

AI, Big Data powered method of life expectancy prediction, severe diseases early stages detection, prevention

Dushkin Roman Valerievich

Lelekova Vasilisa Alekseevna

Maksimov Vladislav Sergeevich (✉ vladislav@intime.digital)

Teterin Oleg Olegovich

Research Article

Keywords: artificial intelligence, biomarkers, neural networks, aging, human aging, health monitoring, 4P Medicine, machine learning, biological age, definition of serious diseases

Posted Date: August 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-820273/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Aging is a part of human life, often accompanied by serious illnesses. Nowadays, people sometimes do not live up to the biological aging of the body at all due to late-timed diagnosis of diseases. Unfortunately, the methods of early detection of diseases associated with aging do not yet have the technical equipment that would allow them to be fully implemented. This article provides an overview of methods for defining and analyzing the aging of the body. This is a review article of a novel hardware and software complex for health monitoring developed by a scientific group, which analyzes human bio parameters using artificial intelligence algorithms. The relevance of the proposed system is undeniable due to the used algorithms of artificial intelligence, with the help of which it is possible to quickly and accurately analyze a large amount of data related to human aging. The article will be of interest to developers of artificial intelligence, biostatisticians and scientists working on the definition of aging in the human body.

Introduction

Aging is natural for every living organism, while it is often accompanied by diseases that affect both the body and the human psyche. From a biological point of view, aging is the result of the influence of metabolic errors and external stress factors on the individual development of the body [Moskalev A., 2019]. For different people of the same age, the rates of aging can differ significantly, as well as the rates of aging of systems and organs within one organism - the degradation of one system causes changes in many others [Fedintsev A., 2017]. Aging is not a simple factor of time. Some animals age quickly; others age over a very long time. This is exciting because it is not simply an accumulation of time and toxic factors. Instead, there is some kind of physiologic clock.

Many scientists are working on the goal of extending human life and determining the biochemical processes that occur during aging. For each organism, aging occurs individually, therefore, an unambiguous way to track the process itself has not yet been found. However, there are many different ways to determine biological age, different from the calendar age. The data that identifies aging in the body are called biological markers of aging, or simply biomarkers of aging. Biomarkers can be found throughout the body: in the microflora of organs, in cells, and especially in DNA and RNA chains.

Due to the global change in lifestyle in recent decades, premature deaths are recorded much more often. According to statistics, many people die much earlier than the physical death of their organs. The main reason is serious diseases that were not diagnosed and cured in time [Kupryushin A.S., 2016]. Usually, doctors conduct a comprehensive examination only if the standard tests do not correspond to normal values, while most latent diseases are not tracked using the type tests available today [Vasilkov V.G., 2017]. Artificial intelligence methods can help make the current situation better. As an example, the Watson for Clinical Trial Matching from IBM Watson Health can collect and link structured and unstructured data from electronic medical records, medical literature, trial information and eligibility criteria from public databases such as ClinicalTrials.gov [Aggarwal M., 2017]. Mayo Clinic's early use of

Watson for CTM resulted in an 80 percent increase in the number of participants in clinical trials of systemic therapy for breast cancer within 11 months of implementation [Helgeson J., 2018].

Biological age and methods for determining it, including using artificial intelligence (AI), machine learning and artificial neural networks - these are the opportunities to find additional biomarkers of aging (natural and digital) - will be discussed in the first part of this article.

The methods used everywhere often do not help in detecting the disease in the early stages. In addition, due to cost constraints, fewer doctors are using modern equipment, such as mass spectrometry instruments, to analyze biomaterials. It's not just the cost of the equipment. There is also a lack of expertise to evaluate the results and the lack of phenotypes for corresponding data, i.e. not enough people have been tested for a particular condition to have meaningful use. Considering all of the above, it is safe to say that there is a need for health monitoring systems in which data can be analyzed using artificial intelligence algorithms. The second part of this article tells about the developed set of solutions.

1. Methods And Data

1.1. Biomarkers and biological aging

Biomarkers of aging are molecular, cellular and physiological parameters of the body that predictably change with age - qualitatively or quantitatively [Moskalev A., 2019]. In the scientific community, such a unit of measurement for biological age is considered the most relevant and promising for early diagnosis, prediction and monitoring of chronic, including age-related, diseases. Many pathologies develop in a latent form for a long time - it is extremely difficult to detect them at an early stage, since their manifestations are nonspecific for the general clinical case. Biomarkers make it possible to determine the deviation from the norm associated with the risk of developing a specific disease, and to prescribe effective prevention at an early stage [Pyrkov T.V., 2019].

According to Butler et al. [Butler R.N., 2004], the biomarker should:

- change with age;
- predicting death is better than calendar age;
- to determine the early stages of a specific pathology, in particular - an age-related disease;
- be minimally invasive - does not require major surgery or painful procedure.

Later, Moskalev [Moskalev A., 2019] added the list:

- have a high sensitivity to early signs of aging of the body;
- be predictable over the foreseeable time frame;
- have low analytical variability - be reliable and reproducible.

To date, the most complete database of biomarkers of human aging is the Digital Aging Atlas [Craig T., 2015].

1.2. Practical determination of biomarkers and biological age

The problem of biological aging was presented by Veytsman et al. [Veytsman B., 2019] as a trajectory in a multidimensional space, each dimension of which represents the characteristics of the functioning or work of the organism. These trajectories are different for different people and, moreover, each person follows them at a certain speed, which can increase or decrease depending on specific circumstances. In fact, each trajectory can be displayed in a set of natural coordinates, in this case, biomarkers.

Consequently, it becomes possible to compare points on a specific set of trajectories, that is, to compare individual people and come to an understanding of how a person's trajectory is built and why the speed of a person's movement along its own trajectory changes.

The main problem in determining such a "trajectory" is the duality of approaches and the need for serious technical equipment. The various approaches are divided into "theoretical" and "practical". The first uses only confirmed data, therefore it is limited, since the current scientific knowledge about the human body is still far from complete. The second takes into account all the data, but requires voluminous calculations, which is why you have to resort to machine learning methods to analyze a large amount of information and filter out useless samples. Consequently, to build correct correlations for the selected biomarkers, technical equipment is required corresponding to the amount of data and the complexity of calculations. A variant of the hardware and software complex developed for the analysis of such data is presented in the third part of this article.

1.3. Promising biomarkers

Chromatin. The basis of chromosomes, it is in the composition of chromatin that the implementation of genetic information, replication and DNA repair occurs. Complex connection of DNA, proteins and RNA; the smallest component is the nucleosome. There are many modifications, in particular, heterochromatin, which is characterized by a condensed (compact) state and a low ability to synthesize RNA.

Effects on aging and disease:

- It has been proven that a decrease in heterochromatin in the body leads to a number of chronic age-related diseases (for example, Werner's syndrome [Shumaker D.K., 2006; Zhang W., 2015]).
- Assumptions about the relationship between aging and a decrease in the amount of heterochromatin in the human body are based on two factors. The first of them is the relationship between mitochondrial stress (the result of incorrect conformational folding of proteins, leading to a change in the work of various genes), in which the concentration of heterochromatin in the blood decreases, and aging [Tian Y., 2016]. The second is the trend of scientific research [Scaffidi P., 2006; Larson K., 2012; Wood J.G., 2010].

- Another confirmation is associated with the accumulation of nucleosomes (a product of chromatin degradation, its smallest constituent part) during aging in mammals [O'Sullivan R.J., 2010].

Histone modifications. Histones are proteins involved in the packaging of DNA strands and the regulation of processes such as the formation of RNA, the creation of daughter DNA, and the repair of damage in DNA molecules. Histone modifications are chemical changes in the original proteins that affect the formation of RNA.

Impact on disease and aging:

- Modifications of histones obtained with the help of nucleosomes accumulate during aging [Contrepois K., 2017; Piazzesi A., 2016].
- The variability of histone modifications changes during aging [Benayoun A., 2015].
- Methylated and acetylated histones are used as the main markers of aging. The amount of methylated histones decreases along with heterochromatin in the same Werner syndrome [Zhang W., 2015], while the amount of acetylated ones correlates with age-related deterioration in cognitive functions [Peleg S., 2010].

DNA methylation. The methylation reaction occurs at CpG dinucleotides - DNA regions that play a key role in the activation of innate immunity in vertebrates [Branda R.F., 1993]. CpGs are usually unevenly distributed along the DNA chain, and, nevertheless, there are sequences of successive CpG regions - CpG islands [Schubeler D., 2015]. Most of the individual CpGs are methylated, with the exception of the CpG islands [Neri F., 2017; Yang X., 2014].

Impact on disease and aging:

- Methylation of CpG islands is a feature of some cancer cells [Deaton A.M., 2011].
- It is possible to track age-related changes in mammals using CpG islands [Fraga M.F., 2007].
- Practical advances in the development of a "biological clock" based on the DNA methylation reaction. Some groups of scientists were able to measure the chronological age of a person [Horvath S., 2018; Nevalainen T., 2017; Wagner W., 2017], and the group of Levine et al. [Levine M.E., 2018] was able to develop a clock that predicts biological age using multiple linear regression algorithms and machine learning capable of analyzing and predicting life expectancy.
- It was confirmed that the data obtained from the analysis of methylated CpG groups are accurate up to the 0.95 level of the correlation coefficient between the predicted age and the present [Horvath S., 2018].

Micro-RNA. A class of non-coding (not translated into protein) RNAs that cause tissue degradation [Bartel D.P., 2004].

Impact on disease and aging:

- The following scientific groups have shown the effect on aging: muscle cells [Singh J., 2016; Drummond M.J., 2011], germ cells [Battaglia R., 2016], human blood [Olivieri F., 2017] and blood serum [Smith-Vikos T., 2016].
- Shown the existence and proved the influence of two micro-RNAs (let-7 and mir-34) on DNA damage [Kato M., 2009; Kato M., 2011], neurodegradation [Lehmann S.M., 2012], aging [Tazawa H., 2007] and cell death [He L., 2007].
- With the help of the above micro-RNAs, it is possible to measure life expectancy [Ewald C.Y., 2016; Fitzenberger E., 2014].
- The process of micro-RNA influence on the aging of the organism through the central conductor of aging signals [Tatar M., 2003] - the insulin signaling pathway [Inukai S., 2012] has been shown.
- Micro-RNAs are extremely promising as biomarkers due to: mobility [Hoy A.M., 2012; Arroyo J.D. 2011; Hunter M. P. 2008; Vickers K.C., 2011], stability, quantification, protection against degradation [Huang X., 2013] and ease of detection [Machida T., 2015; Weber J.A., 2010].
- The most accurate method for non-invasive RNA isolation is from blood serum [Cheng L., 2014].

Microflora. A set of microorganisms in symbiosis with humans.

Impact on disease and aging:

- Critically differs in different age categories, but changes gradually over time [Mangiola F., 2018], which is an important factor for a biomarker.
- Not relevant for data analysis of older people due to too much variability [Claesson M.J., 2012].

Visual image.

Impact on disease and aging:

- Used as a quantitative assessment of phenotypes and definition of biomarkers in databases [Zhao Q., 2016].
- Groups of scientists [Gunn D.A., 2008] proved that the person's age perceived from the image is not only considered a biomarker, but is also associated with variations in human genes [Liu F., 2016].
- Pigmentation of the face that gradually manifests itself with age, which is strongly associated with certain age-related diseases, for example, with atherosclerosis of the carotid arteries, is also considered a biomarker [Miyawaki S., 2016].
- Using MRI and skull scans, it has been proven that the density of the skull tissue [Colcombe S.J., 2003] and the volume of the brain [Driscoll I., 2009] decrease with aging, which is consistent with the decline in human cognitive functions with age.

1.4. Comparison of modern and clinical methods

Biological studies involving large numbers of samples are costly and logistically complex. The necessary process optimization consists in the ability to obtain a sufficiently complete database to describe the aging process, not only qualitatively, but also quantitatively. Another problem lies in the strong difference between biological and clinical markers for assessing human health [Mitnitski A., 2019], however, it can also be solved.

Integration of a large number of biomarkers into clinical trials is possible using the fragility index - the simplest and most reliable method available. The fragility index was presented in 2001 [Mitnitski A.B., 2001] as a means of quantifying the general health of the elderly, and then extended to adults [Rockwood K., 2011]. The fragility index is defined as the ratio of the number of human health disorders accumulated by a person to the total number of violations in a database or study [Mitnitski A., 2015]. According to earlier estimates obtained using cross-sectional demographic analysis, violations accumulate exponentially, at a rate of 3.5% per year [Mitnitski A., 2005]. Also, Minitsky et al. Showed that the fragility index works with cellular biomarkers of inflammation, cellular aging and genetic markers [Zhavoronkov A., 2013].

However, for such a simple method, a contradiction arose - the fragility index takes into account any health deficiency, but does not take into account its characteristics, for example, diseases, disabilities and symptoms. This is convenient for practical use, but contradicts the primary principles of clinical teaching, in which the accuracy of the diagnosis of the disease is primarily important, since what helps in condition A can be harmful in a similar condition B. The simplest example of this is diabetes - the administration of insulin will help a patient with an excessive amount of sugar in the blood, but it will be harmful for a patient with a lack of sugar in the blood.

Nevertheless, as a quantitative assessment of aging, the fragility index fully justifies itself, after all, it does not matter which disorders have accumulated, it is important how they affected the body. So, for example, how could a skin problem and a heart attack be quantified? Not every heart attack is fatal, and not every rash can be benign. To the extent that such disturbances affect the body, they are added to the number of disturbances. The fragility index reflects the actual damage to the human body, without considering what kind of disease caused the damage, and, therefore, as a quantitative characteristic, it is extremely practical.

1.5. Artificial Intelligence Methods for Studying Aging

Assessment of aging is the first step towards taking measures to reduce the morbidity, social and economic burden associated with aging [Belsky D.W., 2017]. One of the main hypotheses is the assumption that chronic diseases associated with age have common components of the genetic architecture and, therefore, are associated with aging and the assessment of health risks in general.

However, for longevity medicine to be officially considered a branch of medicine, it must be practiced by doctors. Clinical practice requires a lot - clinical protocols and guidelines for diagnosis and treatment with defined outcome indicators, official biomarkers and drugs approved by regulatory authorities. To develop

at least preliminary clinical recommendations, it is necessary to monitor aging and consider it as a medical condition, conducting special studies to verify the effectiveness and safety of specific interventions in the process.

First of all, clinical trials are designed to predict the long-term consequences of the use of pharmaceutical or medical drugs [Abbas I., 2016]. The complexity of such a process lies not so much in monetary costs as in time. Patients die before the end of the study period, while the researchers themselves also age [Loseva P., 2020]. Nevertheless, scientists are working on this problem using artificial intelligence, because if you build a general research model and carry out the remaining calculations with its help, then clinical trials will take weeks, not years [Holford N.H.G., 2000]. With the rapidly growing volume of medical data available to researchers, including data from electronic medical records and wearable devices, sophisticated machine learning algorithms can save billions of dollars, accelerate the development of medicine and expand access to experimental therapies [Woo M., 2019].

The most accurate methods for calculating biological age are the subject of constant debate. Recent studies show that a set of biomarkers, and not any single biomarker, is the most effective means of assessing a patient's health status [Liu Z., 2018].

For analyzing such a large amount of data, AI and machine learning methods are best suited. Currently, there is a growing number of AI-based instruments that, having access to appropriate health parameters, use various aspects of the patient's health status and aging rate to make a prognosis [Zhavoronkov A., 2021].

Popular biological age models are trained to predict chronological age, but often fail to fully reflect the signs of a possible disease. This disadvantage can be eliminated using log-linear risk models that allow linear regression to be applied to the task and using clinical data [Mamoshina P., 2018].

For example, Mamoshina et al. [Mamoshina P., 2016] used deep neural networks (DNN - Deep Neural Networks) to build the aging clock. DNN architecture is promising at present due to their ability to identify hidden patterns in datasets and study multidimensional data in an atypical representation [Putin E., 2016]. This group of scientists considered age prediction as a regression problem, that is, the resulting model takes a vector of indicator values from the blood test and returns one value of the patient's age. To address the problem of DNN interpretability and gain a deeper understanding of the data, the researchers used permutation feature importance (PFI) analysis to rank input blood markers according to their importance in predicting age.

In another work [Bottou L., 2012], age prediction was also considered as a regression problem and the standard coefficient of determination and the ϵ -prediction accuracy were used to assess the effectiveness of the method. The single DNA assessment method has shown itself to be promising for further research due to its high accuracy.

In this work, the algorithms described below were used to train machine learning models (including those used in the development of the team of authors of neural networks), although the development of a system for assessing aging indicators is not limited to them:

- Stochastic Gradient Descent Optimizer is a function optimization method with suitable properties. In fact, this is a gradient descent optimization, since, in the case of a stochastic optimizer, not the entire data set is used, but a randomly selected set of a subset of the data, which greatly reduces the computational load, providing faster iterations in exchange for a lower convergence rate [Kingma D.P., 2015].
- Adaptive optimization of the torque estimate - a method that calculates the adaptive learning rates of the neural network for different parameters [Breiman L., 2001]. It is simple to implement, efficient and suitable for tasks with fluctuating parameters (in the described case, biomarkers).
- Optimization of RMS propagation.
- Linear Regression – A simple estimate of the coefficient values used in the analysis using the available data. Useful for large amounts of data.
- Logistic regression - estimation of the probability of a certain value with the available parameters. Useful for large amounts of data.
- Monte Carlo Methods - A subset of computational algorithms that use a multiple random sampling process to numerically estimate unknown parameters. They allow simulating complex situations in which many random variables are involved, and therefore assessing possible risks.
- Markov models: Markov chain, hidden Markov model, Markov decision-making process, partially observable Markov decision-making process.
- Transformer is a deep learning model that uses the attention mechanism by differentially weighing the significance of each piece of input data. Parallelizing training in a transformer model makes it possible to conduct training on large datasets.
- Support vector machine (SVM) is a linear algorithm used in classification and regression problems. The task of the algorithm is to find a hyperplane that separates the received data into two classes. The main advantage is the ability to work with a large amount of data.
- Linear SVC - optimization of the core of the SVM algorithm, which allows to reduce the number of calculations.
- The k-Nearest Neighbors algorithm is a simple algorithm that classifies all data or observations based on similarity to each other. It is useful in that it is easily amenable to parallel implementation and uses only local information, and therefore is adaptive.
- Naïve Bayes Algorithm - A simplification of the Bayesian classifier, an algorithm that calculates probabilities, making the assumption that values are conditionally independent from each other. This makes the Naive Bayes algorithm extremely fast.
- Simple perceptron - a supervised learning algorithm for binary and multiclass classifiers. It is useful for its adaptability in working with class separation.

- Decision tree classifier is a supervised machine learning algorithm that is used to solve classification problems. In fact, it is a branching chain of "Yes" or "No" questions. Such an algorithm does not require global data preparation, and the memory consumption for the implementation of the output is logarithmically dependent on the amount of data used for training, so a large amount of data does not greatly affect the output speed.
- Random Forests - an algorithm that uses an ensemble of decision trees [Hastie T., 2009]. Fast, easy to implement, adaptive, robust to non-linear functions, and capable of handling unstructured data.
- Gradient Boosting Classifier is a machine learning algorithm for classification problems that builds a prediction model in the form of an ensemble of weak predictive models, usually decision trees [Ke H., 2017]. Intuitive and predictive efficient algorithm.
- Ridge Classifier - a classification algorithm that allows you to work not only with linear differences between classes. With its help, it becomes possible to avoid overfitting data for class values.
- Bagging Classifier - an algorithm that evaluates an ensemble of solutions and fits the base classifying values (each on a random subset), and then combines them, and the individual predictions are used to form the final output. Greatly reduces the variance of the output.

2. Results And Discussion

The use of artificial intelligence algorithms to predict the duration and quality of human life is at the very beginning of its development. The objective of this work was to create a hardware and software complex for health monitoring, which can analyze human bio parameters using artificial intelligence algorithms.

Figure 1 shows a block diagram of the forecasting system being developed for assessing the duration and quality of life, and in table. 1 provides descriptions of specific modules. The forecasting system is configured so that the assessment of a person's life and general health is based on as many factors as possible that affect a person. AI algorithms and, in particular, machine learning, make a significant contribution to:

- monitoring by the system of many parameters of human health in real time, including the dynamic change of parameters;
- monitoring the state of hardware of the Forecasting System;
- the work of a recommendation module that suggests steps to improve a person's condition based on similar incidents from the database and prevent possible negative scenarios even at the preclinical stage;
- Correlation of anomalies in the database to create a subclass for unique situations, their causes and consequences.

Nevertheless, many technical and organizational problems need to be solved to organize the process of data collection and training the system. Among the technical ones is the task of forming a training sample and the target value of life expectancy (human health factor). The prediction system also has a

need to formulate a criterion for assessing the health of existing patient sets. The health factor used in this work is a synthetic measure that depends on the number of systemic diseases, current test values and functional characteristics of the organism.

Input data Forecasting systems are a set of parameters obtained from medical records, portable devices, questionnaires and other sources, including the upload of laboratory analyzes of biomaterials to the server of the System itself. Output data Forecasting systems are presented in the form of a report on the state of the body, possible diseases and the "health factor" of a person, which is directly related to life expectancy. The higher this factor, the longer and better a person's life is predicted.

Table 1
Description of the modules of the patented method of the forecasting system

Module name	Module Description
Intelligent wearable device	Contains at least one sensor for bio parameter registration, for example Garmin (implementation on an aggregation platform such as HealthKit, GoogleFit or other mobile applications for bio parameter registration is possible).
User interface	Allows the user to view data in a variety of ways such as graphs, charts, and tables.
Big data server	Data collected from a medical database and approved by the World Health Organization and / or the World Medical Association and / or other similar associations. The list of disease-specific parameters on a big data server may change based on new decisions by the World Health Organization and other associations.
Communication network	Communication with a big data server.
Database	Stores many parameters of the health of a large number of people.
Medical Database	When combined with a regular database, it can be a relational database that recognizes the relationships between the items in both databases.
Monitoring module	Configured to take into account and redistribute many health parameters.
Assessment module A	An artificial intelligence model trained on data from a database.
Artificial Intelligence Module A	Structures personal datasets of a plurality of individuals and generates a digital profile based on matches obtained from at least one user dataset and a plurality of general population datasets obtained from databases.
Artificial intelligence module B	<ol style="list-style-type: none"> 1. generates a list of necessary parameters for continuous monitoring of possible diseases; 2. develops personalized algorithms for each user health report; 3. with the available information about a specific user for a long period of time, uses the collected data to train AI models in order to predict the risk of a particular user getting sick and create personal preventive calculations without using common datasets of the population.
Assessment module B	Evaluates the training dataset to draw conclusions from a large dataset and determine at least one characteristic from the training dataset.
First stage evaluation submodule	Instant evaluation of data about a person obtained at a certain point in time.
Second stage evaluation submodule	A machine learning model for processing historical data that analyzes the history of human diseases and learns from it.

Module name	Module Description
Third stage evaluation submodule	<p>The tasks of this module are:</p> <ol style="list-style-type: none"> 1. Recognition of a large number of similar parameters in a set of input parameters to assess the individual relationship between a person's quality of life and his corresponding quality of health. 2. Formation of a basic personalized machine learning model, trained on a large sample of data and tuned to a specific person over a long period of time.
Second stage module	Develops a machine learning model. This trained model analyzes all the known medical data of a particular person.
Generation module	Provides an output that is a factor in assessing human health that is directly related to life expectancy.
Recommender module	Provides personal advice on the prevention of serious diseases at different stages of their development.
Report module	Creation of a report on the risk of a possible disease.

In Fig. 2 shows a block diagram of the hardware of the Forecasting System, which includes:

- a processor for performing actions in accordance with instructions, for loading operating instructions from main memory (or data storage) into cache memory, for loading instructions from cache memory into on-board registers;
- block of cache memory for storing instructions;
- coprocessor to support the interaction of the processor and main memory;
- bus - an interface providing data exchange between the internal components of the Forecasting System;
- a block of main memory containing computer-executable code and including a processing circuit, which is configured on the basis of the previously described circuit in Fig. one;
- a network interface controller for managing one or more network interfaces and for connecting to network devices (for example, to access the network);
- input-output interface to facilitate sending and receiving data to various input-output devices (mobile phones, printers, etc.);
- data warehouse.

Prediction and assessment of the quality and duration of human life occur in the steps shown in the block diagram shown in Fig. 3 in accordance with the previously described firmware.

The first database includes various bio-data stored and collected in dynamics individually for each person. In the database, at the end of the first step, there are:

- general population data on which scientific articles are based;

- data collected by the presented system, which are distributed according to the similarity of parameters (for example, only men, 45 years old, living in a metropolis, having had a stroke in the past, smokers, etc.)
- data accumulated over 3–5 years for each user. It is assumed that it is this option that will make the most accurate forecast of the trajectory of a person's life.

In a second step, health bio parameters are monitored and evaluated by a machine learning model trained on a database and a medical database. After the bio parameters are converted into a set of personal data, the AI module generates digital user profiles, which are considered a training sample for a machine learning model.

At the third step, the training sample is evaluated in order to draw conclusions based on it and determine at least one characteristic to describe the relationship between the quality and duration of a person's life. In the presence of a person's medical history, the model is trained on a larger amount of data, and, therefore, in the process of work, it recognizes more patterns between the quality of life and the level of health of an individual.

Then, at the fourth step, a trained machine learning model is formed, which analyzes many human bio parameters for different periods of life, affecting both positively and negatively on the prognosis of life expectancy.

The last step is to generate the output. In particular, the coefficient of human health, which is directly related to life expectancy. For example, the health coefficient indicates a longer life expectancy, when the risk of serious diseases is estimated and low. Personalized recommendations for the prevention of serious illness are provided personally to the person through a generated report using an AI module configured to extract the necessary data from an unstructured set.

3. Conclusion

The method described in this article is a 4P Medicine - Prevention, Prediction, Participatory, Personalized system. The system tracks twenty of the most serious diseases at the earliest known stages and is able to increase the duration and quality of human life. An early analysis of the bio parameters of an individual, while the symptoms are not even visible on traditional analyzes (biochemistry, ultrasound, MRI, etc.), but the development of the disease is already in progress, allows you to significantly save on expensive treatment when a serious illness is already running. Moreover, the described development is equipped with the proposed hardware and software complex and monitors more than fifty parameters from wearable electronics devices in real time, as well as the dynamics of a person's state for more than three hundred and fifty parameters of blood and other biomaterials with regular measurement (over two per year), providing the user with timely advisory health reports. As a result, the described development is of great help in the search for the "butterfly effect" of the life trajectory of every person, when it is still

possible to make the necessary changes in people's lives, giving them the opportunity to live more active and happy years without restrictions in the form of serious diseases.

Declarations

Competing interests: The authors declare no competing interests.

References

- Abbas I. (2016) Modeling and simulation in clinical trials. Proceedings of the Modeling and Simulation in Medicine Symposium (MSM '16). Society for Computer Simulation International, San Diego, CA, USA, 1: 1–6.
- Aggarwal M., Madhukar M. (2017) IBM's Watson Analytics for Health Care: A Miracle Made True. In Bhatt C., Peddoju S. (Ed.) Cloud Computing Systems and Applications in Healthcare (pp. 117–134). IGI Global.
- Arroyo J.D. et al. (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. USA* 108 (12): 5003–5008.
- Bartel D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 116 (2): 281–297.
- Battaglia R. et al. (2016) MicroRNAs are stored in human MII oocyte and their expression profile changes in reproductive aging. *Biol. Reprod.* 95 (6): 131.
- Belsky D. W., Moffitt T. E., Cohen A. A., et al. (2017) Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: Do they measure the same thing? *Am. J. Epidemiol.* 2017; kwx346.
- Benayoun B.A., Pollina E.A., Brunet A. (2015) Epigenetic regulation of aging: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell. Biol.* 16 (10): 593–610.
- Bottou L., Bousquet O. (2012) The Tradeoffs of Large Scale Learning. Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. (eds.). *Optimization for Machine Learning*. Cambridge: MIT Press. pp. 351–368.
- Branda R. F., Moore A. L., Mathews L., McCormack J. J., Zon G. (1993) Immune stimulation by an antisense oligomer complementary to the rev gene of HIV-1. *Biochem. Pharmacol.* 45 (10): 2037–2043.
- Breiman L. (2001) Random Forests. *Machine Learning* 45: 5–32.
- Butler R.N. et al. (2004) Biomarkers of aging: from primitive organisms to humans. *J. Gerontol. A Biol. Sci. Med. Sci.* 59: B560.
- Cheng L. et al. (2014) Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. *J. Extracell. Vesicles* 26: 3.

- Claesson M.J. et al. (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488 (7410): 178–184.
- Colcombe S.J. et al. (2003) Aerobic fitness reduces brain tissue loss in aging humans. *J Gerontol. A Biol. Sci. Med. Sci.* 58 (2): 176–180.
- Contrepois K. et al. (2017) Histone variant H2A.J accumulates in senescent cells and promotes inflammatory gene expression. *Nat. Commun.* 8: 14995.
- Craig T. et al. (2015) The Digital Aging Atlas: integrating the diversity of age-related changes into a unified resource. *Nucleic. Acids. Res.* 43: D873.
- Deaton A.M., Bird A. (2011) CpG islands and the regulation of transcription. *Genes Dev.* 25 (10): 1010–1022.
- Driscoll I. et al. (2009) Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology* 72 (22): 1906–1913.
- Drummond M.J. et al. (2011) Aging and microRNA expression in human skeletal muscle: a microarray and bioinformatics analysis. *Physiol. Genomics.* 43 (10): 595–603.
- Ewald C.Y., Marfil V., Li C. (2016) Alzheimer-related protein APL-1 modulates lifespan through heterochronic gene regulation in *Caenorhabditis elegans*. *Aging Cell* 15 (6): 1051–1062.
- Fedintsev A. et al. (2017) Markers of arterial health could serve as accurate non-invasive predictors of human biological and chronological age. *Aging* 9: 1280.
- Fitzenberger E. et al. (2014) The polyphenol quercetin protects the mev-1 mutant of *Caenorhabditis elegans* from glucose-induced reduction of survival under heat-stress depending on SIR-2.1, DAF-12, and proteasomal activity. *Mol. Nutr. Food Res.* 58 (5): 984–994.
- Fraga M.F., Esteller M. (2007) Epigenetics and aging: the targets and the marks. *Trends. Genet.* 23 (8): 413–418.
- Gunn D.A. et al. (2008) Perceived age as a biomarker of aging: a clinical methodology. *Biogerontology* 9 (5): 357–364.
- Hastie T., Tibshirani R., Friedman J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer Series in Statistics.
- He L. et al. (2007) A microRNA component of the p53 tumor suppressor network. *Nature* 447 (7148): 1130–1134.
- Helgeson J., Rammage M., Urman A., Roebuck C.M., Coverdill S., Pomerleau K., Dankwa-Mullan I., Liu L.-I., Sweetman R.W., Chau Q., Williamson P.M., Vinegra M., Haddad T.C., Goetz M.P. (2018) Clinical

performance pilot using cognitive computing for clinical trial matching at Mayo Clinic. *Journal of Clinical Oncology* 36(15): e18598-e18598.

Holford N. H. G., Kimko H. C., Monteleone J. P. R., Peck C. C. (2000) Simulation of Clinical Trials. *Annual Review of Pharmacology and Toxicology* 40 (1): 209–234.

Horvath S., Raj K. (2018) DNA methylation-based biomarkers and the epigenetic clock theory of aging. *Nat. Rev. Genet.* 19 (6): 371–384.

Hoy A.M., Buck A.H. (2012) Extracellular small RNAs: what, where, why? *Biochem. Soc. Trans.* 40 (4): 886–890.

Huang X. et al. (2013) Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genom.* 14: 319.

Hunter M.P. et al. (2008) Detection of microRNA expression in human peripheral blood microvesicles. *PLoS ONE* 3 (11): 3694.

Inukai S. et al. (2012) Novel microRNAs differentially expressed during aging in the mouse brain. *PLoS ONE* 7 (7): 40028.

Kato M. et al. (2009) The mir-34 microRNA is required for the DNA damage response in vivo in *C. elegans* and in vitro in human breast cancer cells. *Oncogene* 28 (25): 2419–2424.

Kato M. et al. (2011) Age-associated changes in expression of small, noncoding RNAs, including microRNAs in *C. elegans*. *RNA* 17 (10): 1804–1820.

Ke G., Meng Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *NIPS*.

Kingma D.P., Ba, J. (2015). Adam: A Method for Stochastic Optimization. *CoRR*, abs / 1412.6980.

Kupryushin A.S., Markova A.A., Kupryushina N.V., Vishnyakova Z.S., Latynova I.V., Semina M.N. THE STRUCTURE OF THE CAUSES OF DEATH OF THE ELDERLY AND OLD PATIENTS, DEATH AT HOME // Modern problems of science and education. – 2016. - No. 3.

Larson K. et al. (2012) Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet* 8 (1): e1002473.

Lehmann S.M. et al. (2012) An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration. *Nat. Neurosci.* 15 (6): 827–835.

Levine M.E. et al. (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10 (4): 573–591.

- Liu F. et al. (2016) The MC1R gene and youthful looks. *Curr. Biol.* 26 (9): 1213–1220.
- Liu Z., Kuo P.-L., Horvath S., Crimmins E., Ferrucci L., Levine M. (2018). Phenotypic age: a novel signature of mortality and morbidity risk. *bioRxiv*363291 [Preprint].
- Loseva P. (2020) Counterclockwise / Alpina Digital, 2020 - (Primus).
- Machida T. et al. (2015) MicroRNAs in salivary exosomes as potential biomarkers of aging. *Int. J. Mol. Sci.* 16 (9): 21294–21309.
- Mamoshina P., Kochetov K., Putin E., Cortese F., Aliper A., Lee WS, Ahn SM, Uhn L., Skjodt N., Kovalchuk O., Scheibye-Knudsen M., Zhavoronkov A. (2018) Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. *J. Gerontol. A Biol. Sci. Med. Sci.* 73 (11): 1482–1490.
- Mamoshina P., Vieira A., Putin E., Zhavoronkov A. (2016) Applications of deep learning in biomedicine. *Mol. Pharm.* 13: 1445–1454.
- Mangiola F. et al (2018) Gut microbiota and aging. *Eur. Rev. Med. Pharmacol. Sci.* 22 (21): 7404–7413.
- Mitnitski A., Rockwood K. (2015) Aging as a process of deficit accumulation: its utility and origin. *Interdiscip. Top. Gerontol.* 40: 85–98.
- Mitnitski A., Song X., Skoog I., Broe G.A., Cox J.L., Grunfeld E., Rockwood K. (2005) Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. *J. Am. Geriatr. Soc.* 53 (12): 2184–2189.
- Mitnitski A., Rockwood K. (2019) The Problem of Integrating of Biological and Clinical Markers of Aging. In: Moskalev A. (eds) *Biomarkers of Human Aging. Healthy Aging and Longevity*, vol 10. Springer, Cham.
- Mitnitski A.B., Mogilner A.J., Rockwood K. (2001) Accumulation of deficits as a proxy measure of aging. *Sci. World J.* 1: 323–336.
- Miyawaki S. et al. (2016) Facial pigmentation as a biomarker of carotid atherosclerosis in middle aged to elderly healthy Japanese subjects. *Skin. Res. Technol.* 22 (1): 20–24.
- Moskalev A., editor. *Biomarkers of Human Aging*. Springer International Publishing; 2019.
- Moskalev A. (2019) Introduction. In: Moskalev A. (eds) *Biomarkers of Human Aging. Healthy Aging and Longevity*, vol 10. Springer, Cham.
- Neri F. et al. (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543 (7643): 72–77.

- Nevalainen T. et al. (2017) Obesity accelerates epigenetic aging in middle-aged but not in elderly individuals. *Clin. Epigenetics* 9:20.
- O'Sullivan R.J. et al. (2010) Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* 17 (10): 1218–1225.
- Olivieri F. et al. (2017) Circulating miRNAs and miRNA shuttles as biomarkers: perspective trajectories of healthy and unhealthy aging. *Mech. Aging Dev.* 165 (Pt B): 162–170.
- Peleg S. et al. (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* 328 (5979): 753–756.
- Piazzesi A. et al. (2016) Replication-independent histone variant H3.3 controls animal lifespan through the regulation of pro-longevity transcriptional programs. *Cell. Rep.* 17 (4): 987–996.
- Putin E., Mamoshina P., Aliper A., Korzinkin M., Moskalev A., Kolosov A., Ostrovskiy A., Cantor C., Vijg J., Zhavoronkov A. (2016) Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging (Albany NY)*. 8 (5): 1021–33.
- Pyrkov T.V., Fedichev P.O. (2019) Biological Age is a Universal Marker of Aging, Stress, and Frailty. In: Moskalev A. (eds) *Biomarkers of Human Aging. Healthy Aging and Longevity*, vol 10. Springer, Cham.
- Rockwood K., Song X., Mitnitski A. (2011) Changes in relative fitness and frailty across the adult lifespan: evidence from the Canadian National Population Health Survey. *CMAJ* 183 (8): E487 – E494.
- Scaffidi P., Misteli T. (2006) Lamin A-dependent nuclear defects in human aging. *Science* 312 (5776): 1059–1063.
- Schubeler D. (2015) Function and information content of DNA methylation. *Nature* 517 (7534): 321–326.
- Shumaker D.K. et al. (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc. Natl. Acad. Sci. USA* 103 (23): 8703–8708.
- Singh J. et al. (2016) Aging-associated changes in microRNA expression profile of internal anal sphincter smooth muscle: role of microRNA-133a. *Am. J. Physiol. Gastrointest. Liver. Physiol.* 311 (5): G964 – G973.
- Smith-Vikos T. et al. (2016) A serum miRNA profile of human longevity: findings from the Baltimore longitudinal study of aging (BLSA). *Aging (Albany NY)* 8 (11): 2971–2987.
- Tatar M., Bartke A., Antebi A. (2003) The endocrine regulation of aging by insulin-like signals. *Science* 299 (5611): 1346–1351.
- Tazawa H. et al. (2007) Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc. Natl. Acad. Sci. USA*. 104 (39):

15472–15477.

Tian Y. et al. (2016) Mitochondrial stress induces chromatin reorganization to promote longevity and UPR (mt). *Cell* 165 (5): 1197–1208.

Vasilkov V.G., Safronov A.I. (2017) CRITICAL MEDICINE: INTENSIVE DIAGNOSTICS. TOPICAL ISSUES IN DIAGNOSIS, TREATMENT AND REHABILITATION OF PATIENTS Materials of the 19th anniversary interregional scientific-practical conference dedicated to the 40th anniversary of the ISPU - a branch of the FSBEI DPO RMANPO of the Ministry of Health of Russia.

Veytsman B., Cui T., Baranova A. (2019) Practical Detection of Biological Age: Why It Is not a Trivial Task. In: Moskalev A. (eds) Biomarkers of Human Aging. Healthy Aging and Longevity, vol 10. Springer, Cham.

Vickers K. C., et al. (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell. Biol.* 13 (4): 423–433.

Wagner W. (2017) Epigenetic aging clocks in mice and men. *Genome. Biol.* 18 (1): 107.

Weber J. A., et al. (2010) The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56 (11): 1733–1741.

Woo M. (2019) An AI boost for clinical trials. *Nature* 573: 100–102.

Wood J. G., et al. (2010) Chromatin remodeling in the aging genome of *Drosophila*. *Aging Cell* 9 (6): 971–978.

Yang X., et al. (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26 (4): 577–590.

Zhang W., et al. (2015) Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* 348 (6239): 1160–1163.

Zhao Q., et al. (2016) Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nat. Commun.* 7: 12464.

Figures

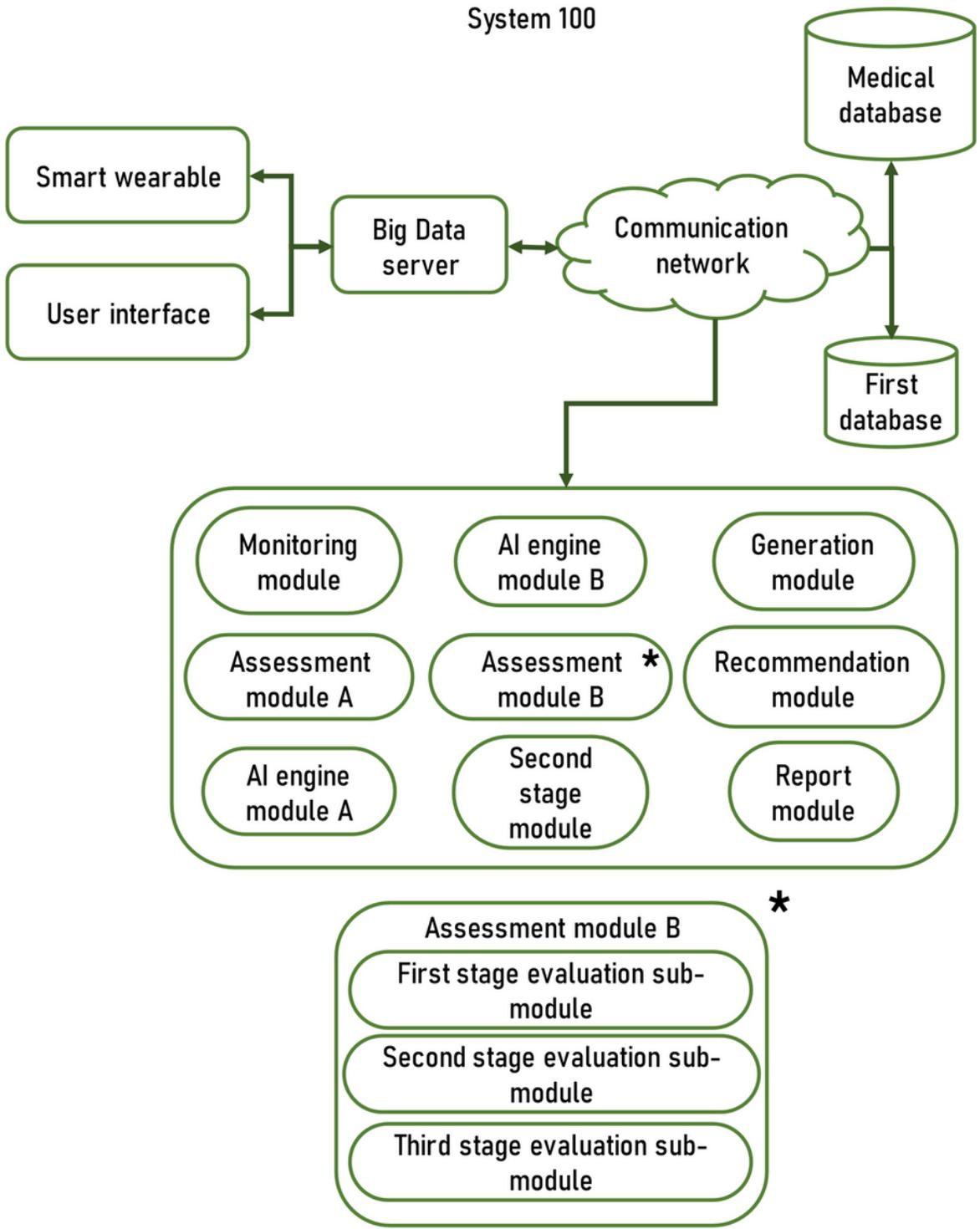


Figure 1

Block diagram of a forecasting system for assessing the duration and quality of life

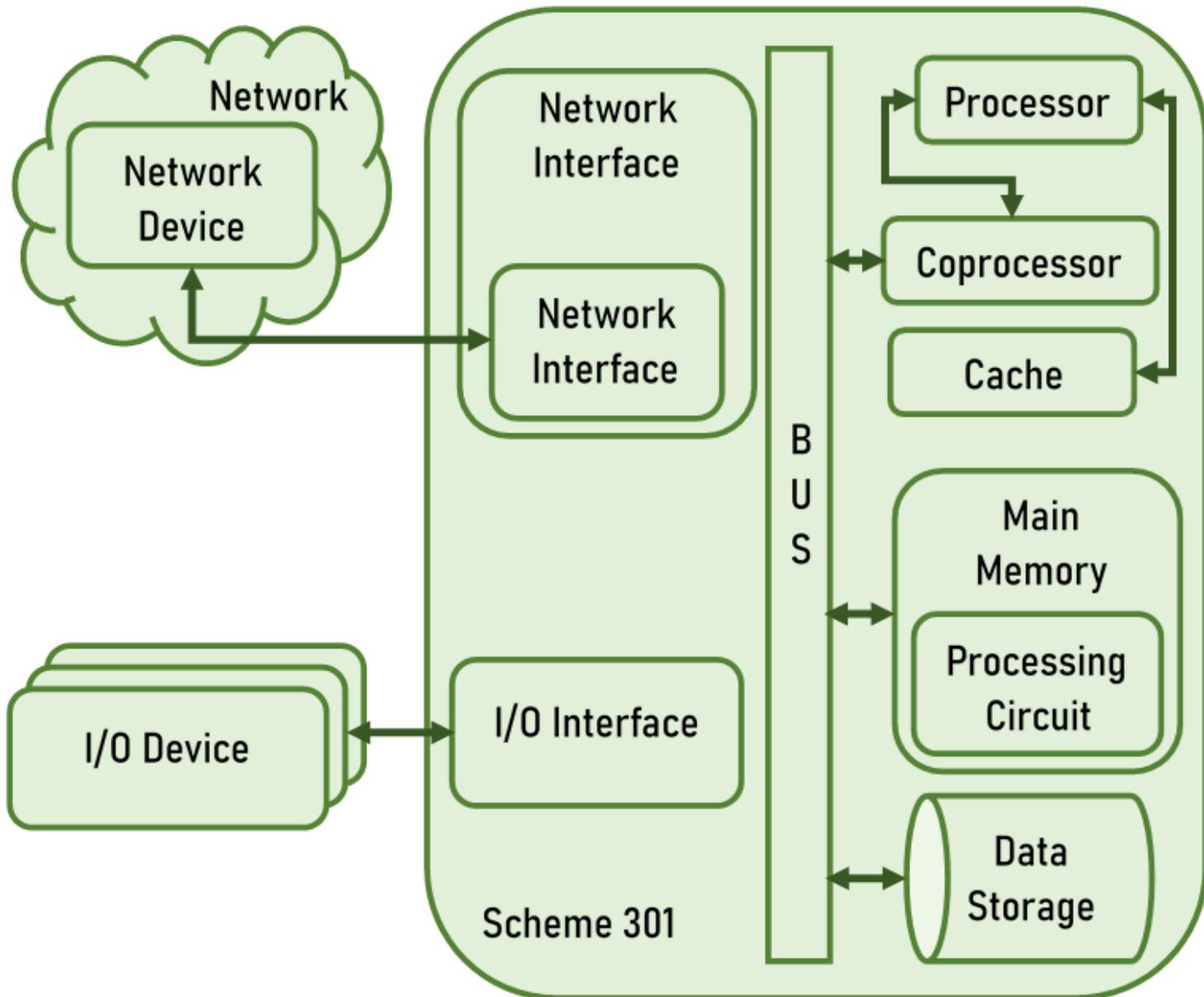


Figure 2

Block diagram of the hardware forecasting system

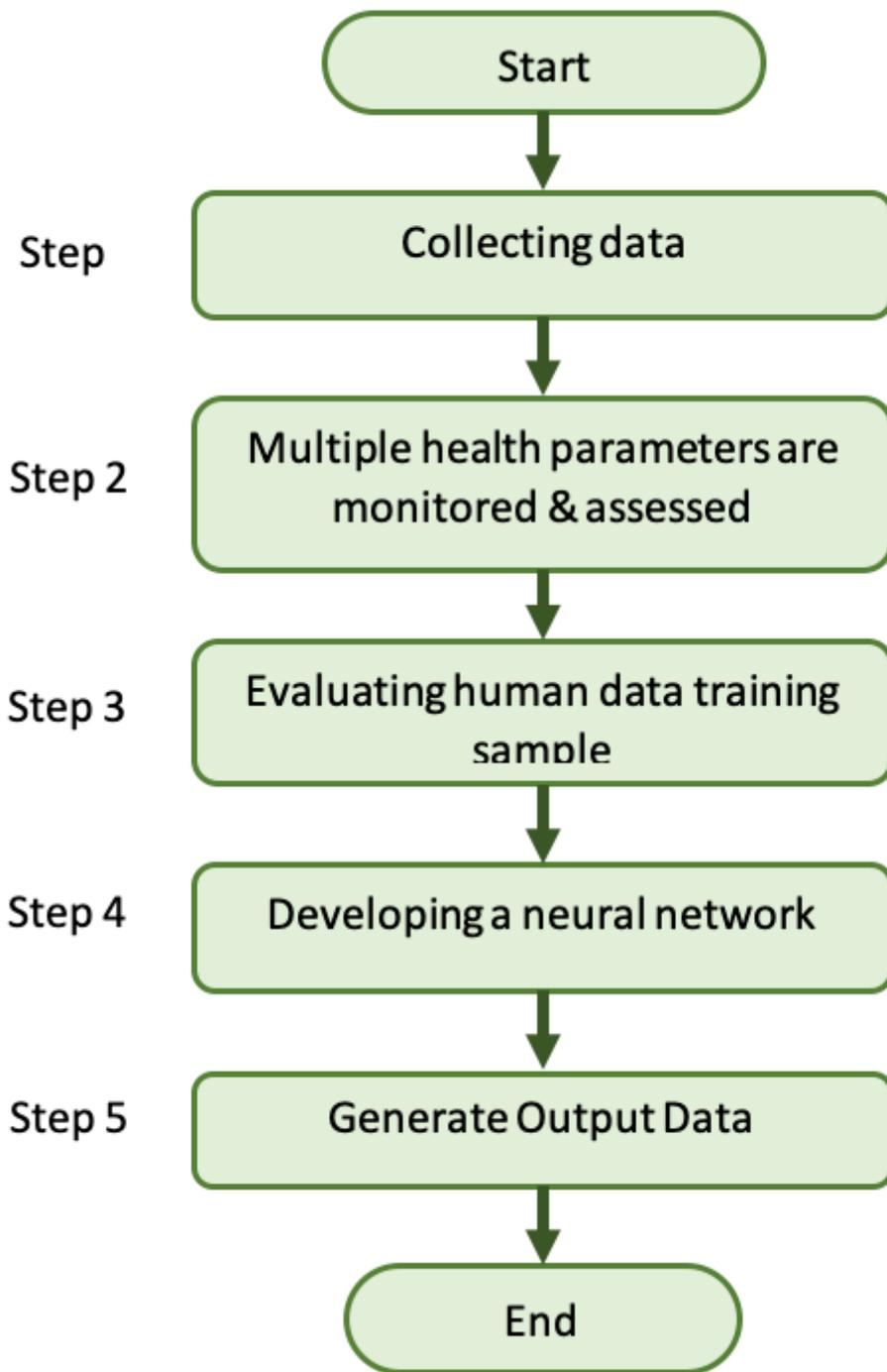


Figure 3

Flowchart of a method for predicting quality and longevity in the present invention