge and a great opportunity for mankind to further reveal the evolution of life in aging, and to reform the social structure and society form. In this paper, we propose an ecological framework from the geographic perspective of the human-environment to manifest the capability and limitation of big data to integrate different research subjects and dimensions across multiple spatial and temporal scales. This study summarizes current studies using big data to benefit statistical analyses in biology and medicine, technological architecture in smart health management, interaction mapping in daily activities, public policy-making, and social security maintenance.

There is loads of work to be done, such as to advance the big data infrastructure for older people (including personal terminals to bridge the digital divide, open databases), as well as to come up with clear protocols for data sharing and privacy protection with comprehensively examination of data collection, utilization, and publication. As for academic research, it is fundamental for human-oriented and human-environment perspectives to focus on the physical, psychological, and behavioural characteristics and patterns of older people, the aging process, and the demands and needs in their later life—especially in vulnerable groups such as elderly nesters, migratory older adults, and the disabled. Therefore, research is important to cope with the economic, political, and cultural changes in the aging society. Although there are constraints, like the digital divide, privacy security, and sparsity, impeding analyses, the interdisciplinary intersection of data science, gerontology, life sciences, public health, urban planning, sociology, economics, and so on, in knowledge participation will foster a human-environment system of aging research, where each and every subject plays both input (analysing big data) and output (generating big data) to facilitate data-information-knowledge.

All in all, big data has great potential, more than high accuracy and efficiency, and we encourage researchers to capture the elements and their interactions using this framework, to set studies in the human-environment system from a macro and dynamic perspective.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/su132111587/s1.

Author Contributions: Conceptualization, X.G. and X.Z.; methodology, X.Z.; investigation, X.Z.; resources, X.G.; writing—original draft preparation, X.Z.; writing—review and editing, X.G., D.W., Z.X. and H.W.; visualization, X.Z.; supervision, X.G.; project administration, X.G.; funding acquisition, X.G., D.W. and Z.X. All authors have read and agreed to the published version of the manuscript.
Funding: This work was funded by the National Natural Science Foundation of China (NSFC), grant number 41871171; and Key Projects of Social Sciences Foundation of Beijing, grant number 19YTA004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The comments and suggestions of reviewers and editors are admired. The valuable discussions from with Xiongwei and Aurora Lee are much appreciated.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

5. Morato, J.; Sanchez-Cuadrado, S.; Iglesias, A.; Campillo, A.; Fernandez-Panadero, C. Sustainable Technologies for Older Adults. Sustainability 2021, 13, 8465. [CrossRef]
8. Chui, K.T.; Alhalabi, W.; Pang, S.S.H.; De Pablos, P.O.; Liu, R.W.; Zhao, M. Disease diagnosis in smart healthcare: Innovation, technologies and applications. Sustainability 2017, 9, 2309. [CrossRef]


86. Diraco, G.; Leone, A.; Siciliano, P. AI-Based Early Change Detection in Smart Living Environments. *Sensors* 2019, 19, 3549. [CrossRef]


89. Tseng, S.W.; Hsu, C.W. A Smart, Caring, Interactive Chair Designed for Improving Emotional Support and Parent-Child Interactions to Promote Sustainable Relationships Between Elderly and Other Family Members. *Sustainability* 2019, 11, 961. [CrossRef]


112. Zhou, J.; Zhang, B.; Tan, R.; Tseng, M.L.; Lim, M.K. Using Neighborhood Rough Set Theory to Address the Smart Elderly Care in
111. Matsumoto, Y.; Ogata, K.; Kajitani, I.; Homma, K.; Wakita, Y. Development of IoT robotic devices for elderly care to measure daily
108. Hossain, M.S.; Rahman, M.A.; Muhammad, G. Cyber-physical cloud-oriented multi-sensory smart home framework for elderly


125. Chen, Y.; Yue, W.; Rosa, D.L. Which communities have better accessibility to green space? An investigation into environmental inequality using big data. *Landsc. Urban Plan.* 2020, 204, 103919. [CrossRef]


