**Abstract:** With the emergence of technologies such as electronic health and mobile health (eHealth/mHealth), cloud computing, big data, and the Internet of Things (IoT), health related data are increasing and many applications such as smartphone apps and wearable devices that provide wellness and fitness tracking are entering the market. Some apps provide health related data such as sleep monitoring, heart rate measuring, and calorie expenditure collected and processed by the devices and servers in the cloud. These requirements can be extended to provide a personalized life expectancy (PLE) for the purpose of wellbeing and encouraging lifestyle improvement. No existing works provide this PLE information that is developed and customized for the individual. This article is based on the concurrent models and methodologies to calculate and predict life expectancy (LE) and proposes an idea of using multi-phased approaches to the solution as the project requires an immense and broad range of work to accomplish. As a result, the current prediction of LE, which was found to be up to a maximum of five years could potentially be extended to a lifetime prediction by utilizing generic health data. In this article, the novel idea of the solution proposing a PLE on an individual basis, which can be extended to lifetime is presented in addition to the existing works.

**Keywords:** Life Expectancy (LE); Personalized Life Expectation (PLE); Predicted Life Expectancy (PrLE); Mobile Health (mHealth)

**1. Introduction**

The problem of processing datasets such as electronic medical records (EMR), and their integration with genomics, environmental factors, socioeconomic factors and patient behavior variations have posed a problem for researchers in the health industry. Due to the evolution of data science technologies such as big data virtualization and analytics, data wrangling and with the cloud, health workers now have an improved way of processing and developing meaningful information from huge datasets that have been accumulated over many years. For example, a case study [1] shows that big data and machine learning techniques can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy and anxiety disorders. They used the demographics of selected regional areas and multiple behavioral health disorders across regions to find correlations between individual behavior indicators and behavioral health outcomes. Smart environment and wireless network technologies [2–5] have also been used to improve the monitoring of chronic diseases with the evolutions in the Internet of Things (IoT) and cloud computing by building smart cities and homes, which allowed the rapidly growing elderly population to access healthcare resources in a cost-effective way.

Banaee et al. [6] proposes the idea of a potential application model that can be elaborated on as an application of using personal health status. As described in Figure 1, it may be possible to create a prediction of personal life expectancy, which can be further used to calculate health indexes on a generic level for which the individualized expectancies may be compared against.
This model can also depict the life expectancy predicted by an inference system, which transmits health data over wireless sensor networks [8]. Generic life expectancies may be calculated from multiple sources obtained and analyzed by big data. These values can be used to create a personalized graph that most resembles the individual in question, with consideration for personal characteristics such as their age, gender, ethnicity, living environment and current comorbidities or lifestyle habits. There may be an enormous number of variables to consider, of which increasing the number may obviously increase the accuracy of a LE prediction. This personalized graph can then be compared with other individuals who may be living similar lifestyles and share similar traits to provide an idea of the generic life expectancy. This concept is described in Figure 1 with the red and black lines superimposed on each other. During point A, the graph shows that the user was ill and there was a decrease in the overall health status until point C when the user recovered, followed by another case of illness at point C until they recovered finally at point E. Point F describes the present moment in time, at which point any values following this point would be an inferred prediction of the individual’s health status for the future. This prediction would be an inference made of physiological data analyzed by the cloud computation. Whilst the red line shows a trend line of the history of the user’s health status and a rough estimate of his or her future health, the graph may also show fine trends as depicted by the green line. This would be used to inform short term information about whether the user’s health is improving or declining on a more detailed level.

\[ F(x) = \frac{d(g[PQ])}{d(x)}, \text{ where } x = \text{ age, and } g[PQ] = \text{ health index function from point } X1 \text{ to } X2, \quad (1) \]

The Health index function \( g[PQ] \) is based on the differential of health status between age \( X1 \) and \( X2 \). For example, when \( F(x) < 0 \), the health index function would suggest a decline in health whereas \( F(x) > 0 \) suggests that the health status is improving (recovery) during the period of \( A \) (\( X1 \)) and \( B \) (\( X2 \)). Periods of illness may be reflected in the health index by a sum illness cycle, which would describe the number and the total durations of illness days. Inference should take into account the information held within its own database along with that received from external agents, such as from monitoring centers (MC) or caregiver terminals (CT) in mHealth networks.

As presented in [8], the health status of each user would vary from the average values and thresholds of other users, and fine calibration or optimization may be required and could be undertaken by a trained professional, for example a physician. A combination of variables may alter the inference result, as some sensed data may be interdependent and affect one another; e.g., insulin levels and blood pressure. Thus, in inferring whether a change in the user’s health status should trigger a health alarm warning of a potential health event, sensed data should be interpreted together with related variables that are known to affect each other. As time progresses, a user’s profile data is continuously optimized due to the number of inferences made over a longer period of time, and inform a database as
to the PLE prediction of each person. As this data is increasingly updated, inferences can be completed more accurately and efficiently.

Smartphones act as a manager within a mHealth network architecture and as an agent in an IoT network, thus exist as an important link for communication between MC, CT and IoT networks after receiving information from sensor nodes. This article describes a new approach towards individualized life expectancy predictions. It discusses health status calculations from inferring data collected by sensor nodes along with calculations done using big data in the cloud. Inference is suggested to be done within the sensor nodes, in order to have control over data sent to the cloud. Smartphone applications can be used to track health status histories and to access future predictions created from inferred data.

As raw and processed data consist of sensitive information, strong security measures are required to protect user privacy such as personal details (e.g., date of birth, identification). Security aspects are therefore critical in dealing with health-related information such as privacy in wireless body area networks (WBAN), and has been included for review in this article despite extensive and comprehensive research having being completed by many researchers in network and information security in the health area [9–11]. During app implementation, development of algorithms and logics on the predictions for individual and generic data need to be considered as this is crucial to analyzing and producing meaningful and useful information. These aspects have not been completed yet as network infrastructure and merging of technologies continue to be immature for these purposes. This article reviewed existing works and identified that no work has been attempted to predict a personal life expectancy and contributes to the area with an idea of how life expectancy can be predicted on a personal level by providing detailed variables of required data and practical network topology that can be used for producing the predictions.

2. Analysis of Existing Works

As a result of the evolution of biotechnologies and related technologies such as the development of sophisticated medical equipment, humans are able to enjoy longer life expectancies than previously before. For example, a clinical research center claims that in 10 to 12 years from now, for every year that humans live, science is extending the life for more than a year using health intelligence platform integrating genomics, advanced clinical imaging and robust machine learning in a spa-like setting [12]. Predicting a human’s life expectancy has been a long-term question to humankind [13], and there have been many attempts to make the prediction accurate and popular since the prevalence of smartphones and apps. However, the effectiveness of those apps is limited due to the constraints of developing a classification of meta-data, such as the complexity and variety of environmental, geographic, genetic, and living factors of humans. For example, a report showed that people living in a village called Yuzurihara in Japan, also known as “the village of long life”, were ten times more likely to live beyond the age of 85 than anywhere in North America. These people also had similar traits such as smooth skin, flexible joints and thick hair [14]. This implies that geographic and living environments affect the longevity of human life, and the use of statistics can make it possible to forecast a life expectancy of a person who lives in a similar environment village with a similar lifestyle.

Whilst the calculation of life expectancy is a complicated process and requires many variables and circumstances to take into account, there have been several attempts to create an equation despite it being impractical to simplify these variables into one equation. For example, Bhosale and Sundaram [15] suggested a simple equation to compute a line of curve using only three health data including heart rate, blood pressure level and respiration rate.

Jafelice et al. [16] attempted to compute the life expectancy of the HIV/AIDS (Human immunodeficiency virus infection and acquired immune deficiency syndrome) population using a fuzzy set-based model. Based on the assumption that viral load and CD4+ (cluster designation positive cells) level are the most important variables to characterize HIV infection, they concluded that AIDS has a direct influence in the mortality rate of a population.
Agrawal et al. attempted to calculate the survival probability to estimate a 5-year survival of a patient from a recent hospital visit. Their approach of using an online web tool taking 24 patient attributes as inputs to generate a PrLE of five year survival calculation is the most relevant and accurate so far [17]. To select the best modeling technique, they analyzed the top five modelling schemes, which had accuracies from between 89.82% and 90.29%.

Figures 2 and 3 shows a screen menu to calculate a five year life expectancy calculation and its result.

![Figure 2](image1.png)

**Figure 2.** Questionnaires for five-year survival probability. The results will estimate a given patient’s probability of surviving at least five years from the last hospital admission (Reproduced with permission from Ankit Agrawal [18]).

![Figure 3](image2.png)

**Figure 3.** Results of the Questionnaires. Healthy patient—Median model-predicted probability of 5-year survival of patients who actually survived 5 years. Given patient—This corresponds to the chances of survival of a patient for at least 5 years, calculated based on the provided values of the patient attributes. Sick patient—Median model-predicted probability of 5-year survival of patients who did not survive the 5 years (Reproduced with permission from Ankit Agrawal [18]).
Gil-Herrera et al. presented a theory to predict the life expectancy of terminally ill patients for life expectancy prognostication. They used retrospective data of over 9000 patients to assist terminally ill patients in making end-of-life care decisions [19]. However, these life expectancy have only been shown for a specific group of patients such as terminally ill patients only.

There is an association of economic and environmental factors to life expectancy as discovered by Kerdprasop et al. [20]. They tried to show the connection between natural resources and the economic growth to the life expectancy of people in the countries along the Mekong river, and the outcome reveals a strong relationship between environment and GDP growth to the life expectancy of the population.

Pascariu et al. [21] proposed a methodology to forecast life expectancy of males and females using the double-gap (DG) approach: (1) the gap between female life expectancy and the best practice trend in the world, which determines future female life expectancy; and (2) the gap between male and female life expectancy to obtain the country specific male life expectancy. They compared the results with the Lee–Carter (LC) model [22] and the Cairns–Blake–Dowd (CBD) model [23], both of which generate a matrix of forecast death rates and the forecasted life expectancies that are computed using standard life table calculations. Their research approach includes a vast range of data sources, which contain mortality data for 47 homogeneous populations in different countries and regions. The time period was between 1950 and 2014 with the ages ranging from 0–95 in 38 countries and regions. They constructed a model for forecasting the life expectancy of males and females for a given age, using correlations amongst countries and gender.

Whilst this may show an improved forecast of male and female life expectancy by taking countries into account, it is still far from providing an individualized life expectancy as they more focused on PrLE of correlations among countries and between sexes as opposed to individuals. Figure 4 depicts the comparison of DG against other methodologies showing differences of the results with better prediction.

Figure 4. Actual and forecast life expectancy (USA) at birth and at age 65 generated by the DG (double gap), LC (Lee and Carter model) and CBD (Cairns et al.) models for females and males, 1950–2050. Prediction intervals at 80% and 95% levels are shown only for the DG model (Reproduced with permission from Marius Pascariu [21]).

There have been many works done in the prediction of LE within the context of specific diseases such as predicting LE in men diagnosed with prostate cancer [24], schizophrenia [25], lung cancer [26,27], breast cancer [28], diabetes [29], limb ischemia [30], IgA nephropathy [31], oral...
squamous cell carcinoma\cite{32}, acute myeloid leukemia\cite{33}, and psychiatric disease\cite{34}. These models have also been researched with consideration of geographical and regional areas\cite{35–37}, income\cite{38}, relationship with economic growth\cite{39}, and globalization\cite{40}. There are multiple models of LE including Sullivan\cite{41}, Lee and Carter\cite{22}, Cairns et al.\cite{23} and Jafelice\cite{16}. Table 1 shows LE prediction works based on each category such as diseases, geographical areas and regions, relationships with other factors such as income and economical variables as well as different modelling approaches. Table 2 shows summary of variables used for the modelling technique with contributions made by up to date methodologies.

Table 1. Summary of Reviewed Life Expectancy (LE) Predictions.

<table>
<thead>
<tr>
<th>Diseases Based</th>
<th>Geographic Areas</th>
<th>Relationships</th>
<th>Modelling Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer\cite{24}, schizophrenia\cite{25}, lung cancer\cite{26,27}, breast cancer\cite{28}, diabetes\cite{29}, limb ischemia\cite{30}, IgA nephropathy\cite{31}, oral squamous cell carcinoma\cite{32}, acute myeloid leukemia\cite{33} and psychiatric disease\cite{34}</td>
<td>America, Europe, and Asia\cite{35–37}</td>
<td>Income related\cite{38}, relationship with economic growth\cite{39} or globalization\cite{40}</td>
<td>Sullivan\cite{41}, Lee and Carter\cite{22}, Cairns et al.\cite{23}, Jafelice\cite{16}, Agawal et al.\cite{17} and Pascaliu et al.\cite{21}</td>
</tr>
</tbody>
</table>

Table 2. Summary of Modelling Techniques.

<table>
<thead>
<tr>
<th>Modelling Techniques</th>
<th>Authors</th>
<th>Variables</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>\cite{41} Sullivan</td>
<td>Skin color, sex, at birth and age of 65</td>
<td>Used a life table model to calculate LE based with/without disability for mortality and morbidity.</td>
<td></td>
</tr>
<tr>
<td>\cite{22} Lee and Carter</td>
<td>Age groups, period-specific intensity index, direct time series methods</td>
<td>Used singular value decomposition model to account for almost all the variance over time in age-specific death rates as a group.</td>
<td></td>
</tr>
<tr>
<td>\cite{23} Cairns et al.</td>
<td>Mortality-rate dynamics at all ages to simulate the distribution of a survivor index over various time horizons</td>
<td>Proposed calculating the market risk-adjusted price of a longevity bond using two-factor model stochastic for the development of longevity through time.</td>
<td></td>
</tr>
<tr>
<td>\cite{16} Jafelice and Barros</td>
<td>Viral load and CD4+ level</td>
<td>Developed fuzzy set-based model to compute the life expectancy of HIV infected populations to determine the average number of individuals and the life expectancy for specific population groups with no anti-retroviral therapy</td>
<td></td>
</tr>
<tr>
<td>\cite{21} Pascaliu et al.</td>
<td>Male or Female at the age of 0 and 65</td>
<td>LE forecasts based on the double-gap life expectancy forecasting model and compared with the Lee and Carter approach and the Cairns et al. strategy. The Double-Gap model approached the use of combination of models for the most promising forecasting tools.</td>
<td></td>
</tr>
<tr>
<td>\cite{17} Agrawal et al.</td>
<td>Electronic health records (EHR), Age, sex, health data including vitals</td>
<td>Used predictive models for five-year life expectancy of patients, built on electronic health records (EHR) of nearly 7500 patients aged 50 and above with more than 75 modeling configurations. They developed an online tool which takes a non-redundant subset of 24 patient attributes as the input and generates a patient-specific prediction of 5-year survival.</td>
<td></td>
</tr>
</tbody>
</table>

As a result, for the purpose of this article it is concurred that Agrawal et al.\cite{17} provides the most accurate and useful methodology to calculate individual LE despite it only showing five years of LE.
This has so far been the only tool that has been proposed to calculate a PLE for an individual utilizing personalized attributes entered in to questionnaires provided by a web application, whilst others are based on predictions of LE for a group based on larger demographic classifications. However, the variables required to enter to their questionnaire on the website are complex and users are unlikely to enter most of the data due to the difficulty of technical data to obtain. Data that are of a technical nature would be reasonably expected to be entered in by a health practitioner as shown in Figure 2.

Challenges appear in integrating wireless technologies such as wireless local area networks (WLAN) and wireless personal/body area networks (WPAN)/WBANs with existing health networks to meet the quality of service (QoS) requirements for different kinds of medical devices and applications in an integrated and ubiquitous network [42]. Whilst most mHealth users are exposed to wireless networks such as WiFi, which connects to public networks such as cellular, there are compliance issues of using medical sensor devices that may affect WiFi access points as well as cellular networks in terms of electromagnetic field exposure regulations. A government agency such as the European Council issued a directive as to the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields) [43,44]. As there have been advances in wireless technologies for healthcare applications and networks, many works in areas such as characterizing electric field (E-field) exposure within the wireless (WiFi) environments [45,46] and the evaluation of electromagnetic dosimetry of wireless systems in complex indoor scenarios with human body interaction [47] or in sensor networks [48] have already been done. Thus, it is not within the scope of this review article to review the impacts to humans by sensor devices and mHealth networks environment whilst it is important to consider those challenges along with managing privacy issues.

Another challenge is the access of health data of general users, whose data may not be stored in medical facilities or databases as many users may not have visited the hospital in the past as well as the lack of uniformity in how data is stored. Additionally, each country and state may have their own regulations in the handling of patient data and security management. It sounds simply to centralize health data by integrating heterogeneous health networks and to combine them in one location. However, the aforementioned laws and regulations may present to be a big hurdle as it relates heavily to the issue of privacy and consent. For example, a user in a developed country may not agree to provide his or her data to be used in another country’s facilities or networks. Once data have been integrated to a centralized cloud server, a separate issue remains to develop algorithms of how to classify and process big data to create meaningful information. Whilst the latter is a technical issue, the former will be critical as it requires government policies and regulations with multiple parties.

When dealing with health-related information and applications converged with IoT which may interact with health devices, adaptive security management techniques are essential to provide a sufficiently secure and robust system. Savola et al. [49] developed a context-aware Markov game theoretic model for security metrics risk impact assessment to evaluate and validate the run-time adaptivity of IoT security solutions by analyzing objective decomposition strategies for IoT eHealth applications. When healthcare information systems have been implemented and used along with IoT, which are deemed to be extremely sensitive as they contain patients’ private information, the healthcare system should guarantee adequate protection of the confidentiality and integrity of patient information. At the same time, the patient information and their health data also needs to be available for authorized health service providers, so that they can provide the proper treatment of the patient [50]. These information and data should be available immediately for urgent cases and the method and ways of accessing them should not be complicated. In other words, ease of access to the data and providing security should be accomplished at the same time, which can be a dilemma. To achieve the goal, it requires changing the business models and the way the IT infrastructure is being delivered as well as the underlying architecture of how they deliver the applications. Zhang and Liu attempted to develop an electronic health record (EHR) security reference model using a use-case scenario describing the security countermeasures and state of the art security techniques as a basic security guard [51]. As mHealth is increasing in prevalence with smartphone applications,
there have been works done in the underlying networks and application security services integrated into health networks by looking for mobile specific requirements \[52,53\]. Due to the demand, IEEE 11,073 workgroup formed a sub-workgroup for cybersecurity to support secure health data exchange between personal health devices (PHD) and their clients. Their objective of the cybersecurity standards is to specify the process and capability of preventing unauthorized access, modification, misuse, denial of use, or the unauthorized use of information that is stored on, accessed from, or transferred to and from a PHD \[54\]. Therefore, it is required to define and distinguish the constraints of communication and collaboration when developing policies of processes and systems \[11\] since healthcare systems cannot provide security as a one for all solution in the context of heterogeneous networks and different environments and regulations in each country and states. It also requires the consideration of complexities such as heterogeneous networks, organizational silos and applications requirements, compliance for regulations, security and network capabilities, and costs when building a security solution. For example, light-weight security may be developed for security functions on a smartphone application of group life expectancy against a personal application that may need to protect private information as the size of software directly affects the performance of small devices, which have power constraints.

3. Review Conclusions

Whilst there have been a few attempts to calculate the LE for classified groups and diseases, little has been done to foresee and predict generic or personalized life expectancy, and could be attributed to several contributing factors.

1. The overwhelming number of variables that could be considered in predicting LE;
2. Lack of accumulated data in one storage location for data processing and analysis to generate meaningful data;
3. Difficulty of centralizing heterogeneous networks from different countries and regions;
4. Unpredictable and fast changing lifestyle of humans with the increase of sophisticated technologies;
5. Limited methods of health data collection such as data only from patients in a healthcare environment (hospitals, insurance companies), which may exclude the general population.

Due to the fast-emerging technologies of big data, machine learning, artificial intelligence, and virtual and smart environment, it may be possible to handle some of the constraints above. Moreover, the prevalence of smartphones in the working generations even in developing countries makes it a possibility to collect health data anytime and to predict how long a person can live based on their health and medical history. Security aspects are essential as it deals with sensitive personal information and should be considered in PLE solution with various security levels for the nature of information. This paper proposes an idea of how prediction of life expectancy to an individual user can be achieved as well as exercising how mHealth can be utilized to help collect and process health data from users.

4. Work in Progress

This section describes an individualized information of life expectancy through health data collected, processed and created based on big data. Information is provided by two ways: (1) smartphone app for brief summary; and (2) a report with a full and comprehensive report processed by big data in the monitoring center of the health service providers. There has been a great level of work in health data processing including knowledge management, managing medical images and data, integration of hospital care data, clinical research and bioinformatics data and information system security and data protection \[55\]. We are not focusing on existing works in data processing, which have already been identified and reported. Similarly, existing health networks can be used to provide PLE service along with mHealth applications. Therefore, established networks and systems that have been broadly researched have not been discussed \[56\]. Rather, we look at the data collection by mobile users who can use their smartphone and application to access the PLE services.
4.1. Raw Data Collection

Data to collect for building generic and personalized LE information include many parameters to integrate such as behavioral parameters, health status monitoring, physiological signals (e.g., blood pressure, etc.), diet, physical activities, and genetics. Not to mention, narrowing down the sampled groups into low-level sub-groups can help increase the accuracy of predicting LE as attempted by Fries [57] for eliminating premature death to calculate the average life span as premature death rate significantly affects the whole LE rate.

Smartphone users download the app and complete the initial data entry, which include generic and personal criteria. These data are added to cloud servers, which classify and group based on the user’s category. Questionnaires and answers to be collected are comprehensive and can vary from basic profile to technical contents that may not be answered by the user but their physicians as below examples. Alternatively, these answers can be obtained by the health service providers who are connected to a health network, which provides the user’s medical data by having a digital signature from the user; e.g., terms and conditions agreed to on the app screen as some users may not wish to disclose their medical data due to privacy concerns.

Distribution and collection of data can be major and fundamental requirement of the solution, as these data are used to set out the outcome; i.e., PLE results. During the process of solution design, including requirements definition, high-level design of network and application features and products, there should be clear roles and responsibilities of each party and stakeholder, including users, service providers, and regulators.

4.1.1. General Questionnaire

When a user responds to the questionnaire for the first time, they are requested to enter basic user profile and general information, which can be responded to without consulting their physicians, as shown in Table 3.

Table 3. General questionnaires, which can be completed by users. These data can be provided without consulting a user’s health service provider.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal information</td>
<td>Gender, age race, country of birth, location</td>
</tr>
<tr>
<td>Body physique</td>
<td>Weight, height, Body Mass Index (BMI)</td>
</tr>
<tr>
<td>Life events</td>
<td>Education, marital status, retired status, residential type (urban/rural), living conditions (e.g., alone, with partner), pets, driving habits, relationship with family &amp; friends</td>
</tr>
<tr>
<td>Work</td>
<td>Occupation, income, physical/office work, night shifts, full/part-time employment, working hours</td>
</tr>
<tr>
<td>Fitness</td>
<td>Exercise, general health, frequency, and amounts</td>
</tr>
<tr>
<td>Health conditions &amp; diseases</td>
<td>Diabetes, high blood pressure, cholesterol levels, cardiovascular conditions, cancer, allergies</td>
</tr>
<tr>
<td>Diet</td>
<td>Vegan, vegetarian, meat intake, seafood, gluten, and other diet restrictions</td>
</tr>
<tr>
<td>Family history</td>
<td>Age of death of grandparents and parents, known family history of genetic conditions</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>Alcohol consumption, smoking, hobbies, sleeping habits</td>
</tr>
</tbody>
</table>

4.1.2. Technical Questionnaire

Some data are more technical in nature and cannot be entered by the user in response to the smart device app when the user first registers. Complex data such as blood and medical investigations...
(e.g., Electrocardiogram (ECG), Ultrasound, Computed Tomography (CT), Magnetic Resonance Imaging (MRI)) and the diagnosis and test results of certain medical conditions (e.g., heart failure, kidney disease, cancer) would only be able to be provided by the user’s health service provider and would necessitate their consultation [17]. However, users may still be able to answer limited technical questions depending in their level of knowledge. Table 4 includes technical and non-technical information that can be obtained from a centralized server with the consent of the user in addition to the initial survey by the smartphone app.

Table 4. Technical questionnaires, which can be completed by users by consulting their health service provider. This information would be required to fully access the centralized health service network server.

<table>
<thead>
<tr>
<th>Comorbidity count</th>
<th>Mean creatinine (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of visits for primary care, hospitalization in the past</td>
<td>Lowest sodium (mEq/L)</td>
</tr>
<tr>
<td>Mean diastolic blood pressure (mm Hg)</td>
<td>Lowest calcium (mg/dL)</td>
</tr>
<tr>
<td>Mean albumin (g/dL)</td>
<td>Digoxin prescription</td>
</tr>
<tr>
<td>Highest blood urea nitrogen (mg/dL)</td>
<td>Loop diuretic prescription</td>
</tr>
</tbody>
</table>

Additional questions specific for each individual may be included by the health service provider, such as questions relating to a current diagnosis. All these data come from data sources processed and classified by big data.

A high-level approach may also be used for health determination attribute classifications [58]:

- Individual lifestyle behaviors e.g., spending patterns, exercise, diet;
- Physical and social environments e.g., living density, pollution levels;
- Socioeconomic factors e.g., education level, financial status;
- Health outcomes e.g., illnesses;
- Health systems e.g., health insurance status.

4.1.3. Data Source Selection

The method of selecting samples and training data in the cloud is important as it determines the quality of data accuracy. This may require integrating information from multiple data sets. Table 5 shows a primary dataset (sub-regional area information for important health outcomes) complemented by integration with a second dataset (census-level information).

Table 5. Dataset integration. The first dataset has restrictions of geographic granularity due to privacy reasons. The second dataset can be used to complement to improve the quality of data sources (Adapted from [1]).

<table>
<thead>
<tr>
<th>Sub-Regional Area (SRA)-Level Dataset (First Dataset)</th>
<th>Census-Tract Level Dataset (Second Dataset)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health and human service agency (HHSA) Behavioral Health Data (Hospitalizations &amp; Emergency Department visits for behavioral health conditions)</td>
<td>Census-tract level dataset</td>
</tr>
<tr>
<td>HHSA Demographics (Demographics)</td>
<td>American Community Survey 2012 (5-Year Estimates) (Census demographics)</td>
</tr>
<tr>
<td>Environmental Systems Research Institute (ESRI) Market Potential Data (Consumer buying patterns and behaviors)</td>
<td>CalEnviroScreen 2.0 (Pollution data)</td>
</tr>
<tr>
<td>San Diego Association of Governments (SANDAG) Healthy Communities Atlas (Data on physical and built environment)</td>
<td>Loop diuretic prescription</td>
</tr>
</tbody>
</table>
4.1.4. Other Factors

Other factors (such as in Table 6) should also be considered when the predictive model is designed. It is still unknown which attributes are relevant to maximize the accuracy of the PrLE. With a constantly growing database of health information, it is crucial to assess and prioritize relevant and important health attributes (e.g., with machine learning, artificial intelligence) to improve the quality of the PrLE. Figure 5 depicts the overall and summarized questionnaires.

Table 6. Environmental factors that affect PrLE (Adapted from [1]).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average violent crime rate</td>
<td>Composite score: pedestrian traffic safety</td>
</tr>
<tr>
<td>Total park acreage</td>
<td>Percentage of block groups that interest with parks</td>
</tr>
<tr>
<td>Fast food density per square mile</td>
<td>Percentage of sidewalk coverage</td>
</tr>
<tr>
<td>Block groups with minority areas</td>
<td>Count of grocery stores locations</td>
</tr>
</tbody>
</table>

Figure 5. Questionnaires for Life Expectancy calculation. Questions are customized and further developed for specific groups and individuals. For example, a diabetic patient may have detailed questions relating to their diabetic history, types, and dietary intakes.

4.2. Information Flow of Data Creation

When data are transferred to the cloud server, they are compared against the generic data of the user group to create an individual’s data. These data are fed back to the generic data for training purposes and inference at sensors.

There are two pieces of information to be provided: (1) generic life expectancy graph of the group where the user belongs to; and (2) individualized result of the user. A monitoring center server in the cloud collects physiological data from individual smart devices such as smartphones, which collect and process sensed data in WBAN. These data are classified and analyzed for data accuracy prior to inferring, optimizing and transmitting to smartphones to be used by the app. Information flow of data transmission is shown in Figure 6. Steps 1 & 2 refer to initial data collection and entry followed by Steps 3 & 4 for data inference and optimization. Step 5 is for presenting PLE outcome, and Step 6 refers to data updates for fine-tuning with ongoing data feedback to the cloud server.
Table 6. Environmental factors that affect PrLE (Adapted from [1]).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average violent crime rate</td>
<td>Composite score: pedestrian traffic safety</td>
</tr>
<tr>
<td>Total park acreage</td>
<td>Percentage of block groups that interest with parks</td>
</tr>
<tr>
<td>Fast food density per square mile</td>
<td>Percentage of sidewalk coverage</td>
</tr>
<tr>
<td>Block groups with minority areas</td>
<td>Count of grocery stores locations</td>
</tr>
</tbody>
</table>

Figure 5. Questionnaires for Life Expectancy calculation. Questions are customized and further developed for specific groups and individuals. For example, a diabetic patient may have detailed questions relating to their diabetic history, types, and dietary intakes.

4.2. Information Flow of Data Creation

When data are transferred to the cloud server, they are compared against the generic data of the user group to create an individual’s data. These data are fed back to the generic data for training purposes and inference at sensors.

There are two pieces of information to be provided: (1) generic life expectancy graph of the group where the user belongs to; and (2) individualized result of the user. A monitoring center server in the cloud collects physiological data from individual smart devices such as smartphones, which collect and process sensed data in WBAN. These data are classified and analyzed for data accuracy prior to inferring, optimizing and transmitting to smartphones to be used by the app. Information flow of data transmission is shown in Figure 6. Steps 1 & 2 refer to initial data collection and entry followed by Steps 3 & 4 for data inference and optimization. Step 5 is for presenting PLE outcome, and Step 6 refers to data updates for fine-tuning with ongoing data feedback to the cloud server.

Possible scenarios of health data collection and transfer to cloud servers are described below:

1. A user installs a PLE app on their smartphone and adds personal information and health data as requested by the app. The user also gives a consent to the service provider so that they can access and retrieve their health data from the user’s health record located in a central server.
2. The app communicates with the central server which initiates user registration and basic provisioning processes. It also pairs up (if needed) with the user’s body sensors to be ready for data transfer.
3. The user’s body sensors collect and transfer the sensed data to the smartphone and follow instructions of when and how to collect and transmit data.
4. When data have been collected, the smartphone app transmits them to the central server that will prepare and generate the initial results, and send them back to the app for optimization. The app may collect further data from the sensors as programmed prior to displaying the outcome of PLE for the user.
5. The central server continuously updates and improves data quality to input to the generic group data as classified and grouped for the user. Each element fine tunes the data by inference algorithms before transmission, and discern how to handle the priority; e.g., QoS.

4.3. App Design

Health data are transferred from sensors to a smart device via low-powered wireless protocols such as Bluetooth low energy (BT-LE), WiFi (802.11), and Zigbee. The app displays the received data and transfers them to a server in the cloud for data processing. The screen menu can be combined as a part of existing mHealth mobile applications (as shown in Figure 7). There are two phases of personal data collection: general and technical health information. As the current health and fitness related apps integrate multiple functions, PLE features can be implemented within the existing apps as current mHealth services, which already require handling of physiological data.
5. Applications

Along with existing health applications such as fitness tracking, chronic disease monitoring and real-time patient monitoring, the PLE application can be useful for users to improve their lifestyle and exercise by planning goals on a short and long-term basis. For example, the current PLE outcome of 85 years will be adjusted when the user changes their attributes such as smoking cessation, reducing alcohol consumption, commencing regular exercise, or modifying dietary plans.

The development of wearable devices is evolving rapidly to capture data and for use in applications [59]. By wearing fitness tracking or monitoring devices, those attributes can be automatically updated and sent to the cloud servers, which process and update the health index defined by the user’s physician to assess the user’s health life, which eventually leads to updating the user’s PLE. IEEE P11073 Personal Health Devices (PHD) or sensors attached on or implanted in the user’s body also provide physiological data to the mHealth monitoring center, which is connected to the centralized cloud servers for data optimization with increased data accuracy as shown in Figure 8.

Figure 7. mHealth mobile app screens (example). (a) Existing health mobile app can integrate PLE feature as additional services (b) as both functions require the same physiological data to transfer to a monitoring center for PLE calculation.

Figure 8. Mobile health (mHealth) network, which consists of WBAN, Health service providers, Caregiver network and central cloud servers.
One of the potential stakeholders of this solution may be health insurance companies, who can provide personalized insurance products by monitoring their customers’ health index. They may provide customers with incentives to improve their health condition and lifestyle, for example by revising their premiums in response to improvements in their customers’ health index obtained and processed by the cloud computing program.

A strict policy and regulations should be imposed for third parties such as health insurance companies regarding access to user data such as requiring patient consent and agreement, as well as parameters on how that data can be used. It is essential to develop and agree upon transparent security and privacy policies prior to engaging third parties for the usage of health data of users across states and governments.

6. Conclusions and Future Works

Few works have been done to provide an individually customized life expectancy prediction. We have reviewed existing works and techniques in the prediction of human LE, and reached a conclusion that it is feasible to predict a PLE for individuals using evolving technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearables and mobile health monitoring devices. We also identified that the collection of data will be a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The interworking of a heterogeneous health network is also a challenge for data collection. Despite these challenges, we showed a possibility of a PLE prediction by proposing an approach of data collection and application by smartphone, with which users can enter their information to access the cloud server to obtain their own PLE. No attempt has been made to create this novel idea of using smartphone integrating cloud servers for real-time data entry. We investigated obstacles and barriers that can be resolved by future works described below. Previous works have described a five year LE prediction, however it is not oriented as a personalized prediction but rather utilizes a median model-predicted probability of 5-year survival of patients who are either sick or healthy. It is proposed that this can be extended to a lifetime prediction by using big data to generate a generic data, which can be used to create a PLE based on training data as a future solution. Building a generic database will take a considerable amount of time for data collection and analysis, taken from birth to death for various demographic groups to be useful and accurate in representing each attribute classifications. Whilst current applications attempt to show PLE for smartphone users, they are complicated and difficult to collect technical data requested by the questionnaire, as users are unlikely to be able to provide these data themselves. This can be resolved by connecting the app to the central cloud server with the mHealth networks which provide other health related applications.

A centralized cloud server plays a key role in collecting, processing, and creating meaningful value using big data, which forms the input of the solution as well as creates generic data against each user’s PLE requirements. Service providers shall envisage challenges and hurdles to obtain consent of personal health of heterogeneous health networks across developed countries. This will lead to (3) classification of data based on processing big data and each group’s traits, which can be used as personalized threshold ranges; (4) When this has been completed in a cloud, it can be connected to a smart device app that can provide questionnaires developed by health specialists and collect answers to customize the user’s PLE; (5) Optimization of the generic groups’ data is done by developing an algorithm using machine learning for continuously building and optimizing the user’s generic data.

As the proposed solution requires processing and transmitting health information of users, information security is a key aspect to consider such as privacy as well as ethical requirements recommended by regulation bodies, such as the Australian national health and medical research (NHMRC). The scope of security and ethical requirements need to be clearly defined and specified for future work as challenges are expected to build a centralized database with incorporation of health networks. For example, North America, Asia, and Europe may have their own unique requirements to satisfy in dealing with health data with different health research guidelines.
To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.

**Author Contributions:** J.J.K. conceived, reviewed related works and designed the solution and application, and wrote the paper. S.A. supervised the overall quality of the article and provided comprehensive guidance of the contents.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Abbreviations**

Life Expectancy (LE)—longevity of human life based on diseases or other variables such as geographical region; race or regional traits; Personalized Life Expectation (PLE)—LE focused on personal or individual longevity rather than groups or classified based on traits such as diseases; races and regional areas; Predicted Life Expectancy (PrLE)—LE that has been predicted for the future from past health related data processed by cloud server. Mobile Health (mHealth)—a branch of electronic health (eHealth) being provided by wireless body area networks; in which sensors collect and transmit physiological or biological data to a monitoring center in the cloud for health service providers to access.

**References**


34. Hannerez, H.; Borgà, P.; Borritz, M. Life expectancies for individuals with psychiatric diagnoses. *Public Health* 2001, 115, 328–337. [CrossRef]
41. Soomro, A.; Cavalcanti, D. Opportunities and challenges in using WPAN and WLAN technologies in medical environments [Accepted from Open Call]. *IEEE Commun. Mag.* 2007, 45, 114–122. [CrossRef]


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).