DT-DNA: Devising a DNA Paradigm for Modeling Health Digital Twins

by

Hawazin Faiz Badawi

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Computer Science

uOttawa

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

The potential of Digital twin (DT) technology outside of the industrial field has been recognized by researchers who have promoted the vision of applying DTs technology beyond manufacturing, to purposes such as enhancing human well-being and improving quality of life (QoL). The expanded definition of DTs to incorporate living and nonliving physical entities into the definition of DTs was a key motivation behind the model introduced in this thesis for building health digital twins of citizens. In contrast with DTs that have been developed in more industrial fields, this type of digital twins modeling necessitates protecting each citizen's unique identity while also representing features common to all citizens in a unified way. In nature, DNA is an example of a model that is both unified, common to all humans, and unique, distinguishing each human as an individual. DNA's architecture is what inspired us to propose a digital twin DNA (DT-DNA) model as the basis for building health DTs for citizens. A review of the literature shows that no unified model for citizens' health has been developed that can act as a base for building digital twins of citizens while also protecting their unique identity thus we aim to fill this gap in this research.

Accordingly, in this thesis, we proposed a DT-DNA model, which is specifically designed to protect the unique identity of each citizen's digital twin, similar to what DNA does for each human. We also proposed a DT-DNA-based framework to build standardized health digital twins of citizens on micro, meso and macro levels using two ISO standards: ISO/IEEE 11073 (X73) and ISO 37120.

To achieve our goal, we started by analyzing the biological DNA model and the in-

fluencing factors shaping health in smart cities. The purpose of the first is to highlight the DNA model features which provide the building blocks for our DT-DNA model. The purpose of the latter is to determine the main bases of our DT-DNA model of health DTs. Based on the analysis results; we proposed DT-DNA to model health DTs for citizens. In keeping with our DNA analogy, we have identified four bases, A, T, G, and C, for our unified and unique DT-DNA model. The A base in the proposed model represents a citizen's anthropometric when we build the DT-DNA on an individual level and represents the city's regulatory authorities when we build the DT-DNA on community and city levels. The T base represents different tasks included in the provided health data that are required to model citizens' health DT-DNA on different levels. The G base represents the geographic and temporal information of the city, where the citizen exists at the time of data collection. The C base represents the context at the time of data collection.

To proof the concept, we present our initial work on building health DTs for citizens in four case studies. The first two case studies are dedicated for health DTs at the micro level, the third case study is dedicated for health DTs at the meso level and the fourth case study is dedicated for health DTs at the macro level. In addition, we developed an algorithm to compare cities in terms of their community fitness and health services status. The four case studies provide promising results in terms of applicability of the proposed DT-DNA model and framework in handling the health data of citizens, communities and cities, collected through various sources, and presenting them in a standardized, unique model.

Acknowledgements

This work would not have been possible without Allah (swt). First and foremost, I would like to thank Him for giving me the health, strength, and patience to persevere during my studies.

It is a pleasure to thank those who have also made this work possible... First, I would like to express my sincere and deep gratitude and appreciation to my great supervisor Prof. Abdulmotaleb El Saddik for his help, advice, continued support, patience, encouragement and guidance from the earliest stages of this research. This thesis would not have come to light without your invaluable support. I am extremely grateful to have been supervised by such an excellent and knowledgeable professor. Conducting my graduate studies under your supervision enabled me to enhance my research skills, enrich my academic experiences, and learn a lot of life lessons. Your commitment to weekly lab meetings and your insightful comments and continued encouragement to be active in research community were always valuable chances for me to widen my knowledge in research and in everyday academic life. Thank you very much Prof, for everything.

Also, I would like to express my sincere gratitude and appreciation to my friend Dr. Fedwa Laamarti for her valuable advice, comments, and continuous support during the progress of this work. Special thanks also go to my thesis committee; the external examiner: Prof. Zhongfei Mark Zhang, OCICS examiners: Prof. Hussein Al Osman, Prof. Shichao Liu, and Prof. Amiya Nayak, and defence chairperson: Prof. Davide Spinello. Thank you for your valuable comments and feedback on this thesis.

I would also like to express my sincere gratitude and appreciation to my beloved husband Omar and my sweethearts: Abdulrahman, Dana and Imran. No words can embody my heartfelt gratitude and love to you. Thank you for providing me with endless love and limitless support and encouragement throughout the PhD journey. Thank you for all the days and nights that you spent supporting me to finish the work. Thank you for all the encouragement, hugs and good food prepared with love. It was a challenging and amusing journey loaded with surprising memories and we gratefully now celebrate the fruits of this success. Thank you for believing in me and being proud of me. Thank you for being in my life, filling my heart with joy and warmth; you are always the source of my happiness and strength, and the reason I have made it so far in life.

I would also like to express my sincere gratitude and appreciation to my beloved family. For my father Faiz, who passed away at the beginning of my master's degree journey with the hopes of seeing me as a professor, I did it my love, and our dream became a reality. For my mother Aishah, you are the original source of unconditional love, strength, support, and everything to me in this life. What I have achieved and will achieve in this life is possible after Allah's help because of you and your presence, support, prayers and encouragement. For my beloved brothers and sisters: Abdulmonem, Amina, Sameer, Halah, Heyam, Yosri and Yaser, my beloved father in-law and mother in-law, my lovely nephews and nieces especially Ammar, Ayah, Anwar, and Saeed, may Allah protect and bless all of you. I would like to express my sincere gratitude to all of you for your faith in me and for understanding and supporting me throughout my graduate studies. You have been a source of inspiration for me, and your continued support allowed me to be as ambitious as I aspired to be.

I would also like to express my sincere gratitude and appreciation to my colleagues at the Multimedia and Communications Research Laboratory (MCRLab) for the good times we spent together in this friendly space. Thank you for the wonderful memories during these years that will forever occupy a special place in my heart. Also, a special thanks goes to my best friends: Samah Aloufi, Rajwa Alharthi, Fatimah Alzamzami, Fedwa Laamarti and Jumanah Abusulaiman. Thank you for your help, support, and simply being amazing friends. I always considered you as my second set of sisters. I also want to thank my dear friends and people who have positively touched my life especially here in Canada. Your constant presence always made me feel like I was never away from my immediate family.

I would also like to express my sincere gratitude and appreciation to those behind the knowledgeable experiences I have been engaged in during the PhD journey. Special thanks go to the Peak Center for Human Performance team and especially Ken Brunet for his collaboration and time meeting with us and answering our questions. Also, special thanks go to Dr. Sarah Fraser and her team, especially Talia Salzman for the collaboration opportunity to perform multidisciplinary research. I would also like to thank the team behind the CREATE-BEST program and especially the program manager, my colleague Dr.Faisal Arafsha, for the invaluable experience – I learned a lot. I would also like to thank the team behind Simon Nehme Summer Entrepreneurship school for selecting me to participate in the 2018 school year. This experience enriched my knowledge in the field which, I am confident, will open many doors for me in the future.

Last but not least, I would like to express my sincere gratitude and appreciation to every-

one who lent a helping hand during my PhD journey. A special thanks goes to Heather Ellwood, Lyra Magloughlin, Annik Dion, Philippe Jolette and all the staff in the Faculty of Engineering and at InfoService and OLBI at the University of Ottawa. Thank you for your collaboration and always being there to help and support.

Finally, special thanks go to my beloved country, The Kingdom of Saudi Arabia, and my sponsor, Umm Al-Qura University in my beloved and home city, Makkah Al-Mukarramah, for the generous scholarship, and to the Saudi Cultural Bureau in Canada for their efforts throughout the years of my graduate studies.

الْحَمْدُ لِلهِ الَّذِي بِنِعْمَتِهِ تَتِمُّ الصَّالِحَاتُ

(Praise is to Allah by Whose grace good deeds are completed)

Dedication

To my parents Aishah and Faiz,
For instilling in me the value of education, believing in me,
and for making every possible effort to get me where I am today

To my husband Omar and my children Abdulrahman, Dana and Imran,
For being in my life and providing me with the endless love and limitless support.
You are the source of happiness and strength.

To my brothers and sisters, especially Yaser and Yosri,
To my father in-law and mother in-law, Abdulrahman and Hind,
To my extended Family,

Thank you all for supporting me and being proud of me.
Your faith in me always motivates and inspires me to pursue my dreams.

"وَقُل رَّبِّ زِدْنِي عِلْمًا" سورة طه - الجزء ٢٠:١١٤

"O my Lord! Increase me in knowledge." Surat Taha 20:114

Table of Contents

List of Tables xii

List of Figures xiv

List of Abbreviations xvii

1 Introduction 1

1.1 Motivation . . . 2
1.2 Application Scenario and Requirements . . . 4
1.3 Research Objective . . . 5
1.4 Thesis Contributions . . . 7
1.5 Thesis Organization . . . 8
1.6 Scholarly Achievements . . . 9
1.6.1 Research in Refereed Journals . . . 9
1.6.2 Research in Conference Proceedings and Book Chapters: . . . 10

2 Background and Related Work 12

2.1 Background . . . 13
2.1.1 Digital Twins (DTs) . . . 13
2.1.2 DNA . . . 14

2.2 Digital Twins for Health and Well-being: State-of-the-Art . . . 16
2.3 Existing Models for Citizens' Health . . . 18
2.4 Health Digital Twins and Standardization . . . 22
2.4.1 ISO/IEEE 11073 . . . 22
2.4.2 ISO 37120 . . . 23
2.5 Towards Building Health DT-DNA for Citizens . . . 24
2.5.1 LT: Definitions, Terms, and Categorizations . . . 24
2.5.2 X73-PHD Standards Compliant Systems . . . 34
2.6 Summary Remarks . . . 37

3 DT-DNA Paradigm for Modeling Health Digital Twins of Citizens 39
3.1 Mapping Between DNA & DT-DNA: Requirements Analysis . . . 40
3.1.1 DNA Model . . . 40
3.1.2 Anatomy of Health in Smart Cities . . . 41
3.2 Proposed Digital Twin DNA Model . . . 44
3.3 DT-DNA-Based Framework to Build Health Digital Twin of Citizens . . . 50

4 DT-DNA Paradigm at the Micro Level 56
4.1 Case Study 1: Estimating LT Towards Building DT-DNA of Physically Active Citizens . . . 57
4.1.1 Importance of LT as an Indicator of Fitness Level for Individuals . . . 58
4.1.2 Existing Methods for Estimating LT . . . 59
4.1.3 LT Estimation under DT-DNA-Based Framework: Method Description 60
4.1.4 DT Data Source and Standardization Module . . . 61
4.1.5 Data Analytics Module . . . 62
4.1.6 DT-DNA Modeling and Visualizing Modules . . . 69
4.1.7 Case Study Limitations . . . 72

4.2 Case Study 2: Measuring Gait Speed Towards Building the DT-DNA of Senior Citizens for Fall Prevention . . . 74
4.2.1 Measurement Methods and the Importance of Gait Speed for Senior Citizens . . . 75
4.2.2 Proposed Standard for Smart Shoe Insoles (SI) . . . 76
4.2.3 Measuring Gait Speed under the DT-DNA-Based Framework: Method Description . . . 84
4.2.4 DT Data Source and Standardization Module . . . 86
4.2.5 Data Analytics Module . . . 87
4.2.6 DT-DNA Modeling and Visualizing Modules . . . 94

5 DT-DNA Paradigm at the Meso and Macro Levels 99

5.1 Case Study 3: Building DT-DNA of Community Fitness in Las Vegas and Oklahoma City . . . 100
5.1.1 DT Data Source . . . 101
5.1.2 Data Standardization Module . . . 102
5.1.3 Data Analytics Module . . . 103
5.1.4 DT-DNA Modeling and Visualization Modules . . . 107
5.2 Case Study 4: Building DT-DNA of Health Services in Boston and Quebec City . . . 110
5.2.1 DT Data Source, Standardization and Analytics Modules . . . 111
5.2.2 DT-DNA Modeling and Visualizing Modules . . . 111
5.3 Algorithm 1: Which City Has Better (Community Fitness/Health Services) Towards Enhancing QoL? . . . 115
5.4 Results and Discussion . . . 118

6 Conclusion and Future Work 122

References 127

APPENDICES 143

A Tables **144**

B Visualization of built DT-DNAs - Case Study 1 **155**

C Visualization of built DT-DNAs - Case Study 2 **166**

List of Tables

2.1 Summary of existing citizens' health models in the literature compared to the proposed model in this thesis 21

2.2 Terms to express LT in the literature, according to [37] 27

3.1 List of all continents and the proposed code to be used in the G base . . . 45

3.2 Environmental Context (EC) criteria according to ISO 37120 [120], to be used in the DT-DNA model 47

4.1 Anthropometric features - all subjects 63

4.2 Achieved correlation coefficient by applying MLP on different anthropometric data combinations 68

4.3 Participants' gender, age, previous falling incidents (if any) and health condition(s) 88

4.4 Mean gait speed (MGS) for each participant in SMup, DTup, SMdown and DTdown 91

5.1 Las Vegas and Oklahoma City data collected from the ACSM Fitness Index portal [10] on January 31st, 2020 and coded according to the proposed DT-DNA model 105

5.2 ISO 37120 Health service indicators [120], their codes and coded values for QC and BO 112

5.3 Results of applying Algorithm 1 on LV vs. OC data 119

5.4 Results of applying Algorithm 1 on QC vs. BO data 121

A.1 A complete list of top-level domains (TLDs) of all countries developed and maintained by (IANA) [118] 145
A.2 Year-code to be used in the G base of the proposed DT-DNA model 151
A.3 Government Type to be used in Authority (A) base in the suggested DT-DNA model 152
A.4 Proposed 2-alphabet code for numbers from 0 to 100 to be used in the proposed DT-DNA model 153
A.5 Proposed code for months from January to December to be used in the proposed DT-DNA model 154

List of Figures

2.1 DNA in biology [14] 15
2.2 DTwins Ecosystem for Health and Well-being (adapted from [8]) 17
2.3 LT literature categorization 28
2.4 Classification of personal health systems compliant with the X73-PHD standards in the literature according to the compliant system usage, adapted from [9] 38

3.1 T base - sample code 50
3.2 Proposed DT-DNA-based framework to build health digital twins of citizens 51

4.1 Case Study 1: Estimating LT towards Building DT-DNA of Physically Active Citizens (adapted from Figure 3.2) 58
4.2 Procedure for filling missing values in the dataset 65
4.3 Proposed model for non-invasive LT estimation using MLP (adapted from [86]) 67
4.4 Actual vs. predicted LT scores for 15 subjects of the experimental sample . 69
4.5 DT-DNA visualization for Subject 1 and Subject 2 of the experimental sample 73
4.6 Case Study 2: Measuring gait speed towards building the DT-DNA of senior citizens for fall prevention (adapted from Figure 3.2) 76
4.7 Smart shoe insole used in developing the X73-PHD standard for the SI (adapted from [20]) 78
4.8 Proposed Domain Information Model (DIM) for the SI (adapted from [20]) 80

4.9 Proposed Service Model and Communication Model for the SI (adapted from [20]) . . . 82
4.10 Classification of classes in the standard implementation for the SI (adapted from [20]) . . . 84
4.11 Professor El Saddik is fitted with the various devices to be used in this case study during a preparation session prior to data collection . . . 86
4.12 Plantar pressure pattern of sensor 3 for DT stair descent of participant ID 7 89
4.13 The total distance walked by each participant in this case study . . . 90
4.14 Participants' mean gait speed (MGS) in SM vs. DT during stair climbing . 93
4.15 Participants' mean gait speed (MGS) in SM vs. DT during stair descent . 94
4.16 DT-DNA visualization for Participant 1 and Participant 2 of this study . . 98

5.1 Case Study 3: Building DT-DNA of community fitness in Las Vegas and Oklahoma City based on ISO 37120 (adapted from Figure 3.2) . . . 101
5.2 Suggested 37120 Wrapper . . . 103
5.3 T base - sample code . . . 104
5.4 A, T, C and G bases, values of the proposed DT-DNA for LV and OC and results visualized in DNA for community fitness . . . 109
5.5 Case Study 4: Building DT-DNA of Health Services in Boston and Quebec City based on ISO 37120 [120] (adapted from Figure 3.2) . . . 110
5.6 A, T, C and G bases, values of the proposed DT-DNA for QC and BO and results visualized in DNA sequences . . . 114

B.1 DT-DNA visualization for Subject 1 and Subject 2 of the experimental sample 156
B.2 DT-DNA visualization for Subject 3 and Subject 4 of the experimental sample 157
B.3 DT-DNA visualization for Subject 5 and Subject 6 of the experimental sample 158
B.4 DT-DNA visualization for Subject 7 and Subject 8 of the experimental sample 159
B.5 DT-DNA visualization for Subject 9 and Subject 10 of the experimental sample . . . 160
B.6 DT-DNA visualization for Subject 11a and Subject 11b of the experimental sample . . . 161

B.7 DT-DNA visualization for Subject 12a and Subject 12b of the experimental sample . . . 162
B.8 DT-DNA visualization for Subject 13a and Subject 13b of the experimental sample . . . 163
B.9 DT-DNA visualization for Subject 14a and Subject 14b of the experimental sample . . . 164
B.10 DT-DNA visualization for Subject 15a and Subject 15b of the experimental sample . . . 165

C.1 DT-DNA visualization for Participant 1 and Participant 2 of this study . . 167
C.2 DT-DNA visualization for Participant 3 and Participant 4 of this study . . 168
C.3 DT-DNA visualization for Participant 5 and Participant 6 of this study . . 169
C.4 DT-DNA visualization for Participant 7 and Participant 8 of this study . . 170
C.5 DT-DNA visualization for Participant 9 and Participant 10 of this study . 171
C.6 DT-DNA visualization for Participant 11 and Participant 12 of this study . 172
C.7 DT-DNA visualization for Participant 13 and Participant 14 of this study . 173
C.8 DT-DNA visualization for Participant 15 and Participant 16 of this study . 174
C.9 DT-DNA visualization for Participant 17 and Participant 18 of this study . 175
C.10 DT-DNA visualization for Participant 19 and Participant 20 of this study . 176

List of Abbreviations

ACSM: The American College of Sports Medicine

AT: Anaerobic Threshold

BMI: Body Mass Index

cm: centimeters

CM: Communication Model

CV: Critical Velocity

CVT: Constant Velocity Test

DIM: Domain Information Model

DNA: Deoxyribonucleic acid

DT-DNA: Digital Twin DNA

DT: Digital Twin

fNIRS: Functional near-infrared spectroscopy

FOR: Fat Oxidation Rate

G20: Group of 20 (of the world's largest economies that meets to coordinate global policy)

GDP: Gross Domestic Product

HR: Heart Rate

HRmax: Maximal Heart Rate

HW: Hardware

IANA: Internet Assigned Numbers Authority

IAT: Individual Anaerobic Threshold

IGT: Individual Glucose Threshold

ISO: International Organization for Standardization

lbs: pounds

LT: Lactate Threshold

MDS: Medical Device System

MET: Metabolic Equivalents

ML: Machine Learning

MLP: Multilayer Perceptrons

MLSS: Maximal Lactate Steady State

NASA: The National Aeronautics and Space Administration

OBLA: Onset of Blood Lactate Accumulation

PHD: Personal Health Device

QoL: Quality of Life

RLT: Reverse Lactate Threshold

RPE: Rate of Perceived Excursion

RT-SA: Real-Time Sample Array

SM: Service Model

spo2: peripheral capillary oxygen saturation

T2D: Type 2 Diabetes

TLDs: Top-Level Domains

UMTT: University of Montreal Track Test

Vm3km: The velocity associated with 3,000 km running performance

Vmax: The velocity associated with $\dot{V}$O2max

$\dot{V}$O2max: Maximum rate of oxygen consumption

WCCD: World Council on City Data

WHO: World Health Organization

WLT: Wearable Lactate Threshold

X73: ISO/IEEE 11073

Chapter 1

Introduction

Digital twins (DTs) technology has come a long way in industry since it was first developed by NASA to mirror the state of health of its flying twin [1] and coined as a core concept for the industrial future by Vickers and Grieves in 2002 [2]. However, deployment of digital twins technology has remained specific to industrial fields closer to machines and physical systems, with a paucity of research on deployment in non-industrial fields closer to humans, such as health, environment and recreation. Despite digital twins (DTs) technology emerging as fifth in the top 10 strategic technology trends for 2017 by Gartner [3], fourth in the same list in 2018 [4] and 2019 [5], and fifth in The Gartner Hype Cycle for Emerging Technologies published in August 2020 [6], few researchers have recognized its potential role outside of the industrial field.

The research introduced by El Saddik in [7] is one of the leading works that promoted the vision of applying DTs technology beyond manufacturing, to purposes such as enhanc-

ing human well-being and improving quality of life (QoL) in smart cities. This work led to the expansion of the original definition of digital twins to "digital replications of living as well as nonliving entities that enable data to be seamlessly transmitted between the physical and virtual worlds" [7]. By bridging the physical and the virtual worlds, data are transmitted seamlessly allowing the virtual entity to exist simultaneously with the physical one [8]. The DTs technology facilitates monitoring, understanding and optimizing the functions of the physical entity and provides continuous feedback to improve quality of life and well-being [7], [8]. Hence, it is the convergence of several technologies such as data analytics and artificial intelligence (AI), AR/VR and haptics, IoT, data visualization techniques, cybersecurity and communication networks [7], [8].

The incorporation of living and nonliving physical entities into the definition of DTs [7] introduced the potentially unlimited benefits of using the technology to build health digital twins to monitor and understand citizens' health, with the goal of optimizing the function of health services, particularly in smart cities, and improving citizens' QoL and well-being.

1.1 Motivation

The expanded definition of digital twins was a key motivation behind the proposed model presented in this thesis for building health digital twins of citizens. In contrast with DTs that have been developed in more industrial fields, this type of digital twins modeling necessitates protecting each citizen's unique identity while also representing features common to all citizens in a unified way. In nature, DNA is an example of a model that is both

unified, common to all humans, and unique, distinguishing each human as an individual. DNA's architecture is what inspired me to propose a DT-DNA model as the basis for DTs of citizens. The proposed DT-DNA model is specifically designed to protect the unique identity of each citizen's digital twin, similar to what DNA does for each human.

While the new definition of digital twins now includes living as well as nonliving entities, and the proliferation of data-rich devices expands the capacity of the technology to readily capture information about e.g., citizens' health patterns, there remains a gap in the literature. Currently, no unified model for citizens' health has been developed that can act as a base for building digital twins of citizens living in smart cities while also protecting their unique identity. This research aims to fill this gap, as described in this thesis.

To build the proposed DT-DNA model in this thesis, we elected to use citizens' health data from various sources, based on an earlier DTwins Ecosystem for Health and Well-being [8], which is one of the first research initiatives to introduce the utilization of DTs technology in the health and well-being field. The explosive use of wearable technology and the greater availability of personal health devices (PHDs) for the general public have added new and interesting sources of citizens' health data in addition to traditional sources collected on a city level through interviews, surveys and questionnaires. A recent and systematic literature review [9] defines a PHD as "any device equipped with one or multiple sensors that are able to monitor vital signals of a person's body possibly taking into consideration signals from surrounding environment such as noise." In addition to providing insight into the health of individuals living in smart cities, PHDs can capture information about the health communities to which an individual belongs, e.g., a fitness community, diabetes community, etc.

1.2 Application Scenario and Requirements

Let us consider an application scenario. In a given smart city, local health authorities and stakeholders are planning to model citizens' health by building a unified digital twins (DTs) model that takes into consideration different types of health data, with the overarching goal of enhancing QoL. They want to be able to analyze data at both the individual and the group levels, while preserving the unique identity of each citizen. They recognize the importance of building a model that will enable representing citizens' health data in a unified way, while protecting their identity. To achieve this, local health authorities and stakeholders also need to adopt internationally-recognized standards to standardize different types of health data retrieved from heterogeneous sources. In addition, they need to consider context during data collection to track the dynamic state of their citizens' health and geo-temporal data. Finally, local health authorities and stakeholders need to be able to visualize modeling results for themselves and for those who care about citizens' health and enhancing QoL.

Using the described scenario, we are able to identify several requirements that must be met in order to build a unified model of health DTs, and to highlight the potentially significant benefits of utilizing this technology for modeling citizens' health. A robust model of health DTs for citizens living in smart cities must be:

- *Unified and unique*, accounting for all citizens' health data while protecting the unique identity of each citizen and providing the means to capture their special interests (being unified implies that the model also needs to be standardized to guarantee interoperability);

- *Contextualized*, tracking the dynamic state of the citizens' health and geo-temporal data;

- *Visualizable*, providing visual representations of results and analysis to citizens, stakeholders and regulatory authorities who care about citizens' health, the performance and sustainability of city health services, and the enhancement of those services for improved QoL.

1.3 Research Objective

Our objective in this thesis is to leverage DTs technology in modeling citizens' health. To effectively model citizens' health DTs, a mechanism is needed to handle various kinds of health data collected from heterogeneous sources, such as personal health devices and applications. The mechanism should also be standardized to guarantee the interoperability of the developed model. However, a general, standardized model of citizens' health that can act as a base for modeling citizens' health DTs is still missing. This model must be able to account for all citizens' health data while protecting the unique identity of each citizen and providing the means to capture their special interests.

DNA was the inspiration for our proposed model because it provides a unified and unique way to represent and protect the genetic blueprint of human and other organisms throughout history. Thus, we propose the digital twin DNA (DT-DNA) paradigm. We undertake the following steps to build the DNA for citizens' health DTs:

- Refine the requirements of the proposed DT-DNA model by analyzing two domains.

First, analyze the biological DNA model and highlight the features that allow us to introduce a digital twin DNA (DT-DNA). Second, analyze health in smart cities to define the bases of the DT-DNA model of citizens' health considering that the intended citizens live in a smart city where they interact with the environment and are involved in and affected by the city's health services.

- Discuss the design of the proposed DT-DNA paradigm to model health DTs of citizens in light of the analysis results to meet the requirements listed above.

- Discuss the design of the DT-DNA-Based Citizens' Health Digital Twins Framework based on two ISO standards: ISO/IEEE X73 and ISO 37120.

- Present the proof-of-concept of the DT-DNA-Based Citizens' Health Digital Twins Framework on an individual level by utilizing ISO/IEEE X73 standards in two case studies:

  - First: a case study to build the DT-DNA of physically active citizens who performed fitness assessment tests to estimate their lactate threshold (LT).
  - Second: a case study to build the DT-DNA of elderly individuals who performed a gait speed experiment utilizing standardized smart insoles to compare the speed when performing single vs. dual tasks.

- Present the proof-of-concept of the DT-DNA-Based Citizens' Health Digital Twins Framework on community and city levels by utilizing ISO 37120 standards in two case studies:

- First: a case study to build DT-DNA of community fitness in Las Vegas and Oklahoma City to show which city is more fit, utilizing health data collected from citizens who are engaged in fitness activities in these cities, and documented under the American College of Sports Medicine (ACSM) Fitness Index [10]
- Second: a case study to build the DT-DNA of ISO 37120-based health services in Boston and Quebec City in 2017, to show which city had better services in that year towards enhancing QoL, utilizing health service data collected from citizens across the city and documented under the World Council on City Data (WCCD) portal [11].

Our research question in this thesis is as follows:

> How can we design and develop a unified model of citizens' health DTs that takes into consideration health data at different levels and preserves the unique identity of each citizen towards enhancing citizens' Quality of Life?

1.4 Thesis Contributions

The main contribution of this thesis is the design, development and evaluation of a digital twins DNA (DT-DNA) paradigm that models citizens' health and allows for the comparison of health data sourced from individuals and groups, and at a city level, by emulating the double helix DNA model. This work also makes the following contributions:

- Mapping between the double helix DNA model and the proposed DT-DNA;

- Design and development of an ISO standard-based framework to build DT-DNA;
- Design and development of an algorithm to compare cities to show which city has better community fitness and health services towards enhancing QoL based on the proposed DT-DNA framework;
- Standardization of a smart insole, which is utilized as a personal health device for data collection in accordance with the X73 PHD standard;
- Design and development of a non-invasive model for estimating lactate threshold (LT) using a machine learning (ML) algorithm as a step towards facilitating non-invasive performance monitoring.

1.5 Thesis Organization

The remainder of this thesis is organized as follows:

- In Chapter 2, we review the background related to the work presented in this thesis.
- In Chapter 3, we present the requirements analysis and propose the DT-DNA model and framework to build citizens' health DTs.
- In Chapter 4, we present a proof-of-concept using individual-level health data based on ISO/IEEE X73.
- In Chapter 5, we present a proof-of-concept using group and city-level health data based on ISO 37120.

- In Chapter 6, we conclude the thesis and discuss future work.

1.6 Scholarly Achievements

The work on this thesis resulted in publications in refereed journals, conference proceedings and book chapters that validate our work in the research community as follows:

1.6.1 Research in Refereed Journals

1. **Badawi, Hawazin Faiz**, Fedwa Laamarti, and Abdulmotaleb El Saddik. "Devising Digital Twins DNA Paradigm for Modeling ISO-based City Services." ***Sensors*** 21.4 (2021): 1047. (Impact Factor 3.275).

2. Salzman, Talia, Ahmed Aboualmagd, **Hawazin Badawi**, Diana Tobón-Vallejo, Hyejun Kim, Lama Dahroug, Fedwa Laamarti, Abdulmotaleb El Saddik, and Sarah Fraser. "Prefrontal Cortex Involvement during Dual-Task Stair Climbing in Healthy Older Adults: An fNIRS Study." ***Brain Sciences*** 11, no. 1 (2021): 71. (Impact Factor 3.332).

3. Fedwa Laamarti, **Hawazin Badawi**, Yezhe Ding, Faisal Arafsha, Basim Hafidh and Abdulmotaleb El Saddik. "An ISO/IEEE 11073 Standardized Digital Twin Framework for Health and Well-being in Smart Cities." ***IEEE Access*** 8 (2020): 105950 - 105961. (Impact Factor 3.745).

4. **Badawi, Hawazin Faiz**, Fedwa Laamarti, and Abdulmotaleb El Saddik. "ISO/IEEE 11073 personal health device (X73-PHD) standards compliant systems: A systematic literature review." ***IEEE Access*** 7 (2018): 3062-3073. (Impact Factor 3.745).

5. **Badawi, Hawazin Faiz**, Haiwei Dong, and Abdulmotaleb El Saddik. "Mobile cloud-based physical activity advisory system using biofeedback sensors." (2017): 59-70. ***Future Generation Computer Systems*** 66. (Impact Factor 6.125).

1.6.2 Research in Conference Proceedings and Book Chapters:

1. **Badawi, Hawazin**, Fedwa Laamarti, Ken Brunet, Ed McNeely and Abdulmotaleb El Saddik. "Non-invasive Lactate Threshold Estimation Using Machine Learning." In ***International Conference on Smart Multimedia***, pp. 96-104, Springer, Cham, 2019.

2. **Badawi, Hawazin**, Fedwa Laamarti, Faisal Arafsha, and Abdulmotaleb El Saddik. "Standardizing a shoe insole based on ISO/IEEE 11073 personal health device (X73-PHD) standards." In ***International Conference on Information Technology & Systems***, pp. 764-778. Springer, Cham, 2019.

3. El Saddik, Abdulmotaleb, **Hawazin Badawi**, Roberto Alejandro Martinez Velazquez, Fedwa Laamarti, Rogelio Gámez Diaz, Namrata Bagaria, and Juan Sebastian Arteaga-Falconi. "Dtwins: A digital twins ecosystem for health and well-being." In ***Proc. IEEE COMSOC MMTC Commun. Frontiers***, pp. 39-43. 2019.

4. **Badawi, Hawazin Faiz**, and Abdulmotaleb El Saddik. "Biofeedback in Healthcare:

State of the Art and Meta Review." In ***Connected Health in Smart Cities***, pp. 113-142. Springer, Cham, 2020.

5. Bagaria, Namrata, Fedwa Laamarti, **Hawazin Faiz Badawi**, Amani Albraikan, Roberto Alejandro Martinez Velazquez, and Abdulmotaleb El Saddik. "Health 4.0: Digital Twins for Health and Well-Being." In ***Connected Health in Smart Cities***, pp. 143-152. Springer, Cham, 2020.

6. Fraser, Sarah, Talia Salzman, Hyejun Kim, **Hawazin Badawi**, Diana Tobon Vallejo, Yves Lajoie, Lara Pilutti, and John Farrell III. "Using fNIRS to Capture Cerebral Oxygenation in Older Adults Navigating Stairs." Innovation in Aging 4, no. Suppl 1 (2020): 792.

7. Salzman, Talia; Aboualmagd, Ahmed; Tobón-Vallejo, Diana; Dahroug, Lama; Laamarti, Fedwa; **Badawi, Hawazin**; Hafidh, Basim; El Saddik, Abdulmotaleb; Fraser, Sarah. Going Up and Going Down: How Do Older Adults Manage Dual-Tasking on Stairs? An fNIRS study. In ***Cognitive Aging Conference 2020***, pp. 81. CAC, 2020.

Chapter 2

Background and Related Work

In this chapter we discuss the background concepts that we utilized in the proposed paradigm (Section 2.1) to model standardized health digital twins (DTs) for citizens towards enhancing quality of life. Then, we present a discussion on the state-of-the-art of digital twins for health and well-being (Section 2.2), existing models for citizens' health (Section 2.3), standardization for health digital twins (Section 2.4) and work related to the case studies we conducted as proofs of concept for the proposed health DTs for citizens on the micro, meso and macro levels (Section 2.5). We conclude this chapter with summary remarks (Section 2.6).

2.1 Background

2.1.1 Digital Twins (DTs)

DTs research recently gained attention in academia for non-industrial deployment of this technology in fields that are closer to humans such as health, environment and recreation. According to [7], DTs are defined as "digital replications of living as well as nonliving entities that enable data to be seamlessly transmitted between the physical and virtual worlds. Digital twins facilitate the means to monitor, understand, and optimize the functions of all physical entities and, for humans, provide continuous feedback to improve quality of life and well-being." While this definition highlights promising improvements for human quality of life, there are many challenges associated with having a massive number of DTs that must communicate effectively while also protecting the privacy of real (human) twins. According to El Saddik [7], unique identifiers top the list of DT characteristics that may, along with other characteristics, lead to a gigantic growth in data volume. In its report, Gartner [3] predicted that hundreds of millions of things will be represented by digital twins within three to five years. It described DTs as dynamic software models that include a combination of metadata about real twins. Deloitte's report on DTs [12] highlighted the role of DT users in enriching the DTs experience in general through enabling access to larger volumes of real twins data. The report described the effect of providing more data from digital sources on the DT models as "moving from fuzzy, black-and-white snapshots to colorful, high-definition digital pictures" and formulated it as: models + data = insights and real value. Recent reports prepared and published by Gartner [3] and MarketsandMarkets [13] discussed the growing interest in DTs and how the DTs market is expected to

jump from USD 3.8 billion in 2019 to USD 35.8 billion by 2025. These promising reports on the future of DTs reinforce the need for successful modeling to keep up with the growth of DTs that is expected in different fields.

2.1.2 DNA

DNA (deoxyribonucleic acid) is the hereditary material found in humans and other organisms [14]. DNA molecules are found inside a cell's nucleus and each molecule is tightly packaged in a form called a chromosome [15]. Genes are specific sections of the chromosome that contain the instructions to produce proteins, which do most of the work in our bodies [15]. Genes represent 1% of the DNA sequence in the chromosome and the remainder of this sequence controls the time, the amount, and the method of protein production in the cell [15]. Each organism has a specific number of chromosomes that distinguishes it from others, but all share the unique double helix structure of a DNA molecule proposed by Watson and Crick [16]. The genetic instructions are stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T), which pair up to form units called base pairs, where A pairs with T and C pairs with G, and then arrange into the two long strands that form the double helix [14] depicted in Figure 2.1. These bases are ordered in a specific sequence to determine the information available for building and maintaining an organism, similar to sequencing letters of the alphabet in a certain order to form words and sentences [14]. Each human contains about 3 billion base pairs and 20,000 genes [15]. Approximately 99.9 % of these base pairs are the same in all people, while 0.1% are variants that differentiate humans and make everyone unique [14]. Each cell

in the human body contains 23 pairs of chromosomes. Of these, 22 pairs are identical in both males and females and the 23rd pair is the one that distinguishes them. The double helix structure of DNA is the unique shared model among organisms, but the sequential order of the DNA base pairs along its sides is the distinguishing factor. The ability of DNA to represent and protect the genetic blueprint of humans and other organisms throughout history in a unified and unique way is what motivated us to mimic and utilize this model for modeling health DTs of citizens.

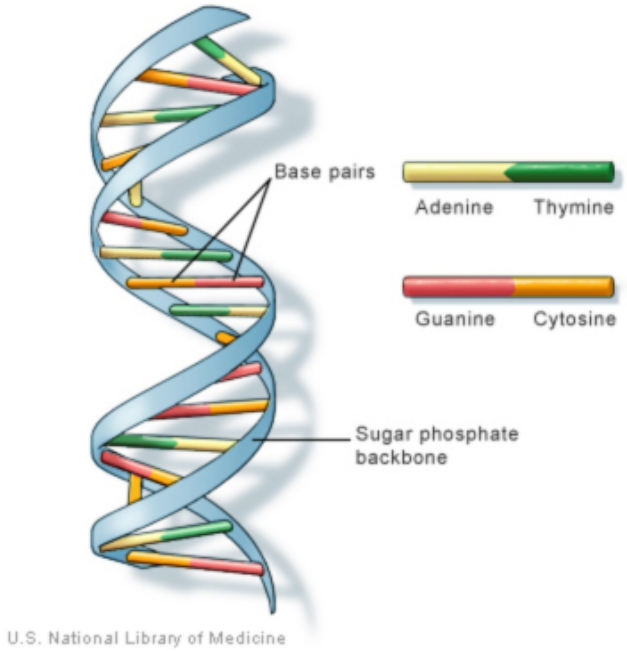


Figure 2.1: DNA in biology [14]

2.2 Digital Twins for Health and Well-being: State-of-the-Art

Due to its prominent role in enhancing individual wellness and quality of life [17], health is one of the main service areas targeted by research to leverage digital twins technology. DTwins Ecosystem for Health and Well-being [8] (depicted in Figure 2.2) was one of the first research initiatives to introduce the utilization of DTs technology in the health and well-being field. It aims mainly to provide a foundation for developing realizable DTs systems in preventive healthcare [8]. Based on this DTwins Ecosystem, research in [18] proposed an ISO/IEEE 11073 standardized digital twins framework architecture for health and well-being, geared towards fostering the use of data collected by personal health devices to benefit the individuals. It also emphasized the importance of standardization to guarantee interoperability of the developed digital twins. ISO/IEEE 11073 [19] is a family of standards developed for personal health devices to facilitate the collecting of health data by these devices. A recent literature review on systems compliant with ISO/IEEE11073 is presented in [9] and the standardization of a shoe insole based on this family of standards is presented in [20].

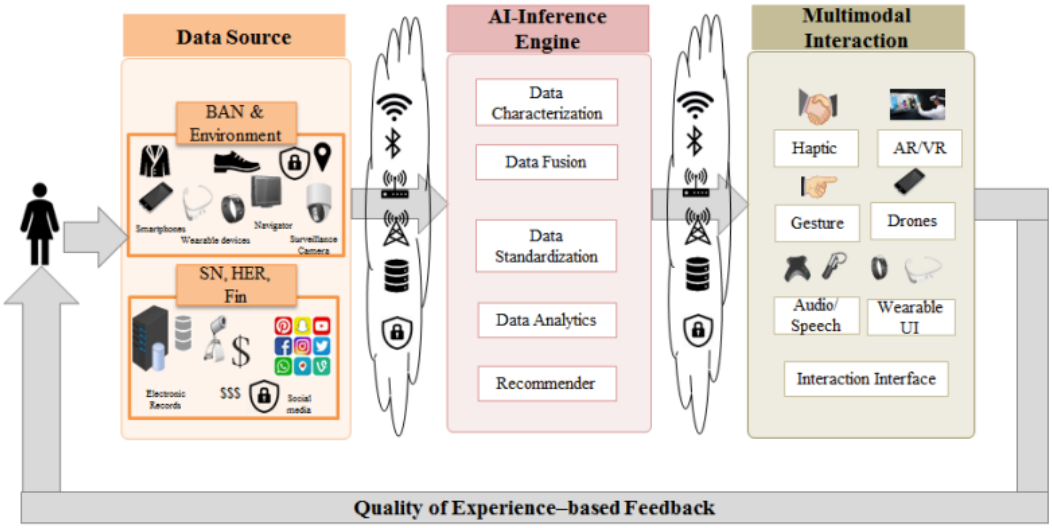


Figure 2.2: DTwins Ecosystem for Health and Well-being (adapted from [8])

Although DTs technology is currently a hot topic in academia and a top priority for several industries, a generally accepted model for standardized health digital twins (DTs) for citizens does not yet exist. Reviewing the literature shows that the majority of existing DTs are suggested implementations for specific services in the manufacturing industry. Conducting a search on Scopus including title, abstract, and keywords for *("Digital Twin∗" AND ("Citizen∗" OR "Human∗") AND ("Health∗" OR "Well-being" OR "Wellbeing") AND "model∗" AND "Standard∗")*, returns only 3 research papers. Analyzing these papers reflects recent interest in utilizing digital twins technology in the medical field in two papers [21], [22], while the third one is out of scope. This result shows the urgent need for standardized models for health digital twins of citizens to pave the road for utilizing digital twins technology in either proactive or reactive health. The following section presents the existing models for citizens' health, thereby highlighting the significance of the proposed model in this thesis.

2.3 Existing Models for Citizens' Health

To emphasize the significance of utilizing DTs technology in modeling citizens' health and the usage of the DNA analogy, we conducted a search on Scopus including title, abstract, and keywords for ("citizens" AND "health model"). We aimed to survey existing models for citizens' health to see if any existing model provides a unified modeling scheme while protecting the unique identity of each citizen, considering the context and visualizing the results. This search returned 27 research papers; five of them are actual models that we considered after our investigation.

Reviewing the literature in search results shows that the majority of existing models for citizens' health focus on patients in a reactive way, to facilitate monitoring patients' health outside of the hospital environment. They are mostly designed for modeling health at the individual level without simultaneously considering the community and city to which each individual belongs. Some of the existing models rely on analogy to model citizens' health and involve citizens' demographics at early design stages. They also consider citizens' context and are developed to serve citizens at specific locations.

Table 2.1 summarizes the five models we considered in comparison with the proposed DT-DNA model in this thesis. The first column shows the title of the research paper with bold text that highlights the purpose of the presented model in the research paper. The second column shows the model type: whether it is reactive (considers patients only) or proactive (considers all citizens). The third column shows the modeling level: whether it considers micro 'I' (individual) level only or meso 'E' (community) and macro 'A' (city) levels as well. The fourth column shows if the proposed model is standardized: 'Y' or 'N'.

The fifth column shows the purpose of the model (bold text in the title): whether it serves the general health and wellbeing purpose 'G' or is dedicated to serve a specific purpose 'S'. The sixth column shows whether the proposed model targets citizens in general 'G' or a specific group 'S'. The seventh column shows whether the proposed model involves various citizens' demographics at early design stages: 'Y' or 'N'. The eighth column shows whether the proposed model considers citizens' context: 'Y' or 'N'. The ninth column shows if the proposed model is developed to be used at specific location: 'Y' or 'N'. The tenth column shows whether the proposed model provides visualization for users: 'Y' or 'N'. The last column shows whether the proposed model uses a specific analogy for modeling: 'Y' or 'N'. According to [23], maintaining familiar shapes and concepts in the newly proposed models is essential to facilitating the acceptance of the proposed model by the targeted population thus enhancing its usability and credibility.

As shown in Table 2.1 only one model [23] other than the one proposed in this thesis is standardized, since it follows WHO standardized recommendations for physical activities. Regarding the considered context, surrounding environment is commonly considered in [24], [25], [26], and [27] in addition to biological factors and social environments where people live, work and play in [23]. Regarding the location, the model in [23] is proposed for citizens in Belgium, whereas models in [25], [26], and [27] are proposed for people in Italy, India and Norway respectively. Models in [23], [25], and [26] provide visualization of the results as a feedback to the various users of the proposed models. Regarding the usage of analogy, the proposed model in [23] uses the analogy with the active food guide pyramid whereas the proposed work in [24] models citizens' health state analogously to the modeling of a product's state from raw good to final product.

Table 2.1 highlights the gap in the literature of modeling citizens' health as well as the significance of the suggested paradigm presented in this thesis, as illustrated in the last row. However, prior to discussing its details, the following section examines the two beneficial standards for developing standardized models for health digital twins of citizens on micro, meso and macro levels.

Table 2.1: Summary of existing citizens' health models in the literature compared to the proposed model in this thesis. Legend: 'I' = "micro", 'E' ="meso", 'A' = "macro", 'Y' = "Yes", 'N' = "No", 'G' = "General", and 'S' = "Specific"

Title -**bold text highlights model purpose-** [reference] (publishing year)	Model Type		Modeling Level			Standardized?		Model Purpose		Model Target		Citizen Demographics ?		Context ?		Location Specific?		Visualize Results?		Use Analogy ?	
	Reactive	Proactive	I	E	A	Y	N	G	S	G	S	Y	N	Y	N	Y	N	Y	N	Y	N
Public health communication and education to promote more physical activity and less sedentary behaviour: Development and formative evaluation of the **' physical activity triangle'** [23] (2021)		✓	✓			✓			✓	✓		✓			✓	✓		✓			✓
Designing Smart Cities for **Citizen Health Well-being** [24] (2018)	✓		✓		✓		✓	✓		✓		✓		✓			✓		✓		
A new ICT based model for **wellness and health care** [25] (2016)		✓	✓				✓	✓		✓		✓		✓		✓		✓			✓
E-health for **rural areas** of Uttarakhand under e-Governance service delivery model [26] (2002)	✓		✓	✓	✓		✓	✓		✓			✓	✓		✓		✓			✓
A socio-ecologic model for **periodontal diseases** [27] (1993)	✓	✓	✓				✓		✓		✓	✓		✓		✓			✓		✓
DT-DNA (proposed paradigm in this thesis)	✓	✓	✓	✓	✓	✓		✓		✓		✓		✓			✓	✓		✓	

2.4 Health Digital Twins and Standardization

Standardization is an essential aspect of developing smart health digital twins to guarantee the interoperability of the digital twins on individual and group levels of citizens. Since our focus in this thesis is on smart health and wellbeing on both individual and city levels, we utilize two standards, ISO/IEEE 11073 and ISO 37120, as discussed below.

2.4.1 ISO/IEEE 11073

Personal health devices (PHDs) enable people to track their health status, manage medicine intake, and transmit collected information to healthcare professionals if required [28]. Many PHDs have been manufactured, such as blood pressure monitors, weighing scales, and insulin pumps. Using these devices is an essential part of addressing issues of widespread physical inactivity and the rapid increase in chronic diseases such as diabetes and high blood pressure [29].

Currently, personal health systems rely on PHDs and wearable technologies to perform their functions efficiently. The development of the X73-PHD standards to facilitate the leveraging of collected health data has been encouraged by the growing use of PHDs. However, utilizing PHDs developed by different manufacturers in different data formats and with different electronic features can lead to integration difficulties [30]. Standardizing PHDs is essential to providing a unified language that can be used to interact with such systems globally. This issue motivated the IEEE and the International Organization for Standardization (ISO) to create the family of standards called the ISO/IEEE 11073 Personal Health Device (X73-PHD) in 2008 [31]. This family of standards targets health

devices designed for personal use by healthy individuals and patients to facilitate health data exchange while providing plug-and-play interoperability. It aims to regulate the manufacturing of PHDs and control the interoperability among PHDs and personal health systems [32], and it provides a cost-efficient standardization solution for personal health systems. Thus, the communication between PHDs and managers such as computer systems and cell phones takes place in a direct manner and away from complex point-of-care systems used in clinics [9]. The results of a systematic literature review on X73-PHD standards compliant systems [9] are discussed in detail in Section 2.5.2.

2.4.2 ISO 37120

This standard, which is officially called "ISO 37120: Sustainable cities and communities – Indicators for city services and quality of life," is the leading international standard on city indicators, developed by the ISO sustainable cities and communities technical committee [33]. ISO 37120 was developed as a tool for city stakeholders to facilitate the implementation of policies and theoretical plans designed to promote sustainable and livable cities and enhance quality of life for all citizens. Since existing indicators are often not standardized, consistent, or comparable over time or across cities, the ISO 37120 technical committee establishes a uniform approach to measuring the performance of city services and quality of life. There are 17 themes or services for ISO 37120 and approximately 100 indicators, either core or supportive, categorized under the themes to assess the performance of various services in a city.

2.5 Towards Building Health DT-DNA for Citizens

In this section, we present work related to the case studies discussed in Chapter 4 towards building health DT-DNA at the micro level. We first discuss utilizing lactate threshold (LT) as an indicator of fitness level towards building DT-DNA of physically active citizens (Section 2.5.1). We then present the results of a systematic literature review on X73-PHD standards compliant systems, which was our initial step towards standardizing a smart in-shoe insole as a PHD for proactive well-being (Section 2.5.2). The actual standardization process is discussed in Chapter 4 under the case study of building DT-DNA of senior citizens for fall prevention (Section 4.2.2).

2.5.1 LT: Definitions, Terms, and Categorizations

A. Different Definitions and Terms

Physical activity and sports play a fundamental role in fighting the sedentary lifestyle that threatens people's health. Indeed, the sedentary lifestyle is one of the main causes of obesity, which, according to the World Health Organization (WHO) [34], is the fifth leading cause of death globally. Much research has been conducted in the areas of physical activity level [17] and sports training [35], [36]. In order to enhance an individual's engagement in physical activity and increase the benefits derived from physical activity and sports, it is important to deploy recent technologies towards increasing awareness of the factors influencing fitness level. Monitoring health and well-being indicators [7] is one of the promising means in this regard, and lactate threshold (LT) is one of the influencing factors

on an individual's physical performance level, regardless of whether they are an athlete or non-athlete. LT is defined as "a point at which blood lactate begins to increase above resting values during a graded exercise challenge" [37]. The importance of LT is that it is an indicator of individual endurance level, and hence used by coaches to predict exercise performance [38]. LT also helps in determining an individual's overall performance level in a specific sport such as running and swimming, and thus facilitates the development of individualized training routines [39]. Dtwins [8], a digital twins ecosystem for health and well-being, can be utilized effectively to increase awareness about LT by estimating, monitoring, and providing biofeedback on this vital indicator. Before discussing this possibility in Chapter 4, in this chapter we highlight the different definitions of LT in the literature, and discuss the biological process (glycolysis) that causes the formation of lactic acid and the chemical interpretation of this physiological feature. We also examine the literature discussing LT in relation to other physiological features such as $\dot{V}O2max$, heart rate (HR), fat oxidation rate (FOR), and rate of perceived excursion (RPE).

A paper published in 1985 [40] is considered a fundamental paper in the area of anaerobic threshold (AT) and lactate production. It provides a description of basic terminologies, validates the AT hypothesis, and defines future research directions based on validation results. The chemical explanation of lactic acid accumulation in cells, which interprets the LT occurrence and its relation with muscle fatigue, are discussed in detail in [37], [41], and [42]. Researchers in [41] discuss lactate accumulation during exercise and its relation with O2 in order to demonstrate that the accumulated lactate during exercise is an anaerobic metabolite. They also discussed the relationship between lactic acid, lactate and muscle fatigue, and highlighted the reasons behind increased lactate accumulation with increas-

ing exercise intensity. Researchers in [37] investigated confusion with relating the lactate with muscle fatigue and energy metabolism. They also discussed the critical role of LT in evaluating endurance levels of athletes as compared to $\dot{V}O2max$, and how to improve LT through training. They concluded their paper by highlighting the challenge of assessing LT for individual athletes compared to a group of athletes. They also differentiated between terminology that are used interchangeably with lactate threshold, leading to confusion in the sport medicine and exercise science communities. This confusion is considered one of the main difficulties in using LT as a training aid or performance predictor in endurance exercise. The terms include "lactate threshold," "anaerobic threshold," "aerobic threshold," "lactate turn point," "onset of blood lactate accumulation (OBLA)," and "maximal lactate steady state (MLSS)." The authors investigate the usage of these terms in the literature, as shown in Table 2.2. We consider the following to be synonymous: MLSS=LT=Anaerobic Threshold. Thus, when we mention lactate threshold or LT, we mean the MLSS according to [43], [44], which compares to a fixed blood lactate concentration of 4 mmol/l as justified in [45]. One of the main definitions of MLSS is "the highest blood lactate concentration (MLSSc) and work load (MLSSw) that can be maintained over time without a continual blood lactate accumulation" [46].

The authors in [42] provide a detailed explanation for blood lactate concentration due to its essential role in clinical exercise and performance testing for athletes. They discuss the biological features of blood lactate concentration to avoid confusion in the clinical tests.

The research in [39] is dedicated to discussing the validity of different LT concepts in addition to the interpretation of blood lactate curves. The authors highlight 25 different concepts found in the literature, and categorize them as aerobic or anaerobic threshold-

related. We find that this work is one of the most informative references for understanding LT-related concepts. The authors discuss the validation of LT by comparing it with current competition performance in an endurance event, referred to as concurrent validity, or by predicting endurance performance in future events, referred to as predictive validity. The authors also discuss the usage of MLSS, which is the highest exercise intensity that can be maintained for a longer period of time without rapid increase in the blood lactate concentration, to assess endurance performance and correlates it with other LTs concepts.

Term	**Definition**	**Other terms**
Onset of blood lactate accumulation (OBLA)	"A point at which blood lactate begins to increase above resting values during a graded exercise challenge." [37]	**Anaerobic threshold** by Wasserman et al. [47], **Aerobic threshold** by Kindermann et al. [43]
Maximal lactate steady state (MLSS)	"A maximal exercise intensity above which a continuous increase in blood lactate is unavoidable" [37]	**Anaerobic threshold** by Kindermann et al. [43], [44]

Table 2.2: Terms to express LT in the literature, according to [37]

B. Literature Categorization

LT plays a critical role in the sport domain in assessing an athlete's performance and determining exercise intensity. Given the increased utilization of this physiological feature in the sport and medicine fields, we classified the literature dedicated to LT into five main categories, and highlighted the related topics, as shown in Figure 2.3.

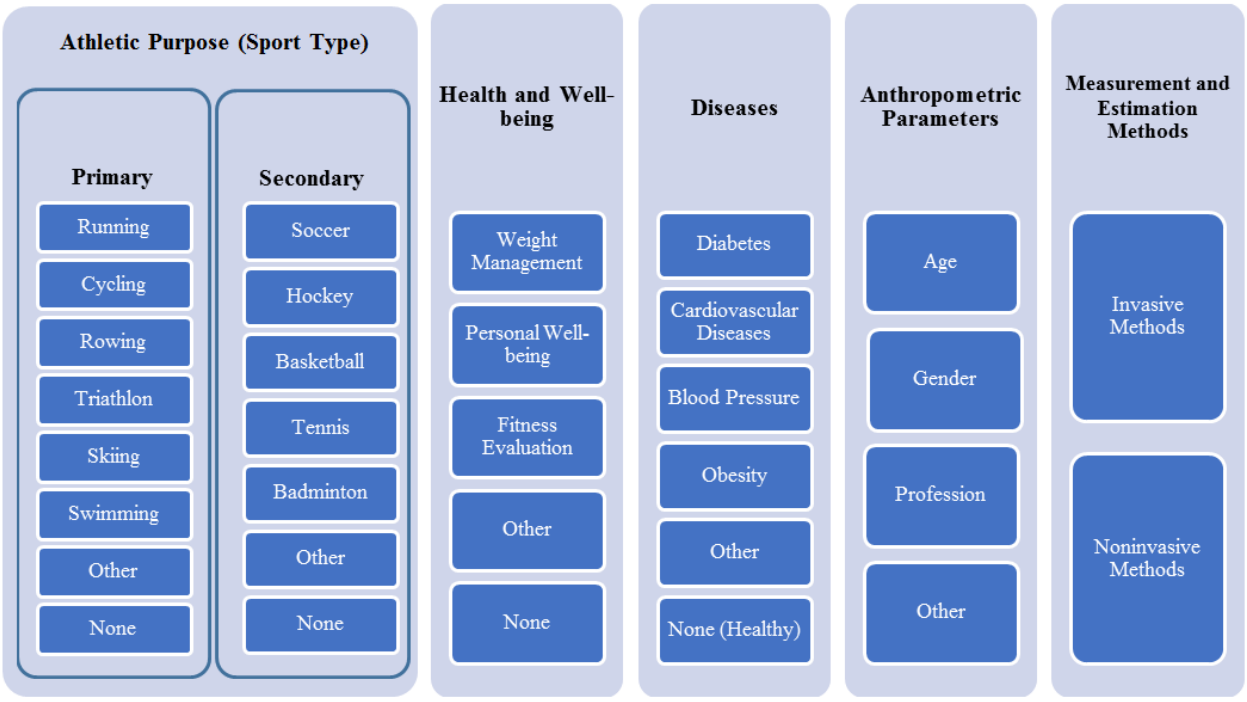


Figure 2.3: LT literature categorization

1) LT and Athletic Purpose (Sport Type)

The Athletic Purpose (Sport Type) category covers the literature discussing the utilization of LT for athletes in different types of sports. For example, literature measuring LT values in sports such as running [48], soccer [49], cycling [50], and swimming [51] is included in this category. Athletic Purpose (Sport Type) is divided into two subcategories which we refer to as Primary and Secondary, based on the importance of LT estimation in the sport in question. Sports that are point-to-point such as running, cycling or rowing are part of the Primary sport category, while other sports such as soccer, hockey or basketball are part of the Secondary category. The "Other" class includes literature examining LT in sports other than those defined specifically in this category, such as professional sports teams

in futsal, handball and basketball, discussed in [52]. The "None" class includes papers discussing LT utilization but not for an athletic purpose.

In research on recreational running [48], a heart rate variability data set is used to detect LT values. Authors in [50] evaluate a new wearable lactate threshold sensor to be used for highly trained cyclists. Authors in [51] discuss the characteristics influencing the estimation of LT values in swimmers and concluded that gender, stroke and distance are the top three influencers on LT value.

2) Health and Well-being

The Health and Well-being category addresses the usage of LT features with individuals who are non-athletes. Such research efforts expand the focus of utilizing LT values to accommodate a population outside the sport community. Work in this category opens doors for future research and aims to utilize LT values for personal health and well-being in general. Research designed to increase awareness of this feature's importance in the non-athletic population is promising. Researchers in [53] compare and verify the relationship between individual anaerobic threshold (IAT), individual glucose threshold (IGT), critical velocity (CV), and the velocity associated with $\dot{V}$O2max (Vmax) and with 3,000 km running performance (Vm3km) in physically active non-athlete individuals. Weight management is another area covered by this category and researchers in [54], and [55] investigate the relationship between FOR and LT.

3) Diseases

The Diseases category covers the literature discussing the utilization of LT in evaluating exercise intensity for patients with chronic disease. Diabetes [56], [57], especially type 2 diabetes (T2D) [58] and cardiovascular disease [59], [60], [61], are the diseases discussed most in the LT literature, while blood pressure [58] and obesity [55] are discussed as related diseases.

T2D patients gain special attention from the LT research community due to the strong relationship between LT and this disease as discussed in [56], where LT values are combined with glucose values to diagnose diabetes mellitus. In addition, the estimation of LT value is important in identifying proper exercise intensity for diabetes in general [57], and to predict post exercise hypotension around LT values [62]. To facilitate the LT assessment task, authors in [63] proposed glycemic threshold as an alternative method to identify AT. Authors in [64] used blood glucose and RPE to predict LT value, whereas authors in [65] used the traditional Dmax method. Authors in [57] studied LT values for T2D women, and authors in [58] provided T2D patients with recommendations for best cycling practices based on their corresponding LT values. The latter study shows that cycling above threshold is more effective for releasing nitric oxide and reducing post-exercise blood pressure. For patients with cardiovascular disease, authors in [55] tested the feasibility of using heart rate variability to determine LT value, whereas authors in [60] actually used heart rate variability to assess LT value. Authors in [59] also evaluated blood LT for well-trained patients.

4) Anthropometric Parameters

The Anthropometric Parameters category covers the literature discussing the relationship between LT and statistical data to characterize the population under investigation. Gender [66], [51], [67], [68], age [69], [54], [68], [70], and profession [71], are examples of the data considered in this LT literature. Authors in [66] assess the effect of gender differences on LT values between female and male soccer players. They asked fourteen females and thirteen males to complete an incremental test and monitored the progression of results. The results reflected differences in the distribution of anaerobic and aerobic metabolic pathways between the two groups. The study in [51] also examined the influence of gender on LT values and characteristics in professional swimmers. From the earliest experiments, researchers have also considered age as an influencing factor on LT, as shown in [69] and [70], where researchers focused on e.g., young soccer players and the elderly. Studies have also take the profession of the experiment sample into account, such as the college students in [71].

5) Measurement and Estimation Methods

The Measurement and Estimation Methods category is the most discussed in the LT literature. The methods can mainly be classified into invasive and noninvasive subcategories. Invasive methods are used most often in the literature, while noninvasive methods are discussed only slightly. Invasive methods include any method requiring blood sample collection and methods capturing other physiological features such as HR. Noninvasive methods include any method utilizing one or more techniques that require other physiological fea-

tures such as HR and RPE to estimate the LT value. The commercial devices could be invasive, such as the LA analyzer used in [70], or noninvasive, such as the Humon wearable device with MetaOx probe proposed in [72] and the wearable lactate threshold (WLT) sensor proposed in [50]. Invasive methods typically provide the most accurate results although they are costly in terms of required equipment, personal effort (coaches), test cost and discomfort conditions. Hence, several research works propose noninvasive methods to estimate the LT, arguing that the data can be captured in an accurate manner with results comparable to those obtained through invasive methods.

- **Invasive methods**

Invasive methods basically rely on blood lactate concentration to determine LT, as we see in the research in [73] and [74]. Authors in [73] aim to measure LT during a resistance exercise using four different methods in order to determine the most accurate method. Twelve men participated in the study and performed a maximal incremental test on a leg press using different intensities. The four methods included visual inspection, log-log, adjustment algorithmic method, and QLac. The LT detected by the first three methods at the same intensity and the adjustment algorithmic method provided the most accurate result. Authors in [74] used eight different methods to measure LT based on blood lactate concentration. Forty-eight well-trained male cyclists participated in the study and performed five maximal graded exercise tests in two week intervals. They aimed to identify the most reliable method for predicting the LT value of cycling performance. The results showed that the Dmax modified method was the most reliable predictor of cycling performance. LT-4MMOL is the best alternative for the Dmax modified method since it is easy

to apply. Other studies including [75], [68], [76] used Dmax in addition to other (mathematical) methods to estimate LT. Authors in [52] predicted the LT values for players of three different professional team-sport: futsal, handball, and basketball, using the maximal heart rate (HRmax) feature in conjunction with traditional LT measurement methods. Authors in [76] aimed to determine the degree of similarity of exercise intensity values at LT points and the authors in [77] proposed a novel method called reverse lactate threshold (RLT) to overcome the pitfalls of a traditional test. Authors in [70] used metabolic equivalents (MET) values corresponding to exercise intensity in combination with other LT measurements to assess LT for older people using a portable LA analyzer. In addition, authors in [78] examined the best fitting functions among six different fitting methods. In terms of least number of errors, the double phase model (two linear regression curves) and the exponential function (single linear regression curve) were the best methods. Authors in [79] also examined the agreement between 4 LT measurement methods used for professional soccer players. ML techniques are rarely utilized to assess LT. Reviewing the literature shows two papers used ML to detect LT value. One of them [80] used an invasive method in which the authors estimated LT for recreational runners.

- **Noninvasive methods**

To minimize invasive method cost and effort, several studies utilized various physiological features, tests, and sensors for estimating LT noninvasively. Heart rate [48], [52], [60], [81], and peripheral capillary oxygen saturation (spo2) [82] are examples of physiological features used in the literature to estimate LT. Researchers in [49] used two tests: an incremental maximal running test conducted using a University of Montreal Track

Test (UMTT), and a constant velocity test (CVT) to assess maximal lactate steady state (MLSS) in a noninvasive way to design endurance training for team sport players. Authors in [83] used microwave sensors for real-time noninvasive monitoring of blood lactate. The results showed that the electromagnetic wave sensors represent a good alternative for invasive blood sampling. The authors highlighted the benefits of the use of this noninvasive method in hospitals, which included reducing infection risks, and increasing measurement frequency. On the other hand, enhancing athlete training and designing prescribing effective training regimes are some of this method's benefits in sport. The literature review includes three papers [84], [85], [86] that used ML to detect LT value in a noninvasive way. A mathematical model to estimate blood lactate is proposed in [84], while a multilayer perceptron (MLP) algorithm, which is a simple neural network, is applied on the constructed datasets in [85] and [86].

2.5.2 X73-PHD Standards Compliant Systems

We conducted the literature review in [9] to survey the existing X73-PHD compliant systems, which are developed for personal use and not for clinical use, by focusing on the utilized devices in each system to determine if they were X73-PHD compliant or not. We reviewed the existing literature using three databases: Scopus, Pub Med and Web of Science. We also proposed definitions of the key concepts: PHD, compliant systems, agent, and manager, as follows:

- Personal Health Device (PHD): Any device equipped with one or multiple sensors that is able to monitor vital signals of a person's body possibly taking into consider-

ation signals from the surrounding environment such as noise.

- Compliant Systems: Personal health systems that adhere to the X73-PHD standards in manager sides by complying with the communication model (CM) of this family of standards as a minimum requirement and optionally with the domain information model (DIM) or service model (SM) as a higher level of conformance.

- Agent: In a personal health system, the device that provides the data is called an agent, which is usually the personal health device.

- Manager: In a personal health system, the device that receives the data is called a manager. This can be a personal computer, a cell phone, etc.

We applied the search query shown in [9] and investigated the results which yielded 53 publications representing the targeted compliant systems. We analyzed the results and draw several conclusions including the compliant system usage as shown below.

We found that the personal health systems can be classified based on the *location* where they are used, the *technology* used to develop them, or their *purpose*, which is determined by the intended users. Thus, we proposed the classification illustrated in Figure 2.4 showing the classes and subclasses described in the following:

1. Location: Personal health systems could be classified as one of three subclasses as follows:

    - U–Health: this subclass refers to ubiquitous personal health systems that can provide health services anytime and anywhere. Systems in [87] and [88] are examples of this subclass.

- In-Home: this subclass refers to personal health systems that are developed to be used at home for monitoring purposes. Systems in [89] and [90] are examples of this subclass.
- Mobile/Tele-Health: this subclass refers to personal health systems that are developed to monitor a patient's health remotely. Systems in [91] and [92] are examples of this subclass.

2. Technology: Personal health systems could be classified as one of two subclasses as follows:

   - Android: this subclass refers to personal health systems that are developed as Android applications or to interact with Android terminals such as smartphones or tablets. Systems in [93], [94], [95], and [96] are examples of this subclass.
   - Adopting Interoperability: This subclass refers to personal health systems that are standardized by utilizing various techniques and protocols to fulfill the interoperability condition. Some examples of these techniques are: an agent with appropriate mapping methods between manufacturer and ISO/IEEE 11073 nomenclature systems [97], a standard message generation toolkit to easily standardize existing non-standard healthcare devices [98], and an implementation model of standardization [99].

3. Purpose: We proposed this classification to show how personal health systems could be used for an individual's well-being. Thus, we suggest that compliant systems could be classified as one of three subclasses as follows. Two of the subclasses names are

derived from the linguistic meaning of "proactive" vs. "reactive" found in [100], [101] to differentiate between a healthy individual's well-being and a patient's well-being.

- Proactive Well-being: Refers to personal health systems targeting healthy individuals to track vital signs and promote healthy practices. Monitoring food intake and daily physical activity level are examples of these practices for healthy lifestyle. These systems may target adults [102] or children [103], [104]. In this review, systems developed to monitor vital signs [30] or physical activity level [105] are examples of this subclass.

- Reactive Well-being: Refers to personal health systems targeting patients to monitor their health continuously and remotely. In our review, systems developed to monitor cardiovascular diseases as in [106] and [107] are examples of this subclass.

- Hybrid Well-being: Refers to personal health systems targeting individuals vulnerable to chronic diseases such as high blood pressure and pre-diabetes conditions. Thus, this subclass includes practices from both other subclasses. Systems developed to monitor elderly people who are vulnerable to many diseases such as [89] and a system developed for postpartum well-being [108] are examples of this subclass.

2.6 Summary Remarks

The state-of-the-art of health DTs of citizens and existing models for citizens' health presented in this chapter shows the pressing need for standardized models for health digital

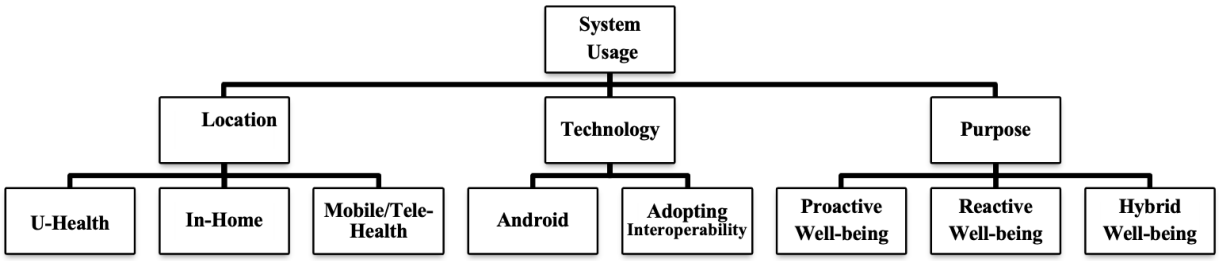


Figure 2.4: Classification of personal health systems compliant with the X73-PHD standards in the literature according to the compliant system usage, adapted from [9]

twins of citizens. Such models have the potential to pave the road for utilizing digital twins technology in either proactive or reactive health. According to the literature, the standards presented in this chapter are beneficial for developing standardized health digital twins to guarantee interoperability. Thus we utilized them in this thesis to build a standardized health DT-DNA model of citizens. As discussed in Chapter 1, this model needs to be unified and unique, contextualized, and visualizable.

The background of the biological DNA model discussed in this chapter represents the first step towards leveraging this model to design the DT-DNA to protect the unique identity of each citizen's digital twin, thereby enabling the fulfillment of the first requirement.

The results of the systematic literature review on X73-PHD standards compliant systems emphasize the role of PHDs as a rich source of personal health data. Different fitness level indicators such as LT that are determined through specific tests represent another source of citizens' health data. Both PHDs and fitness level indicators provide personal health data and contextual data, thereby enabling the fulfillment of the second requirement.

The double helix model of DNA depicted in Figure 2-1 represents a possible way to visualize the built DT-DNA, thereby enabling the fulfillment of the third requirement of the model proposed in this thesis.

Chapter 3

DT-DNA Paradigm for Modeling Health Digital Twins of Citizens

In this chapter, we discuss the analysis and design of the proposed digital twins DNA (DT-DNA) paradigm to model standardized health digital twins (DTs) for citizens living in smart cities, based on the requirements presented is Chapter 1. The research in [109] states that "a city can be defined 'smart' if it enhances the quality of living of its citizens". Our objective is to design and build a unified model of health DTs for citizens: one that handles health data from various sources and preserves the unique identity of each citizen, while building DT-DNA that can be used as an analytic tool for enhancing QoL for all citizens.

Section 3.1 presents an analysis of our DT-DNA model requirements. Section 3.2 presents the proposed DT-DNA model based on the analysis results. Section 3.3 discusses the proposed framework to build health DTs for citizens on micro, meso, and macro levels.

3.1 Mapping Between DNA & DT-DNA: Requirements Analysis

In this section we present an analysis of the biological DNA model and highlight its features which provide the building blocks for our DT-DNA model. Following this (in Section 3.1.2), we present an analysis of the influencing factors shaping smart cities modeling, in order to determine the main bases of our DT-DNA model of health DTs. Our model is based on the 'smart' city definition above. It is designed to capture and analyze health data on citizens living in smart cities who interact with, and are affected by, their environments and the various services they access.

3.1.1 DNA Model

To build effective health DTs for citizens, a mechanism is needed to handle various kinds of data collected from heterogeneous devices and applications. In nature, DNA is an example of a model that is both unified, common to all humans, and unique, distinguishing each human as an individual, and that is what motivated me to mimic DNA for modeling the health DTs of citizens.

The biological DNA model is characterized by many features including the following, which motivated us to consider using this analogy to model digital twins:

- Unified Model: All chromosomes or DNA sequences share a common structure, the double helix DNA, with four bases: A, T, C, and G. The sequential order of these bases causes the difference in the chromosomes' values.

- Unique Model: All chromosomes are identical in the same organism (all cells have an identical copy and number of DNA) but differ from one organism to another. Thus, all humans have the same number of chromosomes but each human has a unique genetic fingerprint and variance.

We propose that an analogy can be made here for modeling health DTs for citizens. Any robust digital twins model of health must be both *unified* and *unique*, similar to the way in which DNA has a common (unified) structure that captures the unique identity of each organism.

3.1.2 Anatomy of Health in Smart Cities

In order to successfully design and model health DTs for citizens, we identified three essential requirements outlined in Chapter 1. We argued that any unified model that could work as a base for building health digital twins for citizens would need to be *unified* and *unique*, *contextualized*, and *visualizable*. Using the DNA analogy as a starting point and incorporating ISO 37120 standards [33], we were able to meet the first of these requirements: design a model that is *unified* and *unique*.

Our next step involved analyzing the literature on the anatomy of cities in order to identify other influencing factors that shape health in smart cities. In our analysis of the literature, we identified humans as one of these influencing factors – citizens who perform various health tasks, and authorities who regulate health services. For example, research in [110] emphasizes that humans (citizens) in different roles and positions are considered

to be a basic influencing factor shaping health in smart cities since they are the original source of health data and the target of health services.

We also identified tasks that a citizen performs or an authority regulates as another influencing factor shaping health. Endurance level tests that physically active citizens perform to determine their fitness level and the walking tests that elderly perform for different health purposes are examples of tasks the citizens perform. Evaluating the fitness level of citizens across city is an example of health tasks that the health authority regulates. Details of tasks included in received health data must be included as part of any complete model, as all tasks in the real twin must be replicated in the corresponding digital twin.

Similarly, contextual data are considered to be an influencing factor shaping health in smart cities. For example, research in [111] states that environmental context has an influence on health in smart cities. The abundance of location-based information that citizens generate by being connected, anywhere and anytime, through smart, wearable devices [112] is an example of this type of contextual data. Indeed, spatiotemporal factor provides location-aware data is a significant contextual data [113] that is also an influencing factor on health in smart cities due to its role in defining tasks that citizens perform or authority regulate.

The literature also highlights the important role of geographical data in smart cities modeling, which can include spatial information gathered by governments, and voluntary information obtained (crowd-sourced) from various sources in smart cities [114].

Based on our analysis of the literature on the anatomy of smart cities, we identified four main components or influencing factors shaping health in smart cities:

- Humans, who represent the members of urban societies, including stakeholders and authorities. Stakeholders consist of all citizens living in smart cities, including individuals, groups, and those who care about health services, such as politicians, researchers, business leaders, planners, designers, and insurance companies. Authorities are responsible for the provision and regulation of health services in the city, including local health authorities, city councils, and city halls.

- Tasks, which represent different health tasks performed by citizens on a personal level, or tasks regulated by the city authority on group and city levels.

- Context, which represents contextual data collected from various sources as defined by [113], including "any information that can be used to characterize the situation of an entity." The sources may be hard sensors, such as those utilized in personal health devices, or soft sensors such as software applications that determine individuals' geographical location.

- Geographical data, which represents the spatial and temporal data of cities that play an essential role in the provision of health services.

Thus, aligning our analysis of the biological DNA model with the results of our smart cities literature analysis reinforces the suitability of the suggested analogy between the biological DNA model and our proposed DT-DNA model. The proposed model, which is discussed in the following section, utilizes the four influencing factors listed above to define the four bases of DT-DNA of health DTs of citizens based on ISO standards, which fulfill the listed requirements in Section 1.2. Thus, this DT-DNA model identifies each citizen's

health DT in a unique way similar to the way in which DNA captures the unique identity of each organism.

3.2 Proposed Digital Twin DNA Model

To meet the requirements listed in the introduction, we designed the proposed DT-DNA model to model health DTs for citizens living in smart cities. In keeping with our DNA analogy, we have identified four bases, A, T, G, and C, for our unified and unique DT-DNA model. We will start by describing our analogous G and C bases, followed by A and T bases. Each base is defined by a set of identifiers, as shown below, and we adopt a "fixed-location" characteristic for these identifiers, maintaining the same location for each, represented by a 2-alphabet code in the DT-DNA strand.

Geo-temporal (G base)

The G base in the proposed DT-DNA model represents the geographic and temporal information of the city, where the citizen exists at the time of data collection. Geo-temporal information plays a critical role for both real and digital twins [115]. We define this base by four main identifiers described in the following:

Continent-Country-City-Year

- ***Continent:*** This identifier represents the code of the continent where the city located. We used a 2-alphabet continent code associated with the list of ISO 3166

Country Codes [116]. The complete list of the seven continents is shown in Table 3.1. The first column of this table shows the continent name, and the second column shows the code according to ISO 3166 [117].

Table 3.1: List of all continents and the proposed code to be used in the G base

Continent	**Code**
Asia	AS
Africa	AF
Europe	EU
North America	NA
South America	SA
Oceania	OC
Antarctica	AN

- ***Country:*** This identifier represents the code of the country where the city is located. We used the complete list of ISO 3166 Country Codes [117], which corresponds to the list of top-level domains (TLDs) of all countries that is developed and maintained by the Internet Assigned Numbers Authority (IANA) [118], as shown in Appendix A, Table A.1.

- ***City:*** This identifier represents the city where the citizen exists at the time of data collection. A 2-alphabet code derived from the city name is used for this purpose. The latitude and longitude coordinates of the city could also be used here.

- ***Year:*** this identifier represents the year when the data have been collected towards

building citizen's health DT-DNA. We proposed a year code shown in Appendix A, Table A.2.

Context (C base)

The C base in the proposed DT-DNA model represents the context at the time of data collection. It is defined by any source of information surrounding the citizens, and influences the status of collected data for a particular health task. It differs according to the level at which the data are collected. On an individual level, the source of information could be hard sensors such as those utilized in personal health devices, or soft sensors such as software applications that determine individuals' geographic location (e.g., Google Maps) [119]. On group and city levels, the context is defined as the Environmental Context (EC) of the city according to the ISO 37120 standards [120], which we adopt in the proof-of-concept described in Chapter 5. Environmental Context is determined by the city's: surface (S), either flat (F) or hilly (H); location (L), either sea (S) or inland (N), and kind of rock (R) upon which it is built, which can be island (I), peninsula (P), valley (V) or deep and hard rock (D). Table 3.2 lists the DT-DNA model codes that we proposed for these criteria. We used a 4-alphabet code for each criterion, with the first two characters representing the initials of the context type.

Table 3.2: Environmental Context (EC) criteria according to ISO 37120 [120], to be used in the DT-DNA model

Environmental Context (EC) Criterion	**Proposed Code**
Surface - flat	ECSF
Surface - hilly	ECSH
Location - sea	ECLE
Location - inland	ECLN
Kind of Rock it is built on - islands	ECRI
Kind of Rock it is built on - peninsulas	ECRP
Kind of Rock it is built on - valleys	ECRV
Kind of Rock it is built on - deep and hard rock	ECRD

Anthropometric/Authority (A base)

The A base in the proposed model differs according to the level in which we build the DT-DNA. When we build the DT-DNA on an individual level, the A base in the proposed DT-DNA model represents a citizen's anthropometric. Consistent with the literature in this area [121], we proposed using anthropometric data, which are an individual's physical features such as height, weight, and body mass index (BMI), and demographic data such as gender and age. We also suggested using the profile/affiliations of special groups of citizens to which the individual under investigation belongs, such as elderly [122], to facilitate future analysis. The research in [122] points to the possibility of linking a citizen's anthropometric data with their usage of certain services and applications in the city (tasks), which would

allow for the mapping of A base with T base data in the proposed model, as is described below.

In contrast, when we build the DT-DNA on group and city levels, the A base in the proposed DT-DNA model represents the city's regulatory authorities. Authority can be defined as those responsible for managing and regulating the provision of different city services, and those participating in decision-making processes. City council, city hall, and local health teams are examples of authorities who regulate health services in the city. When data are collected from citizens as a group and the intended services target them on a city level, the authority who regulates the service provision must be part of the health DT-DNA model to guarantee the reliability of the services provided to citizens, and the data received from citizens through the digital twin.

We suggest using the government type of the country as well as the municipality or city council to represent this A base in the proposed model. Hence, we define A base code as follows:

A base = Anthropometric: ***Physical Features/Age/Gender-Demographic***

A base = Authority: ***Government Type-Municipality/City Council***

- ***Government Type*** This identifier represents the system of governance in the country. We proposed a 2-alphabet code for each government type similar to the internationally-recognized top-level domains (TLDs) of countries worldwide. We referred to the list of basic types of government and their definitions in the World Factbook [123], which is shown in Appendix A, Table A.3. The first column in this table shows the government type, as mentioned in [123], and the second column shows

the code we proposed for each type. According to [123], more than one government type applies for some countries, such that this identifier could be represented by one, or multiple, 2-alphabet code(s). To maintain data integrity, we added the 2-alphabet code "NN" (representing null) to country codes that have one government type, when comparing with countries that have multiple government types (i.e., those with multiple 2-alphabet codes) to guarantee the fixed-location characteristic in the proposed DT-DNA model, as discussed above.

- ***Municipality/City Council*** This identifier represents the organization responsible for managing health services in the city. Municipality (MP), City Council (CC), or City Hall (CH) are examples of 2-alphabet codes for Authority in the city.

Task (T base)

The T base in the proposed DT-DNA model represents different tasks included in the provided health data that are required to model citizens' health DT-DNA on different levels. Similar to A base, the T base identifiers differ according to the level in which we build the DT-DNA. When we build the DT-DNA on an individual level, the T base represents the task(s) through which the data have been collected. Examples of these tasks are presented in the proof-of-concept in Chapter 4. When we build the DT-DNA on group and city levels, the T base in the proposed DT-DNA model represents different tasks of the health service to be modeled. Since we adopted ISO 37120 [120], different indicators under the ISO-standardized health services represent the tasks required to model the DT-DNA of a particular group of citizens at the city level. Using ISO 37120-standard indicators,

our proposed code format (depicted below) requires a core (C) or supportive (S) indicator type, followed by a service initial, followed by an indicator's clause initials.

Indicator Type-Service Initial-Indicator's Clause Initials

For example, the proposed code for a service with the indicator type (C) for *core*, and the service initial (H) for *health*, and the indicator's clause initials (ALE) for *average life expectancy* would be CHALE, as depicted in Figure 3.1:

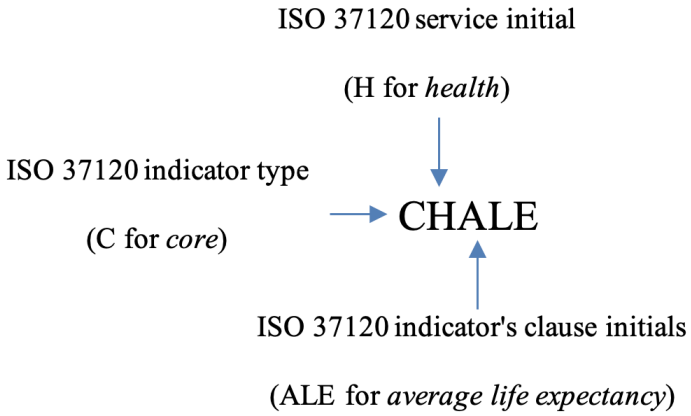


Figure 3.1: T base - sample code

3.3 DT-DNA-Based Framework to Build Health Digital Twin of Citizens

To model the DNA of citizens' health DTs, we needed to take various health data sources into account. Following an analysis of the literature [124], [125], we found that health data may represent a single citizen, a community of citizens such as a community of

physically active citizens or a diabetic community, or citizens from across the city. Thus, the proposed framework models citizens' health digital twins at three different levels. Figure 3.2 illustrates the proposed DT-DNA-based framework to build health DTs for citizens based on our DT-DNA model.

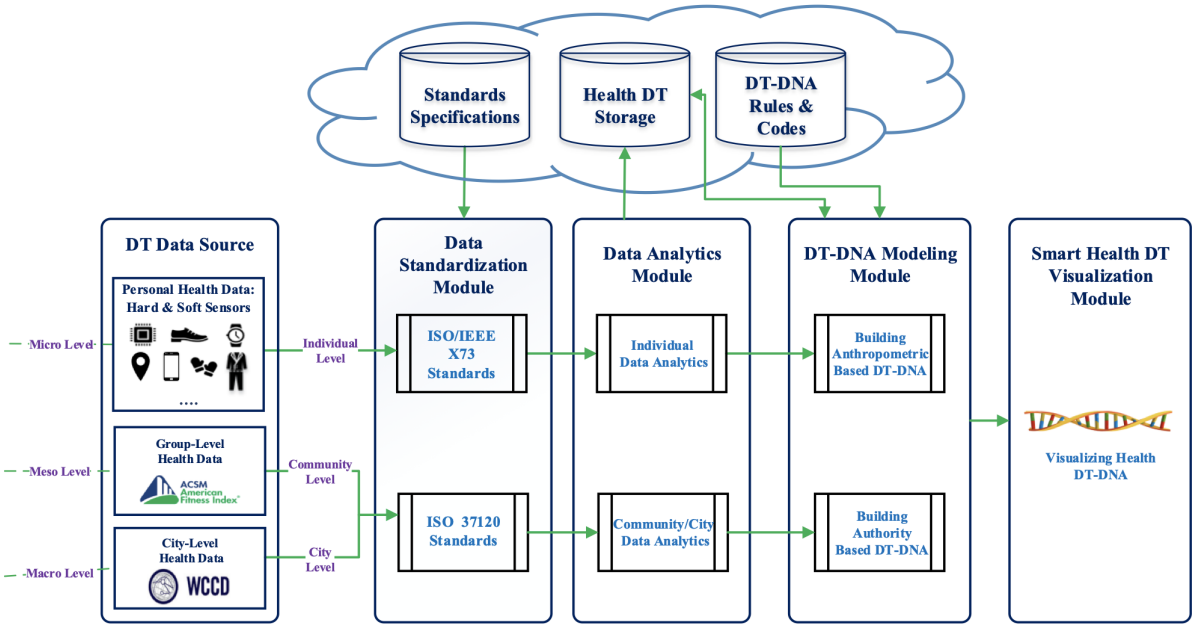


Figure 3.2: Proposed DT-DNA-based framework to build health digital twins of citizens

This framework consists of the following main components (as depicted in Figure 3.2):

- **DT Data Source:** This component represents various sources of health data for citizens' health DTs, collected over time. The proposed framework addresses modeling citizens' health DTs at three different levels: micro, meso and macro. At the micro level, a health DT represents the health of a single citizen utilizing health data collected on an individual level from hard and soft sensors such as physical activity

trackers [126] and location tracking apps e.g. Google Maps[119] respectively. At the meso level, a health DT represents health of a particular community or group of citizens that utilizes health data collected from this community or group of citizens (e.g., physically active citizens , diabetics, etc.). Data from a Fitness Index developed by American College of Sport Medicines [10], which represents the fitness level of the city based on data collected from a group of citizens, is an example of data at the meso level. At the meso level, collected data reflect responses from a group of citizens, through traditional methods of data gathering such as interviews, questionnaires and surveys. At the macro level, a health DT represents citizens' health at a city-wide level, using data collected from both citizens and health services.

- **Data Standardization Module:** This component is responsible for standardizing the received data according to internationally recognized standards such as ISO standards, to guarantee the interoperability of the developed citizens' health DTs. For data collected on all three levels of the DT Data Source, we emphasize the importance of standardization to develop a unified general model of health DTs of citizens as individuals, in communities, or city-wide, that guarantees interoperability. This is due to the fact that each citizen has his/her own needs and preferences for collecting and tracking personal health data, including the types of wearable or personal health devices used. Similarly, each city has its own way of collecting and handling data provided by citizens and used by different health services. Thus, we designed the proposed framework to handle data coming from various sources, regardless of whether it arrives in standardized or non-standardized format. Standardization on an individual level is accomplished according to ISO/IEEE X73 standards [127], which

guarantee interoperability of health devices designed for personal use [9]. In the Data Standardization Module, the received health data are assigned to the appropriate standards using standards specifications stored in the cloud. Standardization on community and city levels is accomplished according to ISO 37120 [120], which determines the set of city services along with different indicators to evaluate each city's performance and its citizens' quality of life. In this module, ISO–based wrappers for both standards are integrated to process the non-standardized data using standard specifications stored in the cloud. ISO-based wrappers allow the DT-DNA model to accommodate valuable health data that do not adhere to the adopted ISO standards but are nevertheless important aspects of a citizen's and a city's health.

- **Data Analytics Module:** Once data have been standardized in accordance with the relevant ISO, the Data Analytics Module is where analysis of standardized health data takes place. Subcomponents of this module include data analytics at the "individual" and "community/city" levels. For example, in the individual subcomponent, analyses may involve the use of athlete and non-athlete citizens' performance test result data to estimate fitness levels, or the use of senior citizens' gait speed data during over-ground walking vs. stair-climbing to explore fall prevention strategies. Similarly, in the community /city subcomponent, analyses of e.g., health services data at the community level, or city-wide fitness data may be conducted. Following analysis, results are sent to the DT-DNA Modeling Module for DT-DNA sequence building. The Data Analytics Module also allows for analysis results to be placed in Health DT Storage in the cloud until they are ready for sequence building in the next module.

- **DT-DNA Modeling Module:** This component is responsible for building the DNA of the health DTs, based on the proposed model in Section 3.2. When the DT-DNA Modeling Module receives data from the Data Analytics Module directly, or retrieves data from Health DT Storage in the cloud, the module begins processing based on the corresponding subcomponent of the Data Analytics Module. If data represent an individual level, the module starts processing under the "Building Anthropometric-Based DT-DNA" subcomponent. The data of the task during which the health data have been collected are assigned to the T base in the DNA of the health DT for this citizen. Physical performance test to estimate fitness level and over-ground walking vs. stair-climbing experiment are examples of these tasks. The anthropometric data of this citizen are assigned to the A base, which are the body individual's physical features such as height and weight, in addition to other demographics and health conditions data. Other data (G base and C base) are determined according to the city represented by the received data, and using the rules and codes retrieved from the DT-DNA Rules and Codes storage (Figure 3.2) in the cloud. If, in contrast, the health data represents group and city levels, the module starts processing under the "Building Authority-Based DT-DNA" subcomponent, where each indicator under the health service in ISO 37120 will be processed following the Task (T base) naming conventions discussed in Section 3.2, and mapping each indicator to its numerical value. Other data (bases A, G, and C) are determined according to the city represented by the received data, and using the rules and codes retrieved from the DT-DNA Rules and Codes storage in the cloud, except that here, A base represents the Authority data. Then, numeric values under all bases are coded to alphabetical

values using a 2-alphabet code for numbers from 0 to 100, which is also stored in the DT-DNA Rules and Codes storage in the cloud. This code is shown in Appendix A, Table A.4. The built DT-DNA sequence is then stored in Health DT Storage (Figure 3.2) for future processing. For example, under the "Building Authority-Based DT-DNA" subcomponent, some assessment algorithms of e.g., health city services or annual fitness levels apply through the comparison of built DT-DNAs. DNA alignment tools such as MatGAT (Matrix Global Alignment Tool) [128] can be used with generated DNA sequences for comparison purposes. The comparison is conducted on the newly-generated DT-DNAs or on the stored ones which are retrieved from the Health DT Storage (Figure 3.2). The algorithm we propose in Chapter 5, called: "*Which city has better services towards enhancing QoL*?" is an example of suggested assessment algorithms in this module.

- **Smart Health DT Visualization Module:** This component is responsible for visualizing the developed health DTs through graphic representations (charts, figures, etc.) of the generated DNA. These visuals provide feedback to authorities and stakeholders and can help them to gain insight into the health of their citizens on a personal, group, or city-wide level, with the goal of enhancing their citizens' overall health and quality of life.

Chapter 4

DT-DNA Paradigm at the Micro Level

In this chapter, we present our initial work on building health DTs for citizens using the DT-DNA-based paradigm. We consider physical wellness in building health DTs and focus on two categories of citizens: physically active citizens, and senior citizens. We conduct two experiments in collaboration with exercise physiologists at the Peak Centre for Human Performance in Ottawa, and researchers at the Faculty of Health Sciences at the University of Ottawa.

In the sections that follow, we present the proof-of-concept of our DT-DNA-based framework for building citizens' health DTs. We begin with two case studies of DT-DNAs, designed on an Individual level using ISO/IEEE X73 standards. Case Study 1 (Section 4.1) uses fitness assessment tests to estimate lactate thresholds (LTs) in physically active citizens. Case Study 2 (Section 4.2) examines gait speed in senior citizens during

performance of single vs. dual tasks.

4.1 Case Study 1: Estimating LT Towards Building DT-DNA of Physically Active Citizens

In Case Study 1, we explain our work on building DT-DNAs for physically active citizens using an examination of Lactate Threshold (LT). We present our proposed DT-DNA framework in Figure 4.1 (adapted from Figure 3.2) to illustrate this case study. Before presenting implementation details, we provide a brief discussion on the importance of LT as a fitness level indicator for Individuals (Section 4.1.1), followed by a discussion on the existing methods used to estimate LT (Section 4.1.2). We then provide a detailed description of our method to estimate LT non-invasively under the DT-DNA-based framework (Section 4.1.3). We conclude by showing the results of non-invasive lactate threshold estimation using machine learning, followed by the visualization of DT-DNA for a sample of subjects.

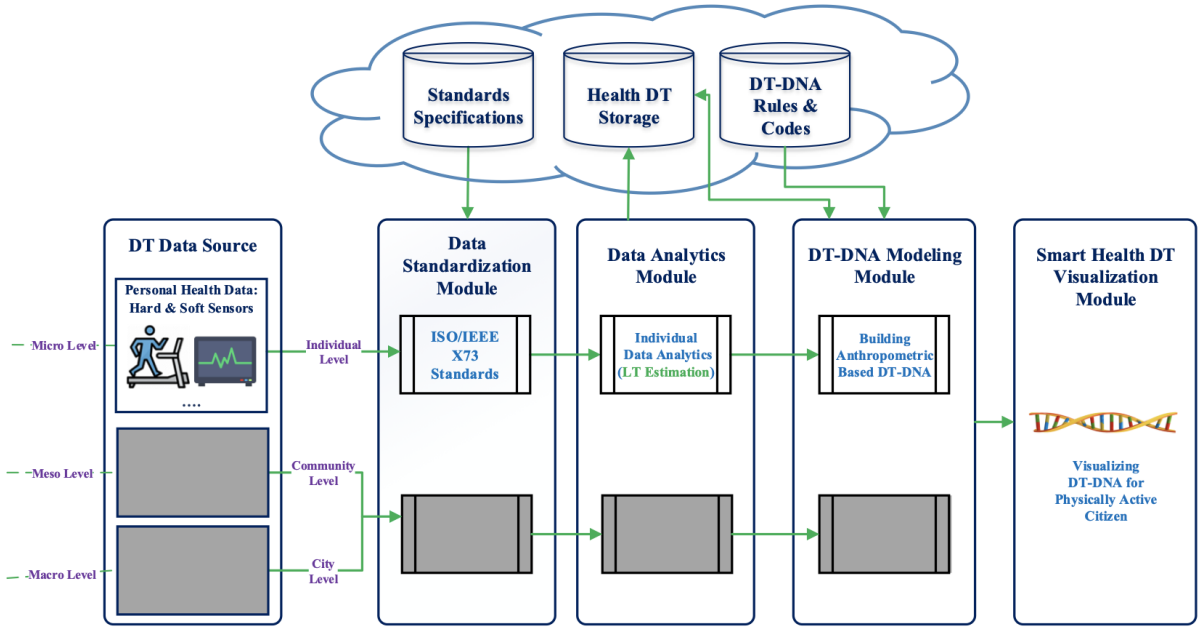


Figure 4.1: Case Study 1: Estimating LT towards Building DT-DNA of Physically Active Citizens (adapted from Figure 3.2; components not relevant to this case study are greyed out)

4.1.1 Importance of LT as an Indicator of Fitness Level for Individuals

Lactate formation has a critical role in health and well-being since it appears in the blood stream with any physical movement [129]. It is produced and consumed by the skeletal muscles and when the produced amount surpasses the consumed amount, lactate threshold (LT) occurs [129]. LT plays a fundamental role in identifying an individual's fitness level and is often used as a guide by coaches and trainers to develop fitness plans for athletes and non-athletes to improve performance. Further, LT is an indicator of an individual's endurance level, and hence used by coaches to predict exercise performance [130]. For

athletes, it aids in determining overall performance level in a specific sport such as running or swimming, and can facilitate designing the recommended training routine [39]. For non-athletes, it provides individuals with information about their state of health and well-being, and may encourage them to be more physically active [86].

Given that LT is one of the accurate fitness level indicators that physically active citizens preferably know about their physical performance, we initiate building the DT-DNA for them. Thus, we estimate LT for a sample of physically active citizens, who represent the individual level in the proposed framework. Before discussing the details of the proof-of-the concept, we present an overview of the existing methods used to estimate LT.

4.1.2 Existing Methods for Estimating LT

Traditionally, estimating personal LT requires performing incremental tests in specific conditions, and includes the collection of blood samples under an expert's supervision [39]. A sports professional's expertise is required to determine an individual's correct LT value, which may create an obstacle for beginner coaches and lead to increased expenses for sports facilities. Such requirement for specialized expertise may affect the quality of provided services in sports facilities in terms of speed and cost on both sides: facility's personnel and clients. To avoid this requirement for specialized expertise, and to minimize discomfort and the associated costs, there have been attempts to develop alternative, non-invasive methods to measure LT. These attempts include a range of approaches such as the use of respiratory gas exchange [47], complex mathematical models [73], and other measuring devices built for this purpose, including a wearable lactate threshold (WLT) sensor [50]

and a Humon wearable device and MetaOx probe [72]. Many of these alternatives are expensive options in terms of required hardware (HW). With advances in machine learning (ML), utilizing ML techniques to predict LT can be a suitable alternative in terms of cost and delivered accuracy, which is what we propose in this case study.

4.1.3 LT Estimation under DT-DNA-Based Framework: Method Description

We conducted an interview with a senior physiologist at Peak Centre for Human Performance in Ottawa, and asked him about existing methods used to assess fitness level and evaluate performance in an individual. He explained that there are multiple fitness assessment tests and performance indicators used to evaluate fitness level in an individual, and that the LT test is one of the most accurate. A LT test depends mainly on blood sample collection from an individual's fingertip (similar to the way blood is sampled in diabetics), during performance of incremental running or cycling tests. However, the collection of multiple blood samples during each of the test stages may cause discomfort for test participants and, despite test accuracy, make it an unlikely option for many. In addition, determining LT value requires the expertise of coaches who have the needed experience and practice to perform the task accurately. Consequently, we discussed the possibility of automating the task, which would facilitate decision-making for coaches and minimize test discomfort conditions for individuals.

Based on information obtained during the interview, we determined that automation would enable us to model lactate threshold as a fitness level indicator in the intended DT-

DNA for physically active citizens. We began building a system to predict LT, including the constructed dataset, the ML model, and the visualized results, through three main stages:

- Data collection, discussed under DT Data Source and Standardization Module (4.1.4);
- Dataset construction and LT estimation using ML, discussed under Data Analytics Module (4.1.5);
- DT-DNA build and results visualization, discussed under DT-DNA Modeling and Visualizing Modules (4.1.6).

4.1.4 DT Data Source and Standardization Module

We used real data collected by coaches at Peak Centre for Human Performance [131] in Ottawa as a source of DT data. We standardized the collected data according to ISO/IEEE X73 by using the X73 Wrapper that we provided in the proof-of-concept developed in our lab and discussed in [132]. The data belong to different physically active subjects (athletes and non-athletes), who performed incremental running tests. All tests were carried out in accordance with relevant guidelines and regulations in a room with a temperature of 18–20°C. Each subject performed an incremental running test with 3-minute running stages. The number of achieved stages differs between subjects due to the differences in fitness level and therefore the stage where the LT occurs. Blood samples were collected from the subject's finger by a coach at the end of each incremental running stage. All subjects

were tested in the laboratory using a treadmill, heart rate (HR) monitor, and blood sample collecting kit, which adhered to the stationary training equipment and health devices standards, according to coaches consents as described in [133].

4.1.5 Data Analytics Module

Dataset construction

We used the incremental running test data collected at Peak Centre to construct the dataset for estimating LT non-invasively. Each record in the constructed dataset represents the results of an incremental running test performed by a specific subject and evaluated manually by expert coaches. The constructed dataset contains data of 100 subjects: 51 male and 49 female, mean age, 35.88 years. For each subject, the *test date*, *subject code* (Subj xxx) where xxx is a sequential number, and anthropometric data (*gender*, *age*, *height* in centimeters (cm), and *weight* in pounds (lbs)) are collected, in addition to *speed* (km/h) and *heart rate* (*HR*) (bpm), which are collected during test stages, to be used as features for training the ML algorithm. The LT value for each subject, which corresponds to running speed and is determined by coaches who are domain experts, is used as the ground truth score for the non-invasive LT estimation discussed below. A description of the anthropometric features of all 100 subjects is shown in Table 4.1.

Table 4.1: Anthropometric features - all subjects

Term	Minimum	Maximum	Mean (+/-) StdDev
Age	11	64	35.88 (+/-) 14.432
Height (cm)	147	188.5	170.801 (+/-) 8.498
Weight (lbs)	85.5	257	160.57 (+/-) 31.126

Non-invasive LT Estimation Using ML

We present the following procedure to estimate LT using ML in two steps:

- **Step 1: Data Pre-processing**

  We investigated the dataset manually to determine the stage at which the LT occurs, which corresponds to the running speed and ensures that the lactate value is 4 mmol/L at that stage. We found that the LT for most of the subjects (95%) occurs between the second and seventh stages of the test, distributed as follow: stage 2 (14%), stage 3 (15%), stage 4 (31%), stage 5 (23%), stage 6 (6%) and stage 7 (6%). For the remaining subjects, the LT occurs as follows: stage 8 (2%), stage 9 (2%) and stage 10 (1%). Stage 1 is the first stage after the warm-up session, so LT occurrence is not expected. This distribution reflects the diversity of our dataset which was not unexpected, given our random sample of athletes and non-athletes who performed the test for various purposes. With respect to lactate value correspondence to the fixed blood lactate concentration (4 mmol /L) in the LT stage, we found 54% of the subjects' lactate corresponded to 4 mmol/L with calibration of (+/-0.5), whereas the remaining subjects

had lactate higher (17%) or lower (29%) than 4 mmol/L (+/- 0.5). This could be used as an indicator of our subjects' fitness levels, where regular-trained (intermediate level) athletes exhibit standard lactate (4 mmol/L), well-trained (advanced level) athletes exhibit higher lactate (>4 mmol/L), and less trained (beginner level) athletes exhibit lower lactate (<4 mmol /L).

Due to the differences in fitness level and therefore the stage where the LT occurs, many missing values must be filled so the ML model can perform its function properly. In each stage, if the *speed* value is missing, it will be the previous speed +1 according to the test setting, where the speed is incremented by 1km per stage. If the *HR* (h) or lactate value (*r.f,* where: r is the real part in the value and f is the fractional part) is missing, the calculation depends on the position of the stage where the value is missing compared to the position of the stage where the LT point occurred, as shown in the equations in Figure 4.2.

Given:

- (h) is HR
- *(r.f)* is lactate value, where: r is the real part in the value and f is the fractional part
- R is the real part r in the last available lactate value A in subject's data
- *("any value"-1)* is any value in immediate previous stage to current stage
- *("any value"-2)* is any value in a stage before the immediate stage
- *("any value"+1)* is any value in immediate following stage to current stage
- *("any value"+2)* is any value in a stage after the immediate stage

- if h is missing in a stage *before* LT stage:
  $h = [(h-1) + (h+1)] / 2$ (1)

- if h is missing in a stage *after* LT stage:
  $h = (h-1) - (h-2) + (h-1)$ (2)

- if $r.f$ is missing in a stage *before* LT stage:
  $r.f = (r.f+1) - [(r.f+2)-(r.f+1)]$ (3)

- if $r.f$ is missing in a stage *after* LT stage:
  Check A and (r.f -1):

  if f *in* $A < 0.5$: $r.f = (r.f-1) + R$ *from* A (4.1)

  else: $r.f = (r.f-1) + [(R+1), R$ *from* $A]$ (4.2)

Figure 4.2: Procedure for filling missing values in the dataset

- **Step 2: LT Prediction Model and Results**

  Based on the gained experience from reviewing the literature, meeting with experts at Peak Centre, and constructing the dataset, we eliminated *test date* and *subject code* to minimize noise in the constructed dataset with the goal of improving LT estimation accuracy. We also concluded that we should use the four anthropometric features,

gender, *age*, *weight* and *height*, in the proposed model, in addition to *speed* and *HR* features from the second to the seventh stages (s2-s7), according to the occurrence of LT as discussed above in step 1. Estimating LT is a regression problem; hence, we applied a Multi-Layer Perceptron (MLP) on the dataset discussed above as depicted in Figure 4.3. According to [134], MLP is "a network of simple neurons called perceptrons. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights." This definition highlights the convenience of MLP for the nature of the problem being addressed and justifies our choice. To avoid overfitting, we chose 10-fold cross-validation for training and testing. Considering our dataset diversity, we were able to conduct several experiments to test the correlation between the LT and the anthropometric data used individually or combined when training and testing the non-invasive model, in addition to using other *speed* and *HR* features to evaluate their effect on LT prediction. We used correlation coefficient score to measure the performance of the generated models in estimating LT compared to the ground truth scores, which are the LT values as determined by the expert coaches in the invasive tests. The results are shown in Table 4.2.

The highest correlation coefficient score using single anthropometric data is achieved when the *height* feature is used, resulting in a correlation coefficient of 0.7381, which suggests a positive correlation between *height* and LT. In dual combinations, the pair (*age* and *gender*) provided the best correlation coefficient of 0.8135, whereas in the triple combinations, (*age* and *height* and *weight*) provided the best correlation coefficient of 0.7901, compared to other pair and triple combination correlation scores. Implementing the experiment including all anthropometric features in addition to the *speed* and *HR*

features from the second to the seventh stages (s2-s7) provides the best correlation coefficient (0.7983) using the LT ground truth scores. Figure 4.4 illustrates the actual (invasive) vs. predicted (non-invasive) LT of 15 subjects in terms of the speed (km/h) at which subjects reach the LT point. For the last five subjects, we used a and b under the subject number to represent the results of the same subject who performed the tests on different dates. The estimated LT for each subject is added to his/her record in the dataset, which is stored in the Health DT Storage (see Figure 4.1) in the cloud.

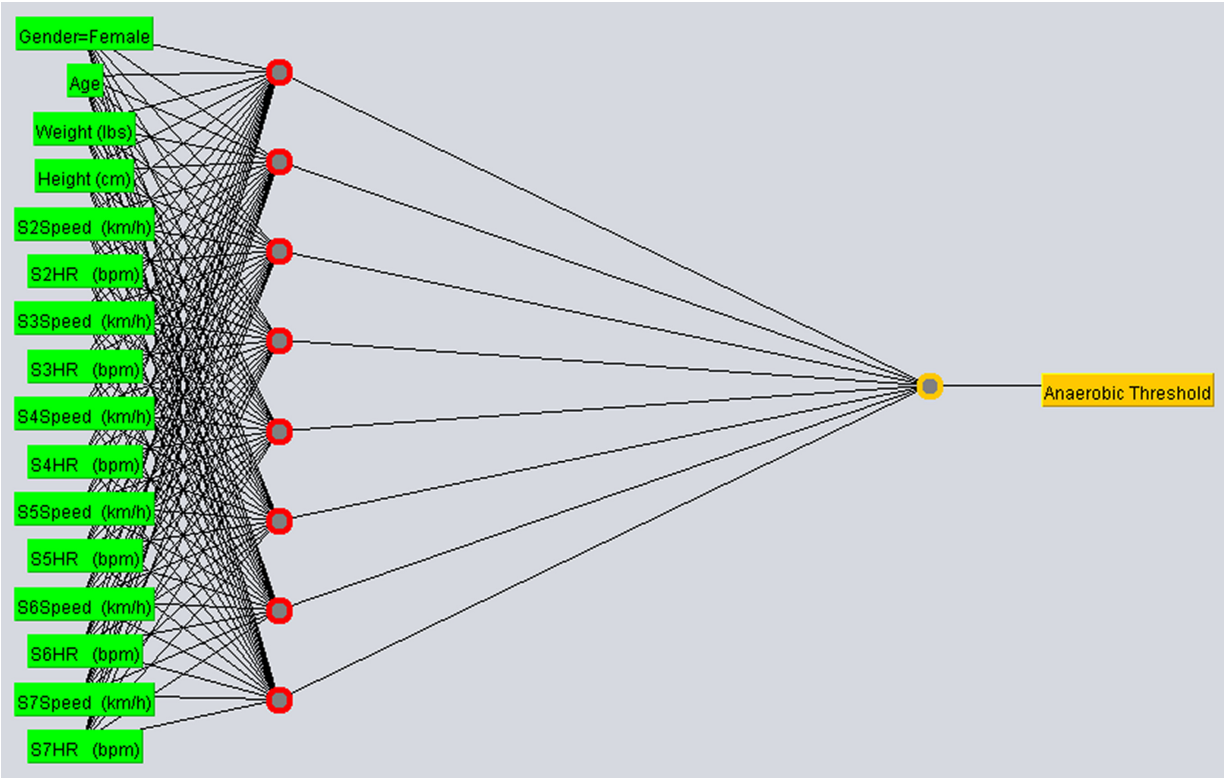


Figure 4.3: Proposed model for non-invasive LT estimation using MLP (adapted from [86])

Table 4.2: Achieved correlation coefficient by applying MLP on different anthropometric data combinations

No	Age	Gender	Height	Weight	$HR_{(s_2\text{-}s_7)}$	Speed $_{(s_2\text{-}s_7)}$	Correlation
1	×	×	×	×	✓	✓	0.6801
2	✓	×	×	×	✓	✓	0.6666
3	×	✓	×	×	✓	✓	0.6204
4	×	×	✓	×	✓	✓	0.7318
5	×	×	×	✓	✓	✓	0.6717
6	✓	✓	×	×	✓	✓	0.8135
7	✓	×	✓	×	✓	✓	0.0806
8	✓	×	×	✓	✓	✓	0.803
9	×	✓	✓	×	✓	✓	0.6565
10	×	✓	×	✓	✓	✓	0.7569
11	×	×	✓	✓	✓	✓	0.6906
12	✓	✓	✓	×	✓	✓	0.6029
13	✓	✓	×	✓	✓	✓	0.7532
14	✓	×	✓	✓	✓	✓	0.7901
15	×	✓	✓	✓	✓	✓	0.6548
16	✓	✓	✓	✓	✓	✓	0.7983

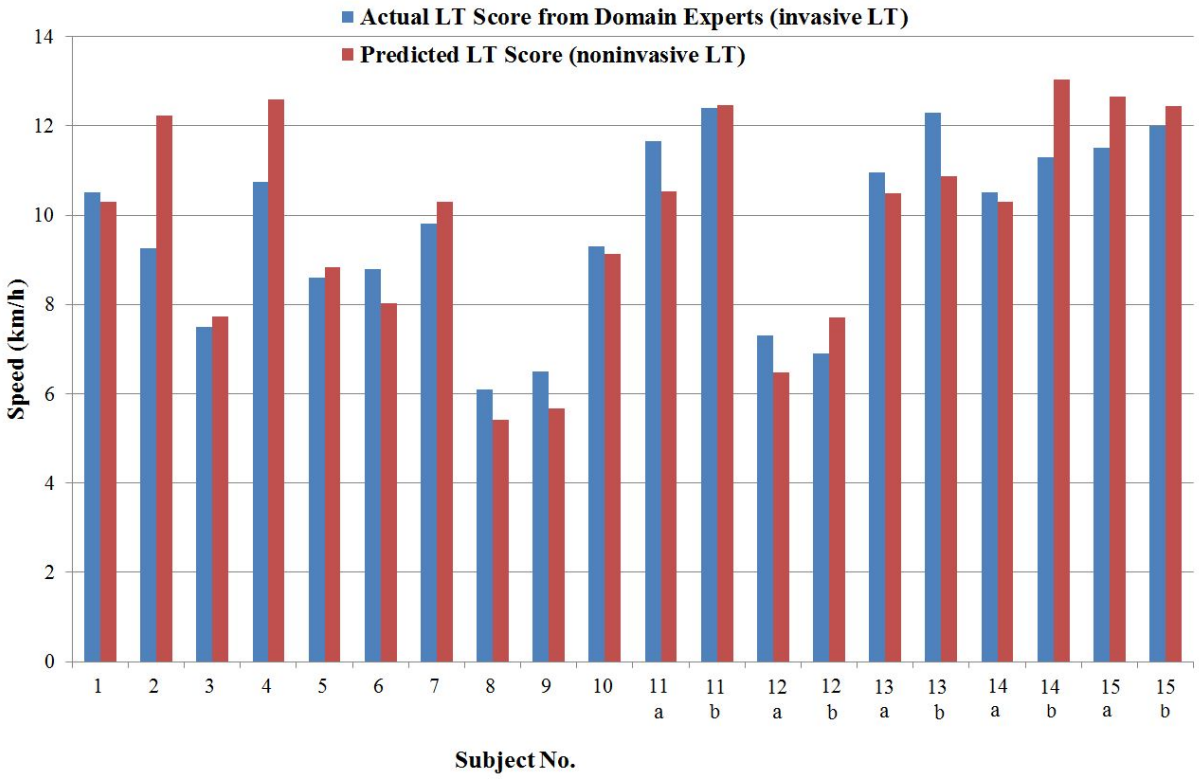


Figure 4.4: Actual vs. predicted LT scores for 15 subjects of the experimental sample; Note: a and b for the last five subjects show results of tests performed on different dates

4.1.6 DT-DNA Modeling and Visualizing Modules

DT-DNA Modeling

To build DT-DNA based on the model we proposed in Chapter 3, that shows both actual and predicted LT points of physically active subjects in this experiment, we retrieved subject data from the Health DT Storage in the cloud, along with DT-DNA Rules and Codes, in order to define DT-DNA bases as follows:

- **Geo-temporal (G base)** in this experiment is defined by the four identifiers, Continent-Country-City-Year. Given that the full date of the LT test is in the form of (Day-Month-Year), it is added under the fourth identifier instead of (Year), using the codes we proposed in Appendix A. This is because some of the subjects performed the test multiple times during the year. Thus, this base is defined as: NA-CA-OT-(LT Test Date), where NA stands for North America, CA stands for Canada, and OT stands for Ottawa. For the test date, since the day and month make a difference in this case study, we used the proposed code for numerical values and preceded them by (D) for the day value and (M) for the month value in order to build the DT-DNA code for each subject. Thus, the values for days range between DAA and DBH, which correspond to the number of days per month between 1 and 31. Similarly, the values for months range between MAA and MAM, which correspond to the months from January to December (1-12). The codes we proposed are listed in Appendix A, Table A.5.

- **Context (C base)** in this experiment is defined by the surrounding context and the hardware used while performing the LT test. As mentioned in Section 4.1.4, Room temperature ranges between 18-20°C and the hardware used includes a treadmill, heart rate monitor, and blood sample kit. Again, since these data are specific for this case study, we used the proposed code for numerical values and preceded them by (RT) for room temperature value, and (HW) for hardware. We used the average for room temperature, so the code for an average room temperature of 19°C is RTAV, since AV is the proposed code for the number 19. For the hardware, we assigned a number value to each, where 1=treadmill, 2=heart rate monitor, and 3=blood

sample kit. The codes for hardware are defined as HWAA for the treadmill, HWAB for the heart rate monitor, and HWAC for the blood sample kit, since AA, AB, and AC are the proposed codes for the numbers 1, 2 and 3 respectively.

- **Anthropometric (A base)** in this experiment is defined by the anthropometric data for each subject. They are *gender*, *age*, *weight* and *height* as discussed in Section 4.1.5. For *gender*, we used (GF) for females and (GM) for males. We used the numerical values representing the identifiers *age*, *weight* in kilograms (kg), and *height* in inches (in), for the remainder of the proposed code.

- **Task (T base)** in this experiment is the task of predicting the LT. It is defined by the actual LT point determined by the domain expert, and the LT predicted by the ML model discussed above (4.1.5). We also include the stage number at which the LT point occurs because it is an important comparison factor in case of future LT tests.

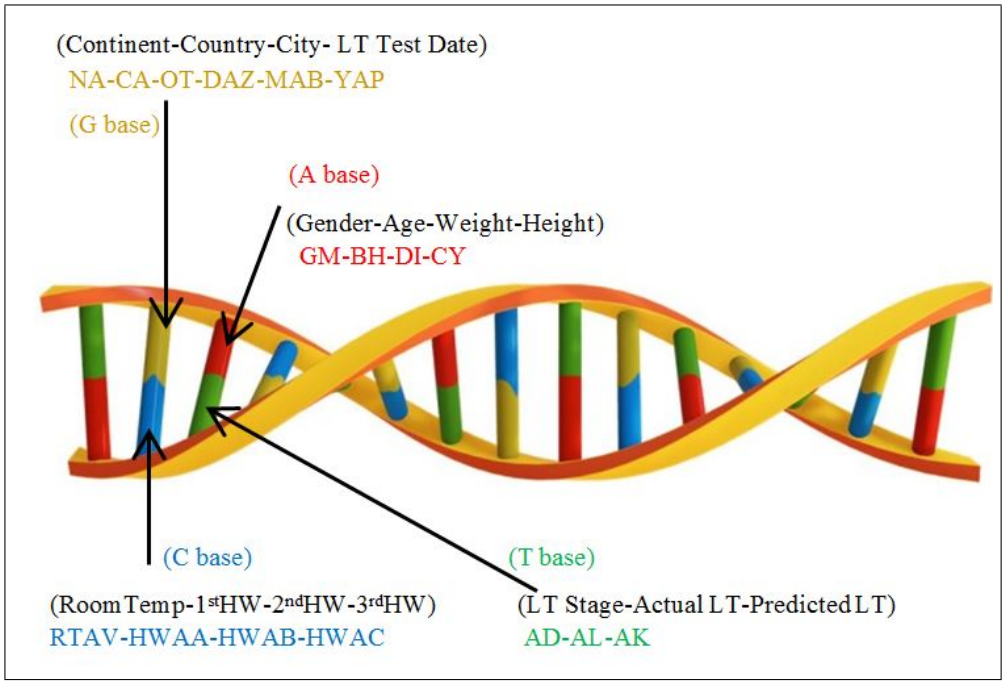
Figure 4.5 visualizes the built DT-DNA for two subjects (Subject 1 and Subject 2) of the experimental sample depicted in Figure 4.4. The visualizations of the built DT-DNAs for all subjects are shown in Appendix B. The built DT-DNA for each subject is shown using the proposed code in base values of the DT-DNA as discussed above. Thus, the built DT-DNA, for example for subject no.1, shows that he performed the incremental running test under the context conditions discussed above as shown in G and C bases respectively. It also shows he is 51 years old, weighs 78 kg and 68 in height as shown in A base. It also shows he reached the LT at the fourth stage of the test. At that point, his actual speed was 11 km/h and predicted speed was 10 km/h as shown in the code

under A base. This visual representation of the DT-DNA provides a description of the results of health task that this citizen performed on that day including the task's context in addition to his anthropometric data. Thus, as per the third requirement, we show that the model is visualizable and this may facilitate the understanding of the results by citizens, stakeholders and regulatory authorities who care about citizens' health. However, we still need to evaluate the usability of this visualization as discussed in the next section.

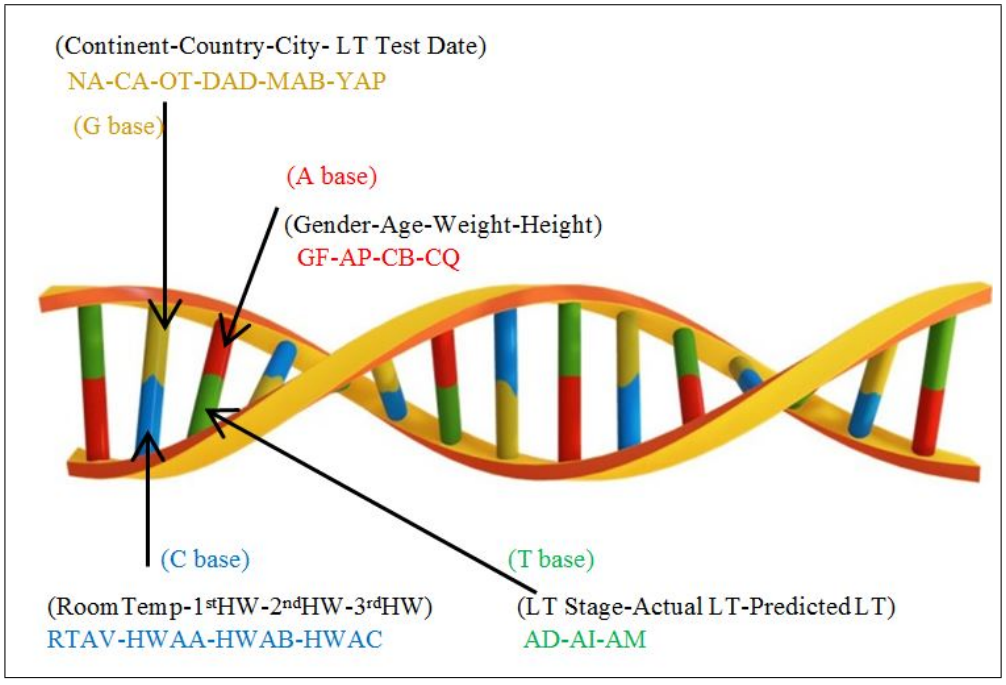
4.1.7 Case Study Limitations

The dataset used in this case study is considered small in size in that it only applies one ML algorithm. Thus, we plan to expand the dataset by adding more test data for various subjects. We also plan to apply more ML algorithms to develop different ML models for training and testing, to minimize the number of input features with the goal of building a general non-invasive model to predict LT. It would be interesting to further investigate the influencing factors on LT through the prediction model results and compare them with experts' knowledge.

The visualization results need to be evaluated by the end users through a usability study to check if they are more or less intuitive for a regular stakeholder, when compared to simple graphs such as that presented in Figure 4.4. We plan to expand the proposed code for numerical values to include fractions, since the current implementation uses rounding.



Subject 1



Subject 2

Figure 4.5: DT-DNA visualization for Subject 1 and Subject 2 of the experimental sample

4.2 Case Study 2: Measuring Gait Speed Towards Building the DT-DNA of Senior Citizens for Fall Prevention

In this case study, we explain our work on building DT-DNA of senior citizens for fall prevention. We focus on fall prevention because, according to the World Health Organization [135], falls rank as the second leading cause of accidental or unintentional injury deaths globally, and citizens who are older than 65 years old experience the highest number of fatal falls. We start with measuring gait speed since it is one of the fall predictors in older adults [136]. We present our proposed DT-DNA framework in Figure 4.6 (adapted from Figure 3.2) to illustrate this case study. Before presenting the implementation details, we provide a brief discussion on the importance of gait speed to assess and prevent fall risk, and highlight how personal health devices can contribute positively in this regard (Section 4.2.1). We then provide a detailed description of standardizing a smart shoe insole (SI), the personal health device we used for collecting data, based on ISO/IEEE 11073 standards since this wearable device is not yet standardized (Section 4.2.2). Next, we present our method to measure gait speed under our DT-DNA-based framework (Section 4.2.3), and discuss data collection from a sample of senior citizens using the standardized smart insoles under the DT Data Source and Standardization Module (Section 4.2.4). We present the initial results of measuring gait speed towards assessing fall risk under the Data Analytics Module (Section 4.2.5). We conclude this case study by modeling and visualizing the built DT-DNAs for this sample of participants (Section 4.2.6).

4.2.1 Measurement Methods and the Importance of Gait Speed for Senior Citizens

Many studies highlight the relationship between gait speed and fall risk assessment in seniors and investigate existing and new methods for measuring gait parameters including speed [136], [137], [138]. Existing methods for collecting gait parameters provide reliable data but utilize stationary or in-floor systems that are expensive and limit the measurement to constrained research environments or clinical settings [139]. Thus, various studies have proposed new metrics and instruments to measure gait parameters such as the average in-home gait speed (AIGS) metric proposed in [137], developed for mobility and fall risk assessment. Personal health devices also present promising new directions for measuring gait speed including smart shoe insoles, which are low cost and reliable wearable devices [140].

In this case study, we used custom in-shoe smart insoles (depicted in Figure 4.6) that were designed in our lab [141] and validated against the Tekscan Strideway mat system, which is used by clinicians for different purposes with successful results [140]. Smart shoe insoles are advantageous for their low cost, and are ideal for measuring stair negotiation since they collect data in real-time and can be used across many different study designs [140].

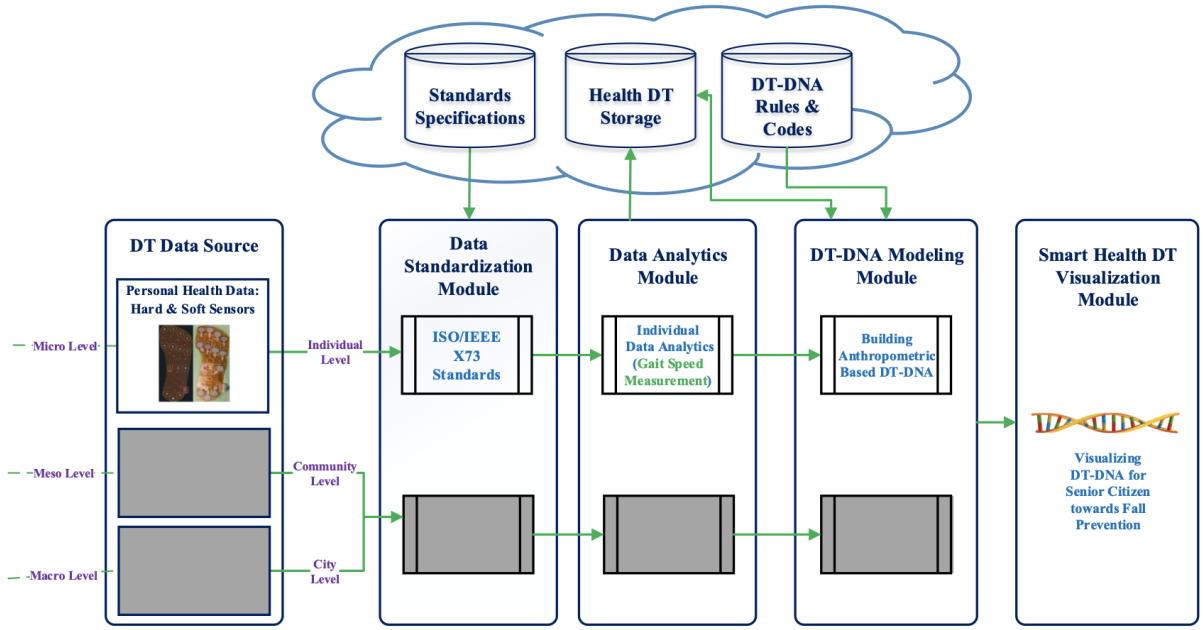


Figure 4.6: Case Study 2: Measuring gait speed towards building the DT-DNA of senior citizens for fall prevention (adapted from Figure 3.2; components not relevant to this case study are greyed out)

4.2.2 Proposed Standard for Smart Shoe Insoles (SI)

A. Design of the X73-PHD Standard for SI

The standard we proposed for smart shoe insoles (SI), based on ISO/IEEE 11073 Personal Health Devices standards, is published in [20]. We designed this standard based on the communication profile defined in IEEE 11073-20601 [32], which is used for the personal health agents and managers that are typically used outside a clinical setting such as mobile or in a person's home systems, to fulfill the specific requirements for personal health agents

and managers. We then referred to the existing standards of three personal health devices: Basic ECG (Part 10406) [142] , Pulse Oximeter (Part 10404) [143], and Weighing Scale (Part 10415) [144]. We reviewed these standards starting with the communication model (CM), followed by the domain information model (DIM), and the service model (SM). The definitions of the three parts of the standard are:

1. Domain Information Model (DIM): according to [32], the DIM illustrates the agent's information (the SI) as a set of classes. The attributes of each class represent the data that are delivered to or retrieved by the manager through objects instantiated from each class. The manager controls the agent's behavior and reports its status by utilizing the methods of each class.

2. Service Model (SM): according to [32], SM defines "the conceptual mechanisms for data exchange services. These services are mapped to messages that are exchanged between the agent and manager". Thus, the SM implementation interprets in the form of exchanged messages between the agent and the manager.

3. Communication Model (CM): according to [32], CM is used mainly to describe the communication between one or more agents and a single manager as a point-to-point connection. A connection state machine is used for this purpose, where the communication between a single agent and the manager is initiated by the connection state, followed by association, configuration, operation and disassociation states – the four states of CM. Also, CM is used to describe how to handle error conditions and convert DIM abstract data into binary messages using medical device encoding rules (MDER).

As shown in Figure 4.7, the custom in-shoe SI that we used in developing this standard has 12 pressure sensors spread across the forefoot (sensors 0 through 7), midfoot (sensor 8), and heel (sensors 9, 10, and 11).

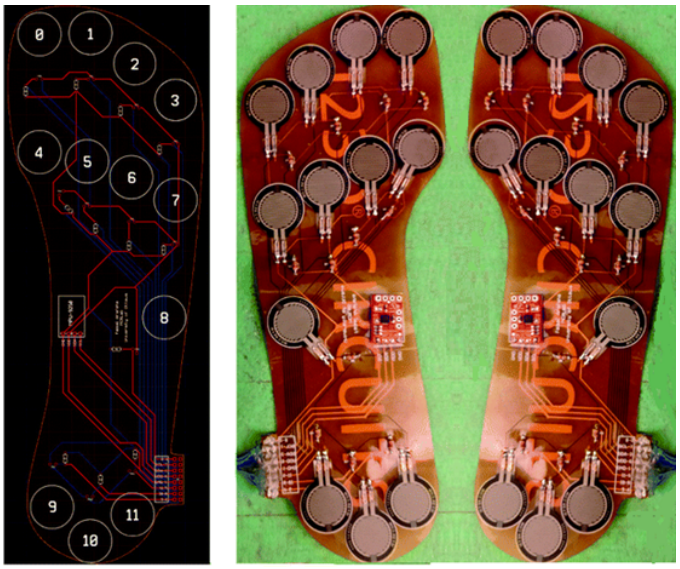


Figure 4.7: Smart shoe insole used in developing the X73-PHD standard for the SI (adapted from [20])

B. X73-PHD Implementation

We implemented the X73-PHD standard for SI, which is the agent, by defining the three main parts as follows:

1-DIM: is defined by a set of classes and objects depicted in Figure 4.8 and discussed in the following:

- Shoe Insole (MDS): this Medical Device System (MDS) object is instantiated from the MDS class and acts as a root object for the personal health device (PHD). Each

agent has only one MDS object according to the X73-PHD standards [32], which we called Shoe Insole in this suggested standard (as illustrated in Figure 4.8).

- Shoe Insole Measurements (Metric): this is the base class for all objects representing measurements and status in the SI agent, including the Pressure Sensors Waveform, Acceleration Data, and Device Status, in keeping with the definition of metric class in [32]. It contains common and shared attributes that are inherited by these classes as defined in [32].

- Pressure Sensors Waveform (RT-SA): this object is instantiated from the real-time sample array (RT-SA) class. It associates with the Shoe Insole object and represents the waveform measurement generated from different pressure sensors. It is specified as an RT-SA object because this object type represents continuous samples or waveforms, as is the case with generated waves.

- Acceleration Data (Numeration): this object is instantiated from the numeration class. It associates with the Shoe Insole object and represents the acceleration values. It is specified as a numeration object because numeration objects represent episodic measurements, as is the case with acceleration data that change occasionally.

- Device Status (Enumeration): this object is instantiated from the enumeration class. It associates with the Shoe Insole object and represents events that occurred while operating the agent. It is specified as an enumeration object because the enumeration object type represents status and/or annotation information.

- Periodic (PeriCfgScanner): this object is instantiated from the scanner class. It is used to facilitate the reporting of agent-initiated data transfers. It imports the

information about agent (SI) status from the Shoe Insole object. A periodic scanner is chosen because enabling it will send the pressure sensors waveform continuously.

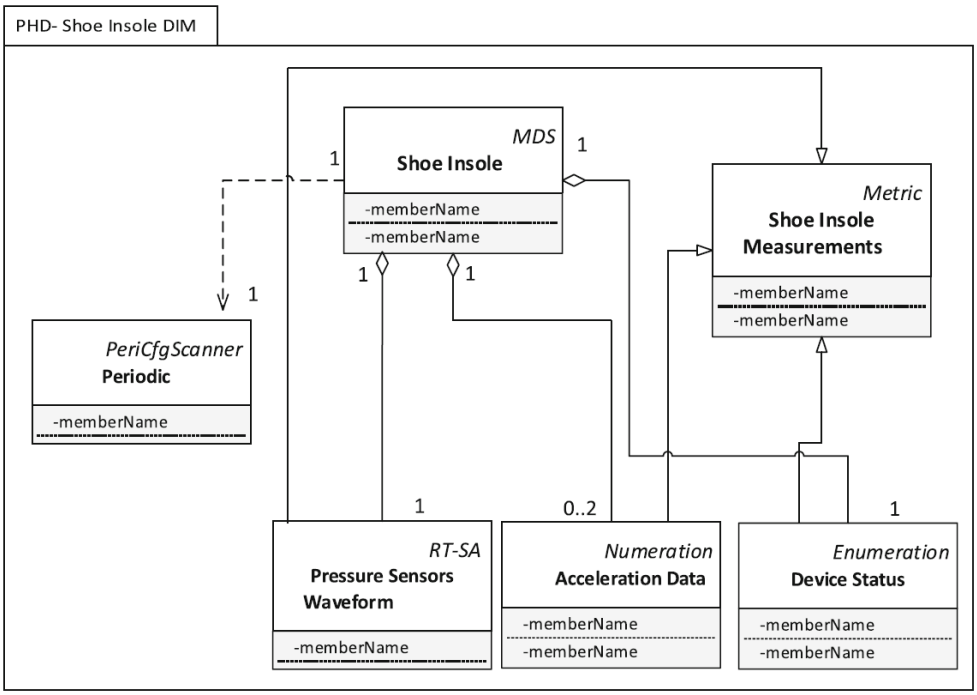


Figure 4.8: Proposed Domain Information Model (DIM) for the SI (adapted from [20])

2-SM: The 20601 protocol structure [32] differentiates between two types of messages: connection-management-related messages (association services), depicted in red in Figure 4.9 and object-related messages (object access services), depicted in green and blue in Figure 4.9. The main association messages are Association Request from the agent (A) (left) to the manager (M) (right): (A→M), Association Response (M→A), Disassociation Request (A→M), and Disassociation Response (M→A). In the event of a failure on either the agent side or the manager side, an Abort message terminates the association imme-

diately, without response. Figure 4.9, which depicts the current implementation of the proposed X73-PHD standard for the SI, illustrates these messages in red.

The data access services are used to exchange DIM data between the agent and the manager. GET, SET, ACTION and Event Report are the fundamental services in each SM. As stated in [32]:

- GET service is "used by the manager to retrieve the values of the agent MDS object". In our implementation shown in Figure 4.9, the GET message and its response are illustrated in blue.
- SET service is "used by the manager to set values of attributes of the agent's object." (This service is not needed in the current implementation.)
- Event Report service is "used by the agent to send configuration updates and measurement data to the manager". The periodic object is responsible for this service in the proposed standard for the SI. In our implementation shown in Figure 4.9, the Configuration message, the Configuration Response, the Data Report message, and the Data Report Response are illustrated in green.
- ACTION service is "used by the manager to invoke actions (or methods) supported by the agent." (This service is not needed in the current implementation.)

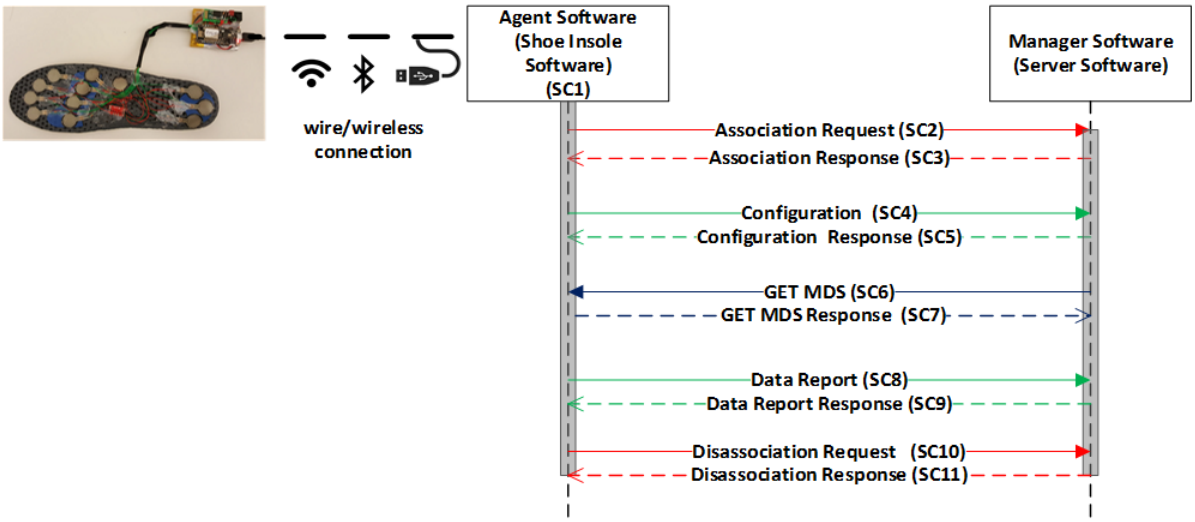


Figure 4.9: Proposed Service Model and Communication Model for the SI (adapted from [20])

3-CM: Based on the specifications in the X73-PHD standard - Part 20601 [32], we implemented the proposed standard for the SI as a set of classes that represents the CM's four states. We refer to these as X73-standard classes (SC1 to SC11, shown in Figure 4.9), in addition to other supportive classes. Figure 4.10 shows the classification of classes in the standard implementation. The X73-standard classes in the current implementation include:

- Shoe Insole: is the main class that contains all class attributes in addition to the main methods, such as methods that handle the received instructions from users. This class is illustrated by (SC1) in Figure 4.9.
- Request Classes: are the classes used to send requests from the agent (SI) or the manager. As illustrated in Figure 4.9, SC2 (Association Request), SC4 (Configuration

Request), SC8 (Data Report), and SC10 (Disassociation Request) are the classes used by the agent to send requests to the manager. SC6 is the class used by the manager to request the SI information. In the current implementation, these classes contain the attributes with their values and act as data storage because they do not contain methods.

- Response Classes: are the classes used to respond to the requests sent by the agent (SI) or the manager. As illustrated in Figure 4.9, SC3 (Association Response), SC5 (Configuration Response), SC9 (Data Report Response), and SC11 (Disassociation Response) are the classes used by the manager to respond to the requests sent by the agent. SC7 is the class used by the agent to respond to the requests sent by the manager. These response classes cause assigning values in the agent.

- Other classes are used to facilitate the implementation of the SI standard. In the current implementation, these four classes include:

  - Sensors Reading: this class is used to get data from sensors, so as to be ready for manager requests.
  - Sensors Data: this class is used to save the read data from sensors.
  - Request Methods: this class contains all request methods to be accessed by request classes.
  - TCP/IP Connection: this class is used to send data from the agent (SI) to the manager, to guarantee the reliability of data transmission.

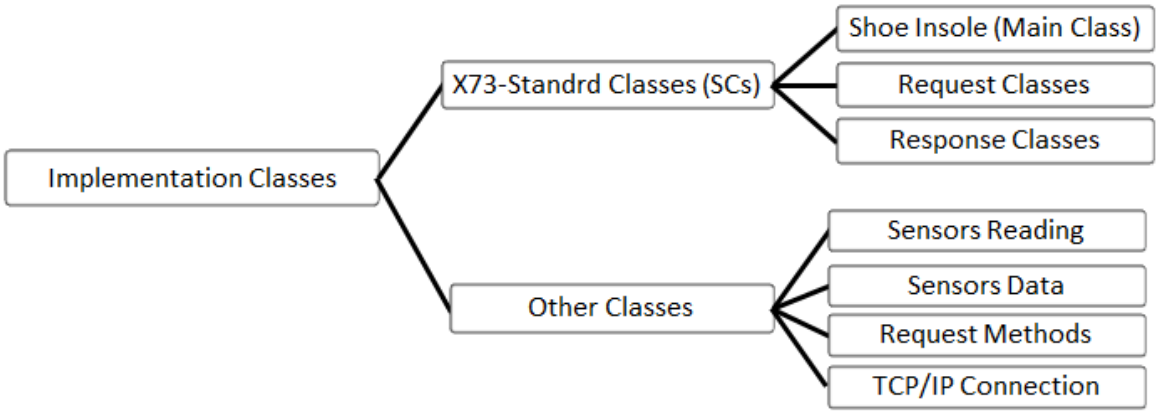


Figure 4.10: Classification of classes in the standard implementation for the SI (adapted from [20])

4.2.3 Measuring Gait Speed under the DT-DNA-Based Framework: Method Description

We conducted this case study in collaboration with the Interdisciplinary School of Health Sciences at the University of Ottawa. The study's aim was to determine whether older adults showed a performance effect while performing single vs. dual-tasks during walking and stair climbing [145]. Dual-tasks consist of two tasks, such as a motor and a cognitive task, that an individual performs simultaneously. Dual-tasks are used to determine an individual's ability to manage the demands of each task [146]. A single task consists of either a motor task, such as walking or climbing the stairs, or a cognitive task, such as a verbal response to a "peep" sound, such as word "Top" in our experiment. For data collection, the research team of the Interdisciplinary School of Health Sciences utilized the functional near-infrared spectroscopy (fNIRS) while using the SI. They draw their

conclusion based on brain signals while we draw our conclusion based on shoe insole signals. The study is designed as multiple (12) blocks with five different conditions, including: single cognitive (SC), single motor up (SMup), single motor down (SMdown), dual-task up (DTup) and dual-task down (DTdown). Several devices were used in this case study, as shown in Figure 4.11. For our purposes, we focused on motor tasks only, which represent 8 blocks out of 12: 2 blocks for SMup, 2 blocks for SMdown, 2 blocks for DTup, and 2 blocks for DTdown. We utilized the SI to collect gait data and calculate gait speed. Our goal was to evaluate the effect of different task(s) on gait speed, in the interests of predicting the chances of fall incidents in senior citizens. A comprehensive analysis of gait data collected during this case study is still underway. We present our initial findings in the sections that follow.

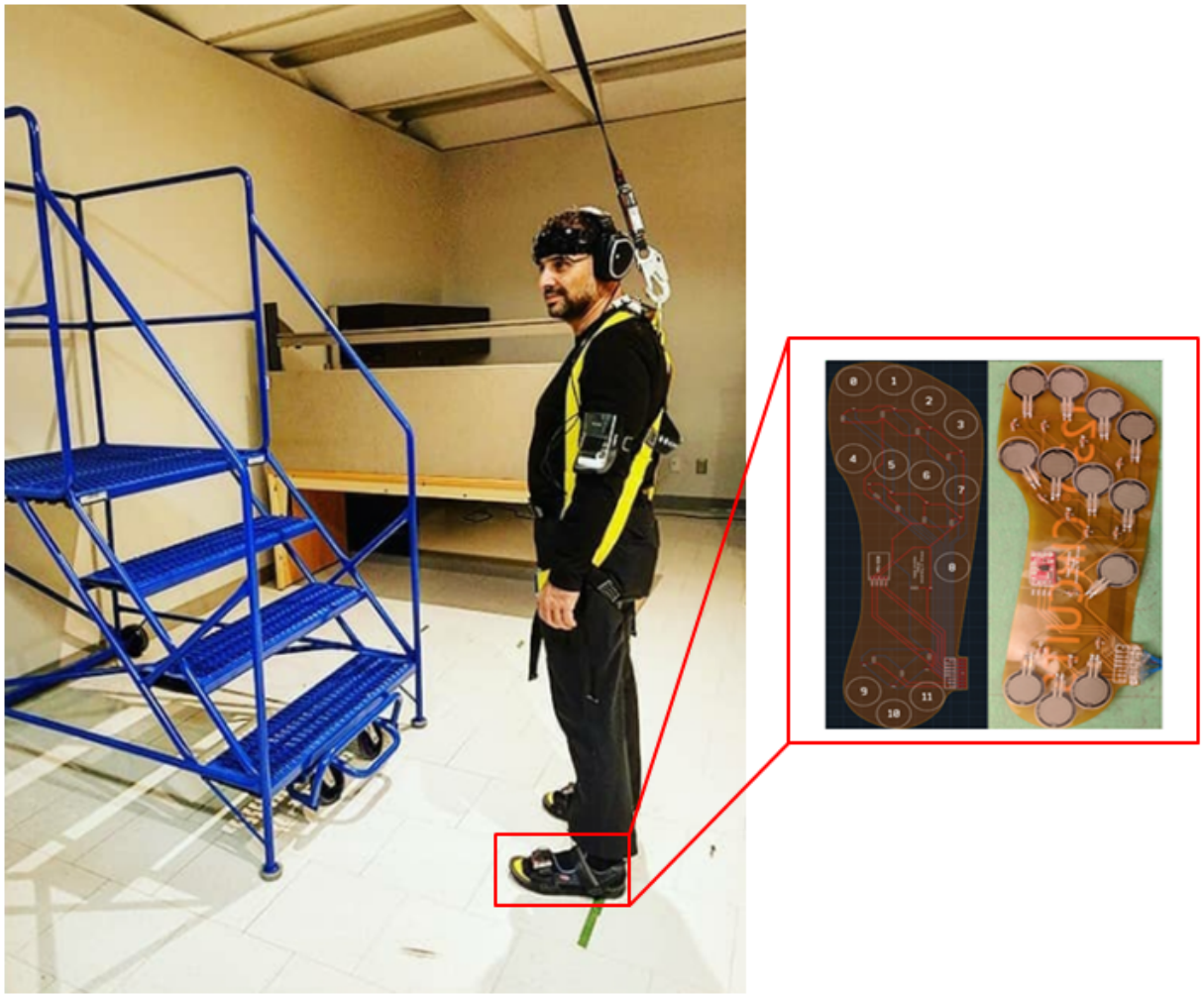


Figure 4.11: Professor El Saddik is fitted with the various devices to be used in this case study during a preparation session prior to data collection

4.2.4 DT Data Source and Standardization Module

We collected gait data from twenty healthy seniors (14 females) over the age of 60, who were recruited from community centers across Ottawa, Canada by the University of Ottawa's

Interdisciplinary School of Health Sciences team. Healthy was defined as being able to walk 15 meters without assistance, having no hearing aids or impairments, and no neuromuscular pain – conditions that would affect participants' ability to perform the cognitive task and/or stair climbing or descent. Participants were asked to disclose their gender, age, and health conditions, as well as information about whether they had fallen previously while walking. We collected this information because, according to the World Health Organization [135], a person's age, gender, and health conditions can increase the risk of fall and affect the type and severity of possible injury. Table 4.3 shows participant data, including gender, age, previous falling incidents (if any) and health condition(s). This case study was approved by the University of Ottawa Research Ethics Board and all participants provided written informed consent prior to participation.

In order to ensure consistent measurements across participants and accommodate different shoe sizes, small and large smart shoe insoles were developed, along with two pairs of adjustable sandals. For data collection, we also used a Bluetooth smartphone application [18] throughout the experiment, designed to communicate with the insoles and receive sensor readings, which consisted of participants' plantar pressure data. The collected data are standardized since we utilized the standardized SI, as discussed in Section 4.2.2.

4.2.5 Data Analytics Module

Data from each insole was collected at a frequency of 10 readings per second from each of the 12 pressure sensors. Different patterns of plantar pressure data emerge given that pressure is applied differently on the sensors of the smart insoles at different stages of the experiment.

Table 4.3: Participants' gender, age, previous falling incidents (if any) and health condition(s)

Participant ID	Gender		Age	Have you ever fallen while walking?		Health Condition(s)?		
	Female	Male		Yes	No	Yes	No	If Yes, what?
1	✓		76		✓		✓	
2	✓		69	✓				Arthritis
3		✓	81		✓		✓	
4		✓	73		✓	✓		High Blood Pressure
5		✓	71		✓	✓		High Blood Pressure
6	✓		72		✓	✓		Thyroid Condition
7	✓		66		✓		✓	
8		✓	65		✓		✓	
9		✓	60		✓		✓	
10		✓	77		✓		✓	
11	✓		61		✓		✓	
12	✓		81		✓	✓		Diabetes, Thyroid Condition
13	✓		80		✓		✓	
14	✓		84	✓		✓		Arthritis
15	✓		77		✓		✓	
16	✓		81		✓	✓		High Blood Pressure, Arthritis
17	✓		67	✓		✓		High Blood Pressure, Diabetes
18	✓		72		✓	✓		Cardiovascular Condition, High Blood Pressure, Diabetes
19	✓		70		✓	✓		Cardiovascular Condition
20	✓		66		✓	✓		Thyroid Condition

Distinct patterns characterize level walking, stair climbing, and stair descent [147]. For each participant, we extracted duration data for walking and stair-related activities during all 8 blocks of SMup, DTup, SMdown and DTdown, following the plantar pressure pattern related to the level walking, stairs climbing and descent. Figure 4.12 shows an example of the plantar pressure pattern for participant ID 7, sensor 3, during walking and DT stair descent. As illustrated in Figure 4.13, we calculated the mean gait speed (MGS) for each of the blocks of walking and stair climbing for each participant based on a total distance walked of 6.233 m. Given that we have 2 blocks for every condition, each participant will have four mean gait speeds: MGS-SMup, MGS-DTup, MGS-SMdown, MGS-DTdown. Thus, we can compare the MGS for each participant in the SM vs. DT conditions and evaluate risk of fall, because slower speed is associated with increased risk of falls according to the research in [136]. Slow gait is defined by a gait speed of less than 70 cm/s [136]. Table 4.4 shows the MGS in (cm/sec) for each participant in the four conditions: MGS-SMup, MGS-DTup, MGS-SMdown and MGS-DTdown.

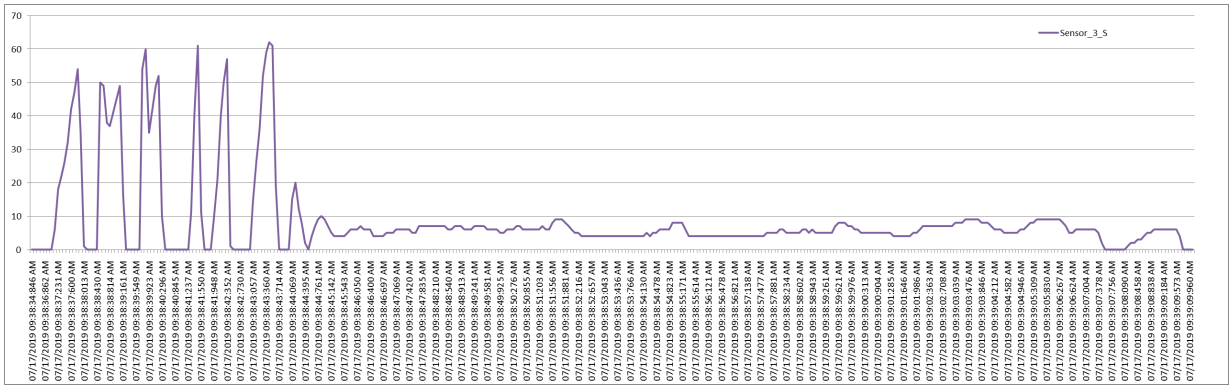


Figure 4.12: Plantar pressure pattern of sensor 3 for DT stair descent of participant ID 7

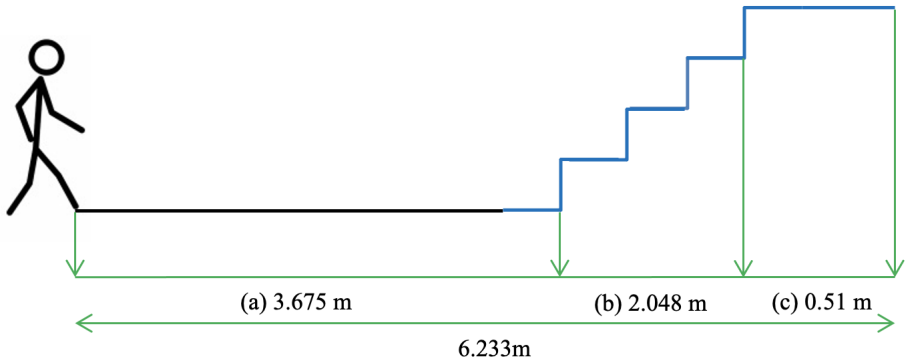


Figure 4.13: The total distance walked by each participant in this case study

Table 4.4: Mean gait speed (MGS) for each participant in SMup, DTup, SMdown and DTdown

Participant ID	Mean Gait Speed (cm/sec)			
	MGS-SMup	MGS-DTup	MGS-SMdown	MGS-DTdown
1	62.33	73.33	77.91	59.36
2	69.26	51.94	65.61	77.91
3	69.26	62.33	77.91	73.33
4	83.11	73.33	65.61	54.20
5	59.36	59.36	54.20	56.66
6	69.26	62.33	54.20	62.33
7	69.26	65.61	73.33	73.33
8	49.86	46.17	54.20	56.66
9	65.61	56.66	73.33	73.33
10	69.26	73.33	69.26	69.26
11	51.94	47.95	46.17	41.55
12	54.20	62.33	56.66	69.26
13	33.69	34.63	41.55	37.78
14	54.20	51.94	54.20	59.36
15	49.86	51.94	62.33	59.36
16	47.95	44.52	49.86	42.99
17	83.11	89.04	77.91	83.11
18	69.26	73.33	77.91	89.04
19	65.61	49.86	54.20	49.86
20	73.33	77.91	69.26	62.33

Figure 4.14 and Figure 4.15 depict the MGSs for each participant during stair climbing (4.14) and stair descent (4.15), which are added to the participant record that is stored in the Health DT Storage in the cloud (see Figure 4.6).

Considering the direction and the value determining slow gait speed, the participants in this study are vulnerable for fall risk during stair climbing more than stair descent. Considering the value determining slow gait speed only (less than 70 cm/sec) implies that the majority of participants in this study are in possibly increased risk of fall because most of the MGSs are less than 70 cm/sec regardless the motion direction. Although, the difference in MGS is not significant between the SM and DT conditions, the MGS for most of the participants is slower during the DT condition compared to the SM condition, which may increase the risk of fall during DT due to the effect of multitasking.

These initial findings need further investigation since this case study is still in progress. For example, we plan to investigate the gait speed of walking (a and c in Figure 4.13) vs. stair-related activities (b in Figure 4.13) and compare our findings accordingly. We also plan to take gender, previous falling incidents, and health condition of the participant into account as another dimension of our investigation.

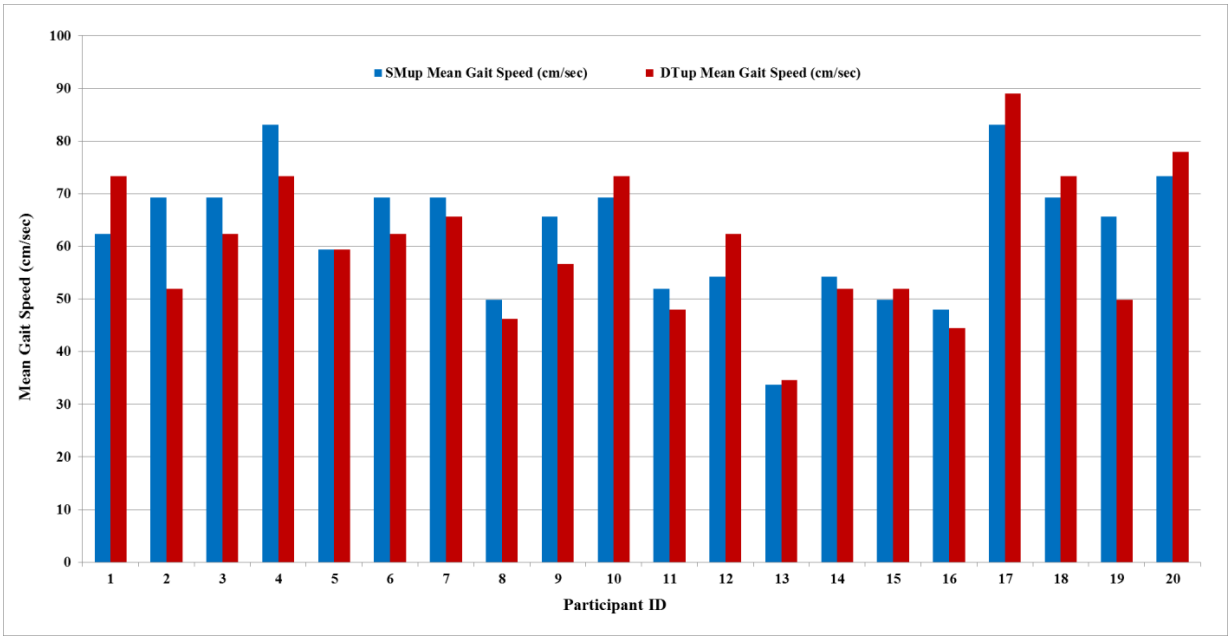


Figure 4.14: Participants' mean gait speed (MGS) in SM vs. DT during stair climbing

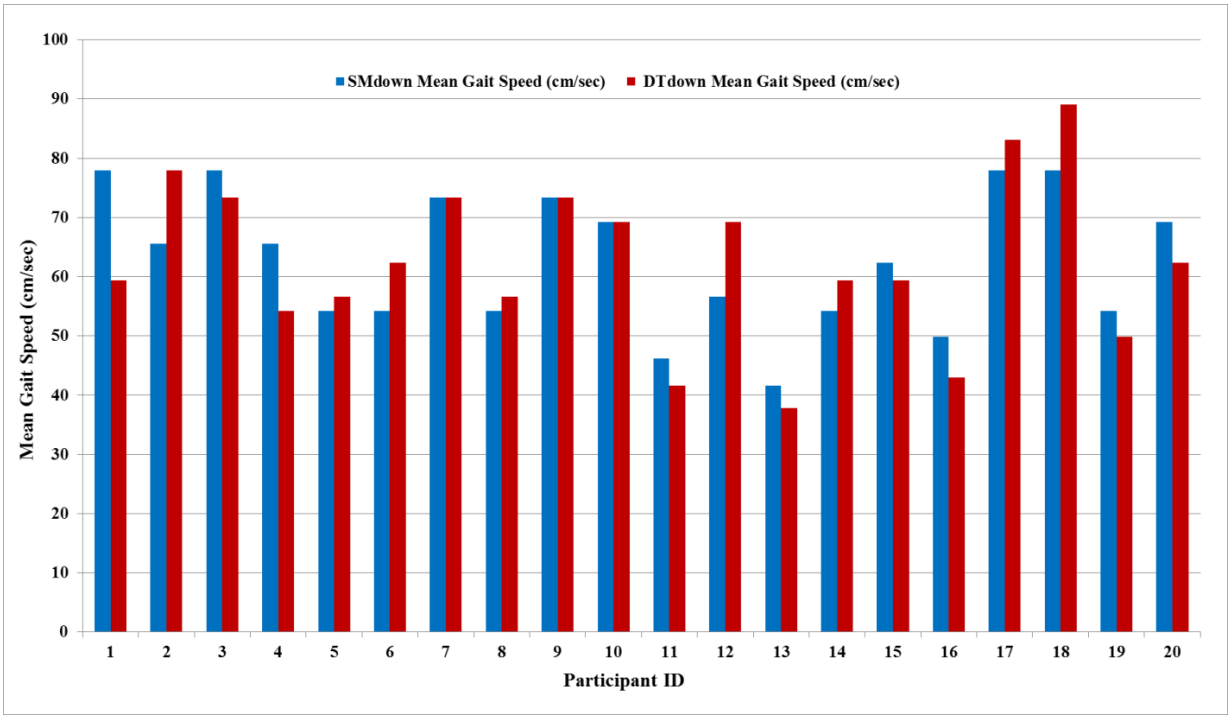


Figure 4.15: Participants' mean gait speed (MGS) in SM vs. DT during stair descent

4.2.6 DT-DNA Modeling and Visualizing Modules

DT-DNA Modeling

To build DT-DNA for senior citizens that incorporates information relevant to fall prevention, we retrieved participant data from the Health DT Storage in the cloud, along with DT-DNA Rules and Codes, and defined DT-DNA bases as follows:

- **Geo-temporal (G base)** in this experiment is defined by the four identifiers, Continent-Country-City-Year. Given that the full date of the experiment is in the

form of (Day-Month-Year), it is added under the fourth identifier instead of (Year), using the codes we proposed (see Appendix A), in order to facilitate data modeling of participants who are recruited for multiple studies in the Interdisciplinary School of Health Sciences. Thus, this G base is defined as: NA-CA-OT-(Experiment Date), where NA stands for North America, CA stands for Canada, and OT stands for Ottawa. For the experiment date, since the day and month make a difference in this case study, we used the proposed code for days and months discussed in Section 4.1.6.

- **Context (C base)** in this experiment is defined by the hardware used for data collection. Since the focus of this study is on collecting plantar pressure data, SI is considered to be the main hardware. Again, we used the proposed code for numerical values as discussed in Section 4.1.6, and preceded them by (HW) for hardware. Since the SI is one of multiple devices used in this study, and this initial modeling result will be stored in the Health DT Storage, we assigned the number value 4 to the SI, to continue the sequence order of hardware stored previously that might be used in future studies and to facilitate the expansion of the hardware code. Thus, the code for SI is HWAD, since AD is the proposed code for the number 4.

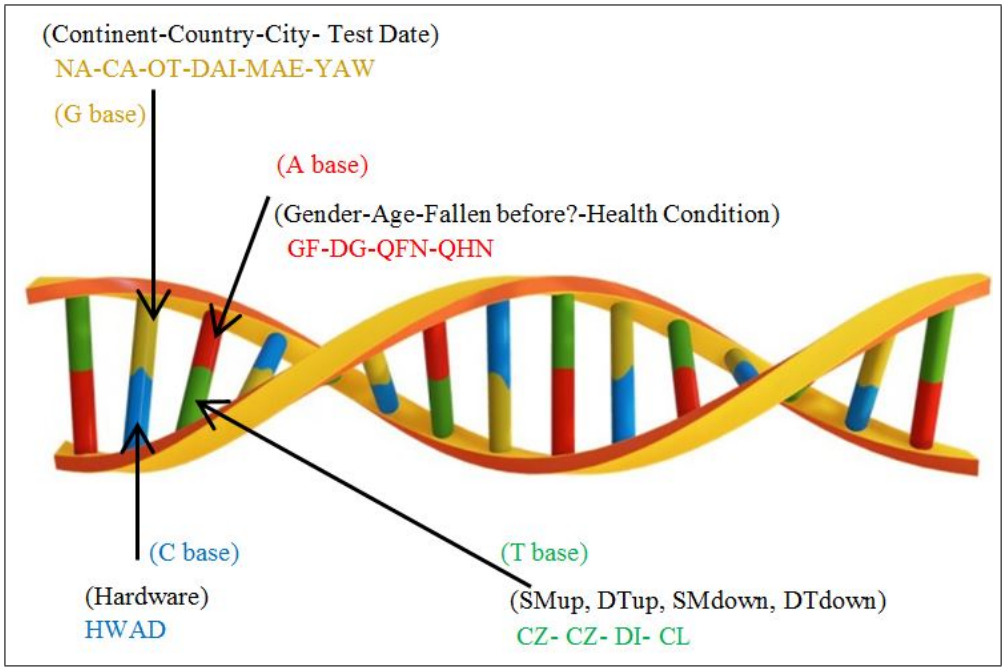
- **Anthropometric (A base)** in this experiment is defined by the anthropometric data for each participant. These are *gender*, *age*, *previous falling incident (if any)* and *health condition(s)* as discussed in Section 4.2.4. We used the same code discussed in Section 4.1.6 for *gender*, and *age*. We proposed a code QF, for *question on previous falling incident (if any)*, followed by Y (for Yes) or N (for No). Thus, if the participant reported a previous falling incident, the code for that participant would be QFY. Similarly, we proposed a code QH, which stands for *question on health*

condition(s) (if any), followed by the proposed code for numerical values, based on the proposed number value for each disease: 1=Arthritis, 2=Cardiovascular Condition, 3=Diabetes, 4=High Blood Pressure, 5=Thyroid Condition, etc. Thus, the code QHAA is used for a participant, who suffers from Arthritis, QHAB for a Cardiovascular Condition, QHAC for Diabetes, QHAD for High Blood Pressure, QHAE for Thyroid Condition, and so on, for other diseases. Multiple codes are used for participants who suffer from different diseases.

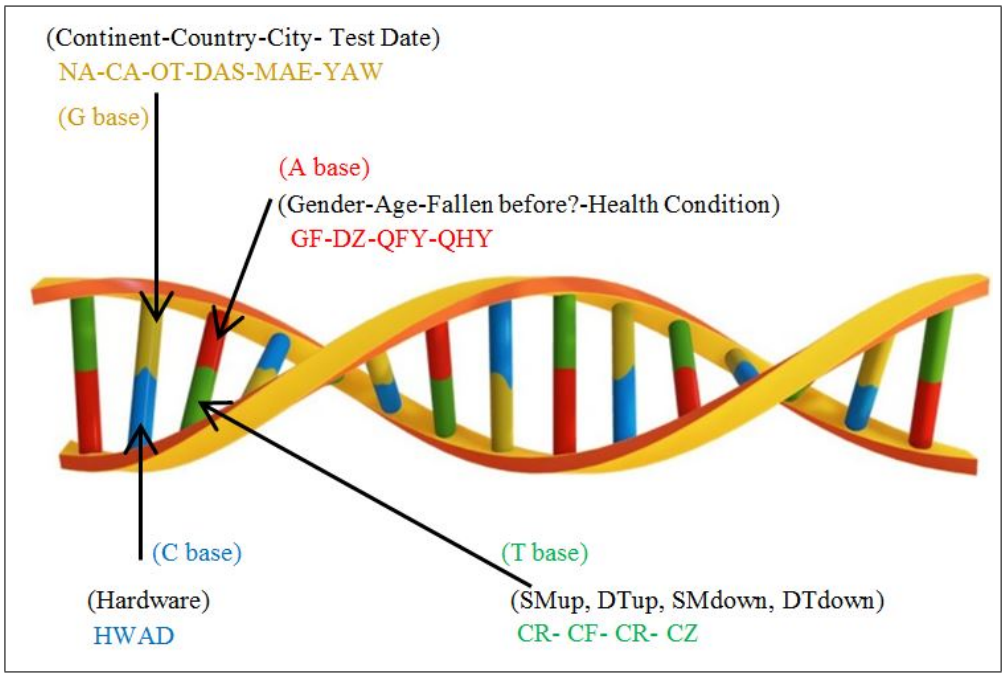
- **Task (T base)** in this experiment is the task of measuring mean gait speed (MGS) in SM vs. DT conditions while climbing (SMup vs. DTup) and descending (SMdown vs. DTdown) stairs, as discussed in Section 4.2.5. Task is defined by the MGS values in the four conditions towards fall prevention, which are shown in Table 4.4.

Figure 4.16 visualizes the built DT-DNA for two participants (Participant 1 and Participant 2) depicted in Figure 4.14 and Figure 4.15. The visualization of the built DT-DNAs for all participants is shown in Appendix C. The built DT-DNA for each participant is shown using the proposed code for the base values of the DT-DNA, as discussed above. The built DT-DNA for the first participant (participant 1) shows that she performed the experiment using the SI as shown in C and G bases respectively. It also shows she is a 76-year-old woman who has never fallen previously and has no health conditions, as shown in A base. It also shows that the MGSs for this woman while climbing stairs are 62 cm/sec and 73 cm/sec, whereas her MGSs while descending stairs are 78 cm/sec and 59 cm/sec in SM and DT conditions respectively, in each direction. This visual representation of the DT-DNA provides a description of the results of health tasks that this citizen performed

on that day, including the task's context, in addition to her anthropometric data. Thus, as per the third requirement, we show that the model is visualizable and this may facilitate the understanding of the results by citizens, stakeholders and regulatory authorities who care about citizens' health. The limitations discussed in Section 4.1.7 regarding the evaluation of this visualization and the expansion of the proposed code for numerical values applied in this case study.



Participant 1



Participant 2

Figure 4.16: DT-DNA visualization for Participant 1 and Participant 2 of this study

Chapter 5

DT-DNA Paradigm at the Meso and Macro Levels

In this chapter, we present our work on building health DT-DNA for communities and cities using the DT-DNA-based paradigm. Collected data in both the case studies presented in this chapter are related to citizens' health in a specific city or community within the city. We present the proof-of-concept of our DT-DNA-based framework for building community and city health DTs based on ISO 37120 standards [148], which reflect health data of citizens in a particular community or city. Case Study 3 (Section 5.1) uses the American College of Sports Medicine (ACSM) Fitness Index [10] to build DT-DNA of community fitness in Las Vegas and Oklahoma City, to show which city is more fit (Section 5.3). Case Study 4 (Section 5.2) uses the World Council on City Data (WCCD) portal [11] to build DT-DNA of ISO 37120-based health services in Boston and Quebec City in 2017, to show which city had better services at that year towards enhancing QoL (Section 5.3). We

discuss the algorithm we developed to compare cities in terms of their community fitness and health services status in Section 5.3.

5.1 Case Study 3: Building DT-DNA of Community Fitness in Las Vegas and Oklahoma City

In this case study, we present the proof-of-concept for the proposed DT-DNA model using data collected at the meso level. We present our proposed DT-DNA framework in Figure 5.1 (adapted from Figure 3.2) to illustrate this case study. This proof-of-concept shows a case where the input data to the framework are not standardized according to ISO 37120, so we present a wrapper designed to handle non-standardized data. We demonstrate how the proposed DT-DNA model provides results that are comparable in their accuracy to the benchmark data available, as described below. Thus, this case study shows the potential applicability of the proposed model to real-world communities (e.g., fitness communities, diabetics communities, seniors' communities, etc.) in a given city.

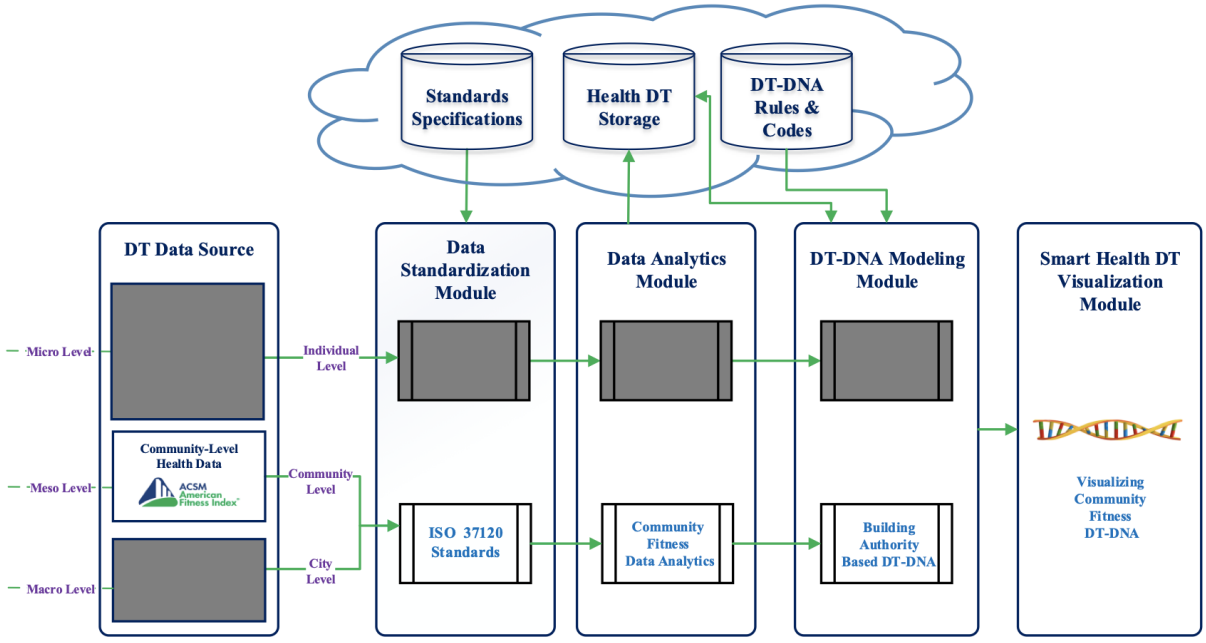


Figure 5.1: Case Study 3: Building DT-DNA of community fitness in Las Vegas and Oklahoma City based on ISO 37120 (adapted from Figure 3.2; components not relevant to this case study are greyed out)

5.1.1 DT Data Source

We collected benchmark data from the Fitness Index developed by the American College of Sport Medicine (ACSM) [10], which represents the fitness level of a given city based on data collected from reports and surveys on personal and community/environment health indicators. These data are then sent to the next module and handled by the ISO 37120 subcomponent, as discussed below.

5.1.2 Data Standardization Module

In this module, an ISO 37120-based wrapper is integrated to process the non-standardized data and map them to the proper service using standard specifications stored in the cloud. It is needed because each city has its own way of collecting and handling community and city-wide health data. We design this ISO 37120-based wrapper to handle the data collected from the ACSM Fitness Index, which are not standardized. The wrapper, depicted in Figure 5.2, serves as a gateway between the non-standardized health data of a given community or city, and the proposed DT-DNA framework for building standardized health DTs. The main role of the wrapper is to accommodate the non-standardized indicators under ISO 37120 services. The ACSM Fitness Index includes 33 indicators categorized under five categories: Health Behaviors, Health Outcomes, Built Environment, Recreational Facilities, and Policy and Funding. The Health Behaviors and Health Outcomes indicators are accommodated under Health (H), Transportation (T), and Safety (V) services of ISO 37120. Built Environment indicators are accommodated under Environment and climate change (E), and Economy (M) services of ISO 37120. Recreational Facilities indicators are accommodated under Recreation (R) and Sport and Culture (O) services of ISO 37120. Policy and Funding indicators are accommodated under Energy (N), Finance (F) and Education (D) services of ISO 37120.

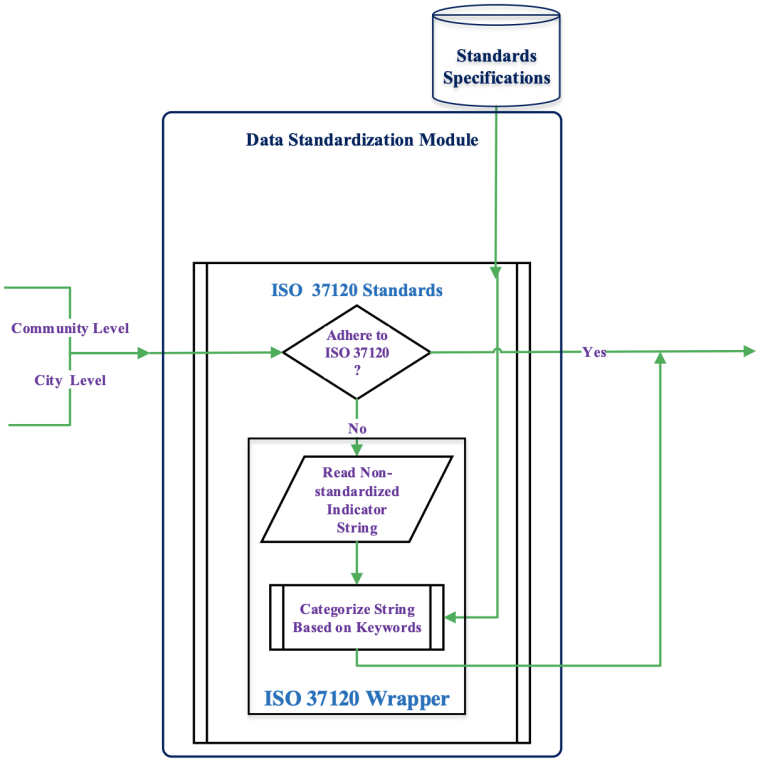


Figure 5.2: Suggested 37120 Wrapper

5.1.3 Data Analytics Module

After standardizing the ACSM Fitness Index indicators according to ISO 37120, we gathered data (scores) of Las Vegas and Oklahoma City from the ACSM Fitness Index portal [10], as shown in Table 5.1. The table also shows the overall score out of 100 for both cities as shown in the index portal. In keeping with the naming scheme discussed in Section 3.2, we named all indicators under the five categories according to the scheme, and mapped each indicator data to its coded value, as shown in Table 5.1. All indicators are supportive

thus their codes start with (S) as illustrated in Figure 5.3 and shown in the Indicator Code column in Table 5.1, since they are standardized using the wrapper and serve a specific community. The second letter in each indicator code represents the ISO 37120 service initial (e.g., H for Health), to which this indicator is mapped in the standardization module as discussed in Section 5.1.2. The last three letters in each indicator code represent the indicator's clause initials. We used the letter (P) to represent the percentage character (%) in the indicator's clause. We also used the letter (R) to represent the mentioned rate in the indicator's clause e.g. /10,000 residents in the indicator *Park units/10,000 residents* and the registered mark. For clarification, we bold the indicator's initials in each clause that are used in the indicator code. Thus, the code for the first indicator in Table 5.1 is explained as follows:

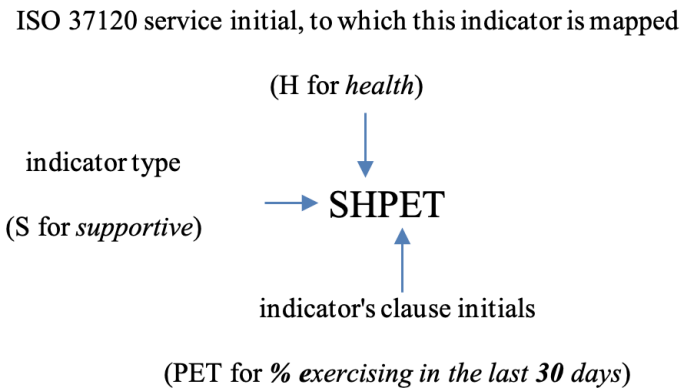


Figure 5.3: T base - sample code

Table 5.1: Las Vegas and Oklahoma City data collected from the ACSM Fitness Index portal [10] on January 31st, 2020 and coded according to the proposed DT-DNA model

Overall Score		**Las Vegas (LV) (34.3)**			**Oklahoma City (OC) (20.8)**		
Health Behaviors	Indicator Code	Actual	Rounded	Coded	Actual	Rounded	Coded
% exercising in the last **30** days	SHPET	70.3	70	DA	71.7	72	DC
% meeting **a**erobic activity guidelines	SHPMA	43.7	44	BX	43.8	44	BX
% meeting aerobic & **s**trength activity guidelines	SHPMS	17.6	18	AT	16.5	17	AS
% bicycling or **w**alking to work	STPBW	1.6	2	AB	1.7	2	AB
% using **p**ublic **t**ransportation to work	STPPT	3.4	3	AC	0.7	1	AA
% consuming 2+ **f**ruits/day	SHPCF	24.1	24	BA	23.4	23	AZ
% consuming 3+ **v**egetables/day	SHPCV	8.9	9	AI	8.9	9	AI
% smoking	SHPSM	17.8	18	AT	17.8	18	AT
Health Outcomes	Indicator Code	Actual	Rounded	Coded	Actual	Rounded	Coded
% in **e**xcellent or very good **h**ealth	SHPEH	42.9	43	BW	46.2	46	BZ
% physical health **n**ot good during the past 30 days	SHPPN	33.5	34	BL	35.6	36	BN
% mental **h**ealth not good during the past 30 days	SHPMH	29.1	29	BF	37.8	38	BQ
% with **ob**esity	SHPOB	26.9	27	BD	34.4	34	BL
% with **as**thma	SHPAS	10.5	11	AL	9.2	9	AI
% with **h**igh **b**lood pressure	SHPHB	32.5	33	BK	34.2	34	BL
% with **a**ngina or **c**oronary heart disease	SHPAC	4.4	4	AD	4.9	5	AE
% with **st**roke	SHPST	2.8	3	AC	3.3	3	AC
% with **di**abetes	SHPDI	10.8	11	AL	10.9	11	AL
Pedestrian **f**atality **r**ate/100,000 residents	SVPFR	2.6	3	AC	2.5	3	AC

Overall Score		**Las Vegas (LV) (34.3)**			**Oklahoma City (OC) (20.8)**		
Built Environment	Indicator Code	Actual	Rounded	Coded	Actual	Rounded	Coded
Air **q**uality **i**ndex	SMAQI	37.3	37	BP	76.3	76	DG
Bike **S**core®	SEBSR	43.9	44	BX	39.5	40	BS
Farmers' **m**arkets/1,000,000 residents	SEFMR	6.2	6	AF	14	14	AP
Park **u**nits/10,000 residents	SEPUR	8	8	AH	2.4	2	AB
% within a 10-minute walk to a park	SEPWR	64	64	CT	42	42	BV
Walk **S**core®	SEWSR	41.1	41	BT	33.1	33	BK
Recreational Facilities	Indicator Code	Actual	Rounded	Coded	Actual	Rounded	Coded
Ball **d**iamonds/10,000 residents	SOBDR	0.8	1	AA	0.5	1	AA
Basketball **h**oops/10,000 residents	SOBHR	1.3	1	AA	1.9	2	AB
Park **p**laygrounds/10,000 residents	SOPPR	3.9	4	AD	1.7	2	AB
Recreational **c**enters/20,000 residents	SRRCR	0.8	1	AA	0.8	1	AA
Swimming **p**ools/100,000 residents	SOSPR	1.9	2	AB	0.8	1	AA
Tennis **c**ourts/10,000 residents	SOTCR	1.1	1	AA	1.2	1	AA
Policy & Funding	Indicator Code	Actual	Rounded	Coded	Actual	Rounded	Coded
Local complete **s**treets **p**olicy	SNLSP	0	0	ZZ	0	0	ZZ
Park **e**xpenditure/resident (adjusted)	SFPER	96	96	ED	51	51	CE
Physical **e**ducation **r**equirement	SDPER	1	1	AA	1	1	AB

5.1.4 DT-DNA Modeling and Visualization Modules

The data in Table 5.1 represent the T base in the intended DT-DNAs of community fitness for LV and OC, as shown in Figure 5.4(1). We removed the indicator codes from the DNA sequence of each city because they are identical in each sequence and assigned a fixed position for each indicator. Thus, each 2-alphabet code represents one indicator in a sequential order, separated by a hyphen from the next indicator value. Other bases follow the same positioning rules.

Using the city's name, we determined A, C and G bases data utilizing stored data and web resources. Since we build the DT-DNA on a community (meso) level, the A base represents the authority. The government in the United States (US) where LV and OC are located, is a constitutional (CS) federal republic (FR) and both LV and OC are managed by City Halls (CH). Thus, the A base of LV and OC are represented by these sequences respectively: CS-FR-LV-CH, CS-FR-OC-CH. The A base sequences are shown in Figure 5.4(2).

LV and OC are fairly identical in their C base data since both cities have flat (SF) and hilly (SH) surfaces, are built on deep and hard rocks (RD) and are inland cities (LN). Thus, the sequence of C base, particularly EC, based on ISO 37120 [33] is: ECSF, ECSH, ECRD, ECLN, as shown in Figure 5.4(3). We removed the indicator's initial from the names because it is the same in each sequence, but we assigned a fixed position for each indicator of EC.

The G base data of both cities is shown in Figure 5.4(4) according to the model description in Section 3.2. However, we will use the full date of data collection from ACSM, which

is Jan 31st 2020 in this case study, following the sitting discussed in 4.1.6 (DBH, MAA, YAX) because the data in the ACSM portal updates frequently. The DT-DNA sequences of LV and OC are shown in Figure 5.4(5) and (6) respectively.

Aligning DNA sequences in MatGAT [128], a DNA alignment tool used to determine the percentage of identity between two or more DNA sequences, took place at this point. Feeding LV DT-DNA and OC DT-DNA sequences into MatGAT provides 66.7 % as the identity percentage of the fitness level of the communities in LV and OC using indicators listed in Table 5.1, after aligning the DNA sequences of both cities. Since they are approximately half identical, this implies that one of the cities is better than the other city in fitness level as shown in Section 5.3. Comparison of our benchmark data as ground truth provides almost the same result (68.04%).

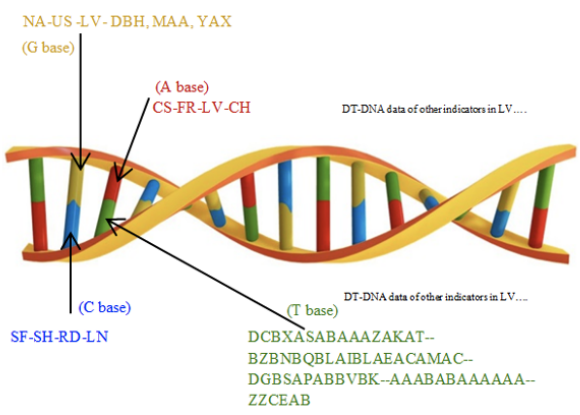
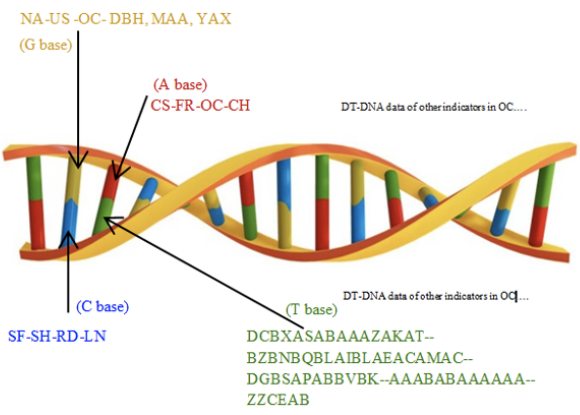
<table>
<tr><td colspan="2">(1) T base for LV vs. OC</td></tr>
<tr><td>>Las Vegas
DABXATABACBAAIAT--
BWBLBFBDALBKADACALAC--BPBXAFAHCTBT--AAAAADAAABAA--ZZEDAA</td><td>>Oklahoma City
DCBXASABAAAZAKAT--
BZBNBQBLAIBLAEACAMAC--DGBSAPABBVBK--
AAABABAAAAAA--ZZCEAB</td></tr>
<tr><td colspan="2">(2) A base for LV vs. OC</td></tr>
<tr><td>>Las Vegas
CS-FR-LV-CH</td><td>>Oklahoma City
CS-FR-OC-CH</td></tr>
<tr><td colspan="2">(3) C base for LV vs. OC</td></tr>
<tr><td>>Las Vegas
SF-SH-RD-LN</td><td>> Oklahoma City
SF-SH- RD-LN</td></tr>
<tr><td colspan="2">(4) G base for LV vs. OC</td></tr>
<tr><td>>Las Vegas
NA-US -LV- DBH, MAA, YAX</td><td>> Oklahoma City
NA- US -OC- DBH, MAA, YAX</td></tr>
<tr><td>(5) LV-DNA
 
</td><td>(6) OC-DNA
 
</td></tr>
</table>

Figure 5.4: A, T, C and G bases, values of the proposed DT-DNA for LV and OC and results visualized in DNA for community fitness

5.2 Case Study 4: Building DT-DNA of Health Services in Boston and Quebec City

In this case study, we present the proof-of-concept of the proposed DT-DNA model using data collected at the macro level. We present our proposed DT-DNA framework in Figure 5.5 (adapted from Figure 3.2) to illustrate this case study. This proof-of-concept shows a case where the input data to the framework are standardized according to ISO 37120. We demonstrate how the proposed DT-DNA model provides results that are comparable in their accuracy to the benchmark data available, as described below. Thus, this case study shows the potential applicability of the proposed model to real-world cities.

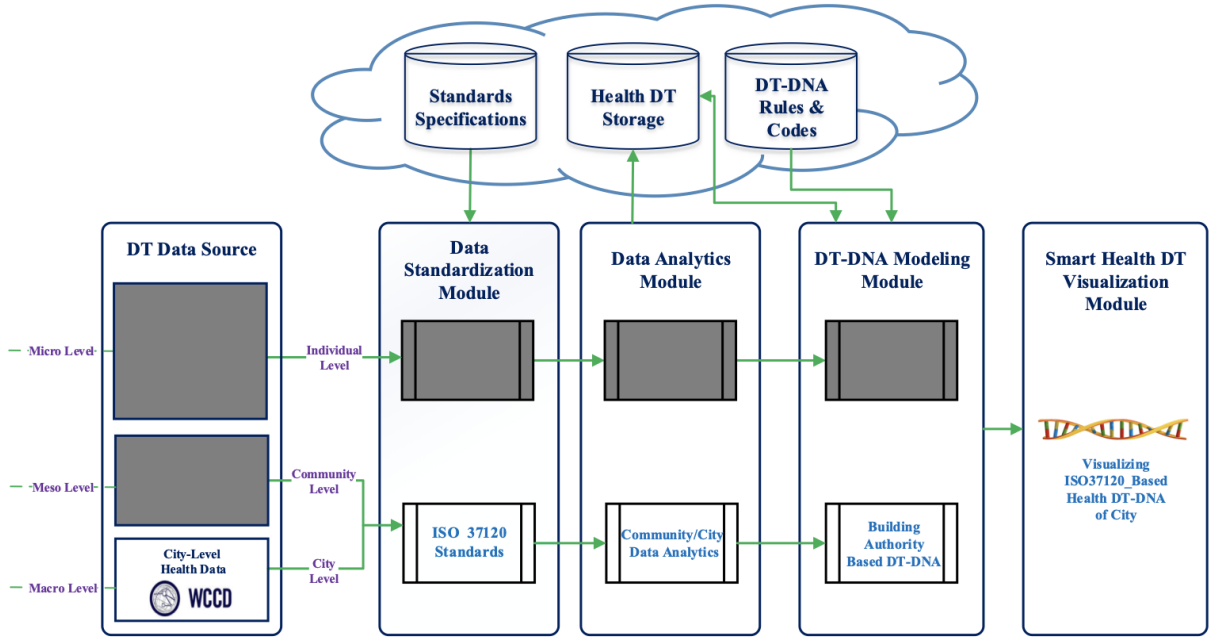


Figure 5.5: Building DT-DNA of Health Services in Boston and Quebec City based on ISO 37120 [120] (adapted from Figure 3.2; components not relevant to this case study are greyed out)

5.2.1 DT Data Source, Standardization and Analytics Modules

We collected benchmark data from the World Council on City Data (WCCD) portal [11], which provides data from ISO 37120-certified cities. As a result, this data was passed to the Data Analytics Module directly. Since city data in the WCCD portal [11] are presented on a yearly basis, we identified 2017 as the year that had the most available health data for two cities, Quebec City (QC) and Boston (BO), and gathered data from the portal for indicators of health service, as shown in Table 5.2. Following the naming scheme discussed in Section 3.2, for all indicators under the health service, we named each indicator according to the suggested scheme and mapped each indicator's data to its coded value as shown in Table 5.2. The indicators codes start with (C) for core indicators and (S) for supportive indicators as shown in the Indicator Code column in the table. The second letter in all indicators codes is H for Health. The last three letters in each indicator code represent the indicator's clause initials (refer to Section 3.2 under Task). For clarification, we bold the indicator's initials in each clause that are used in the indicator code.

5.2.2 DT-DNA Modeling and Visualizing Modules

The above mentioned data represent the T base in the intended DT-DNAs of QC and BO, as shown in Figure 5.6(1). We removed the indicator names from the DNA sequence of each city because they are identical in each sequence and assigned a fixed position for each indicator. Thus, each 2-alphabet code represents one indicator in a sequential order, separated by a hyphen from the next indicator value. Other bases will follow the same positioning rules.

Table 5.2: ISO 37120 Health service indicators [120], their codes and coded values for QC and BO

Health Service Indicators [33]	**Indicator Code**	**QC**	**BO**
Average **l**ife **e**xpectancy (core indicator)	CHALE	ZZ	ZZ
Number of i**n**-patient hospital **b**eds per 100,000 population (core indicator)	CHNNB	BN	BG
Number of **ph**ysicians per 100,000 population (core indicator)	CHNPH	AK	BQ
Under age **f**ive **m**ortality per 1,000 live births (core indicator)	CHUFM	ZZ	ZZ
Number of **n**ursing and **m**idwifery personnel per 100,000 population (supportive indicator)	SHNNM	BI	CR
Number of **m**ental health **p**ractitioners per 100,000 population (supportive indicator)	SHNMP	AM	CD
Suicide **r**ate per 100,000 **p**opulation (supportive indicator)	SHSRP	AV	ZZ

Using the city's name, we determined A, C and G bases data utilizing stored data and web resources. Since we build the DT-DNA on a city (macro) level, the A base represents the authority. The government in Canada (CA) where QC is located is a federal (FD) parliamentary democracy (PD), and QC is managed by Quebec City Council (CC). Thus, the A base of QC is represented by this sequence: FD-PD-QC-CC. Similarly, the government in the United States (US) where BO is located is a constitutional (CS) federal republic (FR) and BO is managed by Boston City Hall (CH). Thus, the A base of BO is represented by this sequence: CS-FR-BO-CH. The A base sequences are shown in Figure 5.6(2).

QC and BO are fairly identical in their C base data since both cities have flat (SF) and hilly (SH) surfaces, are built on valleys (RV) and are coastal cities (LS). Thus, the sequence of C base for both cities, particularly EC, based on ISO 37120 [120], is: ECSF, ECSH, ECRV, ECLS, as shown in Figure 5.6(3). We removed the indicator's initial from the names because it is the same in each sequence, but we assigned a fixed position for each indicator of EC.

The G base data of both cities is shown in Figure 5.6(4), according to the model description in Section 3.2. The DT-DNA sequences of QC and BO are shown in Figure 5.6(5) and Figure 5.6(6) respectively.

Aligning DNA sequences in MatGAT [128], a DNA alignment tool used to determine the percentage of identity between two or more DNA sequences, took place at this point. Feeding QC DT-DNA and BO DT-DNA sequences into MatGAT provides 48.7% as the identity percentage of the performance level of the health service indicators listed in Table 5.2, after aligning the DNA sequences of both cities. Since they are approximately half identical, this implies that one of the cities is better than the other city in health service as

shown in the algorithm proposed in the following section. Comparison of our benchmark data as ground truth provides the same result (48.7%).

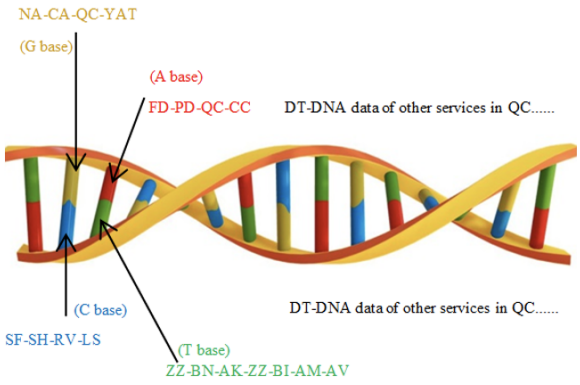
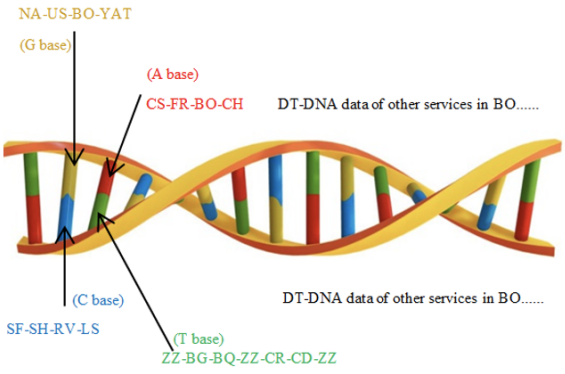
(1) T base for QC vs. BO	
>Quebec City(QC) ZZ-BN-AK-ZZ-BI-AM-AV	>Boston (BO) ZZ-BG-BQ-ZZ-CR-CD-ZZ
(2) A base for QC vs. BO	
>Quebec City FD-PD-QC-CC	>Boston CS-FR-BO-CH
(3) C base for QC vs. BO	
>Quebec City SF-SH-RV-LS	>Boston SF-SH-RV-LS
(4) G base for QC vs. BO	
>Quebec City NA-CA-QC-YAT	>Boston NA-US-BO-YAT
(5) QC-DNA 	(6) BO-DNA 

Figure 5.6: A, T, C and G bases, values of the proposed DT-DNA for QC and BO and results visualized in DNA sequences

5.3 Algorithm 1: Which City Has Better (Community Fitness/Health Services) Towards Enhancing QoL?

We proposed an algorithm to compare community fitness and cities according to the ACSM Fitness Index [10] and ISO 37120-based health services in order to test which city has better community fitness based on the proposed DT-DNA, towards enhancing QoL. We proposed this algorithm to compare which city is more fit and has better health services in LV vs. OC and QC vs. BO respectively, with the goal of enhancing QoL in these cities using their DT-DNAs.

We decided to compare the two cities of each case study due to the data availability limitation discussed above in Section 5.2.1, and the similar Environmental Context (EC) of both cities in each case study. Thus, at the meso level, we took the similarity in the EC into account while collecting data on the same date from the ACSM data portal for both cities. Similarly, at the macro level, we took both the similarity of the EC and the availability of the data for the two cities in 2017 into account when we conducted the comparison, based on ISO-standard-defined services.

The algorithm includes assigning a score to each city based on indicators (T) values. We developed the algorithm so that it utilizes the built DT-DNA of each city. We compared each indicator value (IV) for two given cities X and Y. If it is a core (C) indicator, and X is better than Y, the comparison value (CV) is assigned the value (+2), whereas, it is assigned the value (-2) if X is worse than Y. In a case where the indicator is supportive (S), we apply the same process, but the assigned CV is +1 or -1 respectively. Our reasoning here

is the fact that the core indicator is much more important than the supportive indicator, according to ISO 37120. As there is no assigned weight for either in ISO 37120, we work with the assumption that the core indicator has double the weight of the supportive one in its importance, and have assigned (+/-2) and (+/-1) weights respectively. In the equal case, where X and Y have the same IV for a specific indicator, the CV is assigned the value (0). After comparing all IVs, the sum of the CVs is calculated, and the resulting value (RV) reflects the status of X vs. Y. Positive RV means X is better than Y whereas negative RV means Y is better than X. The RV also shows how big the difference between X and Y is. The bigger the RV value, the bigger the difference and vice versa. The pseudo code of this algorithm is shown as follows.

Algorithm 1: Which city has better (community fitness/health services) towards enhancing QoL?

```
read X = first city, Y = second city, E = Service under comparison;
read N = total number of indicators under E, S = total number of E in the comparison;
set I = indicator under investigation, Type = core OR supportive, IVx = indicator value of first city, set IVy = indicator value
of second city, CVi = (comparison value of indicator i) = 0, RVe = (resulted value of service E) = 0, n = 0, s = 0;

while s < S do
  for each E in S do
   while n < N do
    for each I in E do
       compare IVx AND IVy
          if IVx > IVy AND Type = core then
                    set CVi = +2;
          else if  IVx < IVy AND Type = core then
                    set CVi = -2;
           else if IVx > IVy AND Type = supportive then
                    set CVi = +1;
          else if  IVx < IVy AND Type = supportive then
                    set CVi = -1;
           else //  IVx = Ivy
                     set CVi = 0;
    end for
    n++;
    RVe+ = CVi ;
  end while
  if RVe > 0 then
          X is better than Y in E
  else if RVe < 0 then
         X is worse than Y in E
  else
          X AND Y are equally in E
 end for
 s++;
 RV+ = RVe;
end while
  if RV > 0 then
         X is better than Y in this case
  else if V < 0 then
         X is worse than Y in this case
      else
          X AND Y are equally in this case
```

5.4 Results and Discussion

Comparison results are shown in Table 5.3 and Table 5.4 where EQ means equal, GT means greater than (better) and LT means less than (worse). The results in Table 5.3 show that LV is more fit than OC using the ACSM Fitness Index data collected on January 31st, 2020 and the proposed algorithm, since RV = 1 when we compare LV vs. OC. Similarly, the results in Table 5.4 show that BO has better health services compared to QC using health services data collected from WCCD for the year 2017 and the proposed algorithm, since RV = -1 when we compare QC vs. BO.

The proof-of-concept shows promising results in terms of applicability of the proposed DT-DNA model and framework in handling city health services data collected through various sources, and presenting them in a standardized, unique general model representing the city health services DT. The results of applying the algorithm demonstrate its potential to compare the performance of city health services towards enhancing quality of life for citizens. Results also reflect the benefit of analogy with a biological DNA model since the proposed DT-DNA model can represent any city yet protect its unique identity. The model also fulfills the human-inclusion requirement through its A base for authority, and provides the means to represent fundamental as well as customized services for each city while adhering to ISO 37120 standards. The proposed model is able to handle all 17 services listed in the ISO 37120 standards and provides a chance to accommodate future services through available chromosomes in the city cell. One of the limitations of the current implementation is regarding the indicator's data, where we rounded the numerical value in order to provide the coded one – a limitation we plan to investigate in the future.

Table 5.3: Results of applying Algorithm 1 on LV vs. OC data

Overall Score		**Las Vegas (LV)**	**Oklahoma City (OC)**	**Which City is More Fit?**	
Health Behaviors	Indicator Code			Coded Result	Numerical Result
% exercising in the last 30 days	SHPET	DA	DC	LT	-1
% meeting aerobic activity guidelines	SHPMA	BX	BX	EQ	0
% meeting aerobic & strength activity guidelines	SHPMS	AT	AS	GT	1
% bicycling or walking to work	STPBW	AB	AB	EQ	0
% using public transportation to work	STPTW	AC	AA	GT	1
% consuming 2+ fruits/day	SHPCF	BA	AZ	GT	1
% consuming 3+ vegetables/day	SHPCV	AI	AI	EQ	0
% smoking	SHPSM	AT	AT	EQ	0
Health Outcomes	Indicator Code	Coded	Coded	Coded Result	Numerical Result
% in excellent or very good health	SHPEH	BW	BZ	LT	-1
% physical health not good during the past 30 days	SHPNH	BL	BN	LT	-1
% mental health not good during the past 30 days	SHPMH	BF	BQ	LT	-1
% with obesity	SHPOB	BD	BL	LT	-1
% with asthma	SHPAS	AL	AI	GT	1
% with high blood pressure	SHPHB	BK	BL	LT	-1
% with angina or coronary heart disease	SHPAH	AD	AE	LT	-1
% with stroke	SHPWS	AC	AC	EQ	0
% with diabetes	SHPWD	AL	AL	EQ	0
Pedestrian fatality rate/100,000 residents	SSPFR	AC	AC	EQ	0
Built Environment	Indicator Code	Coded	Coded	Coded Result	Numerical Result
Air quality index	SEAQI	BP	DG	LT	-1
Bike Score®	SECBS	BX	BS	GT	1
Farmers' markets/1,000,000 residents	SECFM	AF	AP	LT	-1
Park units/10,000 residents	SECPU	AH	AB	GT	1
% within a 10-minute walk to a park	SECPW	CT	BV	GT	1
Walk Score®	SECWS	BT	BK	GT	1

Overall Score		Las Vegas (LV)	Oklahoma City (OC)	Which City is More Fit?	
Recreational Facilities	Indicator Code	Coded	Coded	Coded Result	Numerical Result
Ball diamonds/10,000 residents	SSCBD	AA	AA	EQ	0
Basketball hoops/10,000 residents	SSCBH	AA	AB	LT	-1
Park playgrounds/10,000 residents	SSCPP	AD	AB	GT	1
Recreational centers/20,000 residents	SRRCR	AA	AA	EQ	0
Swimming pools/100,000 residents	SSCSP	AB	AA	GT	1
Tennis courts/10,000 residents	SSCTC	AA	AA	EQ	0
Policy & Funding	Indicator Code	Coded	Coded	Coded Result	Numerical Result
Local complete streets policy	SELSP	ZZ	ZZ	EQ	0
Park expenditure/resident (adjusted)	SFPER	ED	CE	GT	1
Physical education requirement	SEDPR	AA	AB	EQ	0
		LV vs. OC		**GT**	**1**

Table 5.4: Results of applying Algorithm 1 on QC vs. BO data

Health Service Indicators [33]	**Health Code**	**QC**	**BO**	**QC vs. BO**	**-1**
Average life expectancy (core indicator)	CHALE	ZZ	ZZ	EQ	0
Number of in-patient hospital beds per 100,000 population (core indicator)	CHNNB	BN	BG	GT	2
Number of physicians per 100,000 population (core indicator)	CHNPH	AK	BQ	LT	-2
Under age five mortality per 1,000 live births (core indicator)	CHUFM	ZZ	ZZ	EQ	0
Number of nursing and midwifery personnel per 100,000 population (supportive indicator)	SHNNM	BI	CR	LT	-1
Number of mental health practitioners per 100,000 population (supportive indicator)	SHNMP	AM	CD	LT	-1
Suicide rate per 100,000 population (supportive indicator)	SHSRP	AV	ZZ	GT	1
		QC vs. BO		**LT**	**-1**

Chapter 6

Conclusion and Future Work

In this thesis, we proposed a DT-DNA paradigm by leveraging the double helix DNA model in the domain of digital twins. Our goal was to model unified health DTs for citizens on micro, meso and macro levels, while protecting the unique identity of each citizen. We were inspired by DNA due to its ability to represent and protect the genetic blueprint of human and other organisms throughout history in a unified and unique way.

The proposed model was designed to capture citizens' health data from various sources collected by traditional methods such as surveys, or wearable devices such as PHDs, all of which provide insights on individual, community and city-wide health. The proposed model was also designed to capture contextual data and provide visual representations of results and analyses to those citizens, stakeholders and regulatory authorities who care about citizens' health.

In Chapter 2, we provided background information on digital twin technology and

DNA. We then presented state-of-the-art on digital twins for health and well-being, to emphasize the importance of utilizing health DTs towards enhancing citizens' wellness and quality of life. We also highlighted the lack of literature on standardized models for health DTs that protect each citizen's unique identity, thereby demonstrating the significance of the proposed model. Finally, we presented the role of standardization towards building effective health digital twins for citizens by discussing the two standards that relate to the proposed work: ISO/IEEE 11073 (X73) and ISO 37120. The first plays an essential role when collected health data represents citizens as individuals and the latter plays a role when collected health data represents citizens as a community or a city in general.

In Chapter 3, we started by analyzing two domains to refine the requirements of the proposed DT-DNA model. First, we analyzed the biological DNA model and highlighted the features that allow us to introduce a digital twin DNA (DT-DNA). Second, we analyzed health in smart cities to define the bases of the DT-DNA model of citizens' health, given that the intended citizens live in smart cities where they interact with the environment and are involved in, and affected by, the city's health services. We then discussed the details of the proposed DT-DNA model, followed by the proposed framework for building health DTs for citizens on micro (individual), meso (community), and macro (city) levels.

In Chapter 4, we presented the proof-of-concept of our DT-DNA-based framework for building health DTs for citizens on the micro level in two case studies utilizing ISO/IEEE 11073 (X73) standards. The first case study estimated lactate threshold (LT) towards building the DT-DNA of physically active citizens. The second case study measured gait speed towards building the DT-DNA of senior citizens for fall prevention. We also presented the details of standardizing smart shoe insoles built previously in our lab according to

ISO/IEEE 11073 (X73) standards.

In Chapter 5, we presented the proof-of-concept of our DT-DNA-based framework for building community and city health DTs in two case studies based on ISO 37120 standards, which reflect health data of citizens in a particular community or across a city. The first case study showed building the DT-DNA of community fitness in Las Vegas and Oklahoma City, and identified which city was more fit using data collected from the ACSM Fitness Index. The second case study showed building the DT-DNA of ISO 37120-based health services in Boston and Quebec City in the year 2017, and highlighted which city had better services towards enhancing QoL that year, using data collected from the WCCD.

For future work, we aim to utilize more personal health devices and wearables as data sources since they are more commonly used nowadays among citizens towards building comprehensive health DT-DNAs for citizens in the smart city. We also plan to incorporate further data sources through hard and soft sensors such as social media, which play a critical role in building personalized and customized health DT-DNAs for citizens as individuals. Regarding the individual-level case studies, we plan to utilize additional physiological features beyond LT, such as $\dot{V}$O2max, heart rate (HR), fat oxidation rate (FOR), and rate of perceived excursion (RPE), to build more comprehensive DT-DNAs of physically active citizens. Moreover, we plan to enhance the proposed ML model to estimate LT by testing and evaluating more algorithms towards deploying LT in various applications that serve purposes other than fitness evaluation, such as weight management and diabetes prediction. We envision these plans as next steps towards building comprehensive DT-DNAs for individuals served by these applications. Regarding the case study for measuring gait speed in senior citizens, we plan to incorporate more factors such as

plantar pressure and other gait features to further investigate the role they play in fall prevention. We also plan to explore the effect of anthropometric data on citizens' samples in both case studies. Since we only considered physical health, we plan to expand the proposed model to cover other health dimensions for individuals such as mental health.

On the meso and macro levels, we plan to expand our leveraging of digital twin technology to build DT-DNAs of city services due to their potential to enhance quality of life in communities and cities. We plan to collect more data to extend our analysis of the progress of a city on an annual basis by comparing the DT-DNAs of its services over time. We also aim to work on enabling the inheritance feature of the DNA model for DT-DNAs of city services towards inheriting a specific DT-DNA by other cities that share some features with the city under investigation. We plan to move towards utilizing a wider range of city data besides health data, towards modeling comprehensive city services DT-DNAs. Further, we plan to extend the module to include federal inputs to enable modeling country DT-DNA, taking into account a country's population, memberships in international forums and unions such as the G20 and European Union, and position in the internationally recognized indexes such as GDP, per capita, and happiness index. We also plan to extend the model to take nonliving entities into account under the A base such as physical entities in the IoT since the current proof of concept covers living entities or humans under A base. We also plan to investigate the efficiency and limitations of applying the model to compare different-size cities. For example, can a small city with a population of 100,000 compare effectively with cities with a few million people using the proposed model? We also plan to investigate the usage of built health DT-DNAs for improving the quality of life of citizens as individuals and as members in communities in the chosen city .

With respect to visualization of the results, we plan to conduct a study to investigate the usability of visualized DT-DNAs illustrated in the case studies, and evaluate the extent to which this visualization provides a meaningful way for citizens, stakeholders and regulatory authorities to understand and interpret results. We also plan to develop automation tools that city heads can use to compare their city with others, and to infer the health and well-being of citizens in the city based on the DT-DNA analysis and compare resulting patterns with available patterns from health authorities.

With these possibilities in mind, the proposed paradigm in this thesis provides a unified model of citizens' health DTs that takes into consideration diverse health data at different levels while preserving the unique identity of each citizen towards enhancing citizens' QoL.

References

[1] Mike Shafto, Mike Conroy, Rich Doyle, Ed Glaessgen, Chris Kemp, Jacqueline LeMoigne, and Lui Wang. Modeling, simulation, information technology & processing roadmap. *National Aeronautics and Space Administration*, 2012.

[2] Michael Grieves and John Vickers. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In *Transdisciplinary perspectives on complex systems*, pages 85–113. Springer, 2017.

[3] Amy Ann Forni. Gartner Identifies the Top 10 Strategic Technology Trends for 2017, 2016. URL https://www.gartner.com/en/newsroom/press-releases/2016-10-18-gartner-identifies-the-top-10-strategic-technology-trends-for-2017. Accessed: 2020-08-22.

[4] Christy Pettey. Gartner Identifies the Top 10 Strategic Technology Trends for 2018, 2017. URL https://www.gartner.com/en/newsroom/press-releases/2017-10-04-gartner-identifies-the-top-10-strategic-technology-trends-for-2018. Accessed: 2020-08-22.

[5] Katie Costello and Gloria Omale. Gartner Survey Reveals Digital Twins Are Entering Mainstream Use, 2019. URL https://www.gartner.com/en/newsroom/press-releases/2019-02-20-gartner-survey-reveals-digital-twins-are-entering-mai. Accessed: 2020-08-22.

[6] Kasey Panetta. 5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020, 2020. URL https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/. Accessed: 2020-08-22.

[7] Abdulmotaleb El Saddik. Digital twins: The convergence of multimedia technologies. *IEEE multimedia*, 25(2):87–92, 2018.

[8] Abdulmotaleb El Saddik, Hawazin Badawi, Roberto Alejandro Martinez Velazquez, Fedwa Laamarti, Rogelio Gámez Diaz, Namrata Bagaria, and Juan Sebastian Arteaga-Falconi. Dtwins: a digital twins ecosystem for health and well-being. In *Proc. IEEE COMSOC MMTC Commun. Frontiers*, pages 39–43, 2019.

[9] Hawazin Faiz Badawi, Fedwa Laamarti, and Abdulmotaleb El Saddik. Iso/ieee 11073 personal health device (x73-phd) standards compliant systems: A systematic literature review. *IEEE Access*, 7:3062–3073, 2018.

[10] ACSM American Fitness Index. ACSM American Fitness Index, 2020. URL https://americanfitnessindex.org/. Accessed: 2020-06-16.

[11] World Council on City Data, 2020. URL https://www.dataforcities.org/. Accessed: 2020-06-16.

[12] Brian Umbenhauer Adam Mussomeli, Aaron Parrott and Lane Warshaw. Digital twins: Bridging the physical and digital, 2020. URL https://www2.deloitte.com/us/en/insights/focus/tech-trends/2020/digital-twin-applications-bridging-the-physical-and-digital.html. Accessed: 2020-05-04.

[13] MarketsandMarkets. Digital Twin Market, 2019. URL https://www.marketsandmarkets.com/Market-Reports/digital-twin-market-225269522.html. Accessed: 2020-05-04.

[14] NIH- U.S. National Library of Medicine. What is DNA?, 2020. URL https://ghr.nlm.nih.gov/primer/basics/howmanychromosomes. Accessed: 2020-05-10.

[15] The National Human Genome Research Institute. Deoxyribonucleic Acid (DNA) Fact Sheet, 2019. URL https://www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet. Accessed: 2020-05-08.

[16] James D Watson and Francis HC Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[17] Hawazin Faiz Badawi, Haiwei Dong, and Abdulmotaleb El Saddik. Mobile cloud-based physical activity advisory system using biofeedback sensors. *Future Generation Computer Systems*, 66:59–70, 2017.

[18] Fedwa Laamarti, Hawazin Faiz Badawi, Yezhe Ding, Faisal Arafsha, Basim Hafidh, and Abdulmotaleb El Saddik. An iso/ieee 11073 standardized digital twin framework for health and well-being in smart cities. *IEEE Access*, 8:105950–105961, 2020.

[19] IEEE Standards Association. 11073-20601-2014 - IEEE Health informatics–Personal health device communication - Part 20601: Application profile- Optimized Exchange Protocol, 2014. URL https://standards.ieee.org/standard/11073-20601-2014.html. Accessed: 2017-08-22.

[20] Hawazin Badawi, Fedwa Laamarti, Faisal Arafsha, and Abdulmotaleb El Saddik. Standardizing a shoe insole based on iso/ieee 11073 personal health device (x73-phd) standards. In *International Conference on Information Technology & Systems*, pages 764–778. Springer, 2019.

[21] Fabien Lareyre, Cédric Adam, Marion Carrier, and Juliette Raffort. Using digital twins for precision medicine in vascular surgery. *Annals of vascular surgery*, 67: e577–e578, 2020.

[22] Puneet Sharma, Michael Suehling, Thomas Flohr, and Dorin Comaniciu. Artificial intelligence in diagnostic imaging: status quo, challenges, and future opportunities. *Journal of thoracic imaging*, 35:S11–S16, 2020.

[23] Katrien De Cocker, Maïté Verloigne, Greet Cardon, and Ragnar Van Acker. Public health communication and education to promote more physical activity and less sedentary behaviour: Development and formative evaluation of the 'physical activity triangle'. *Patient Education and Counseling*, 104(1):75–84, 2021.

[24] Inas S Khayal and Amro M Farid. Designing smart cities for citizen health & well-being. In *2017 IEEE First Summer School on Smart Cities (S3C)*, pages 120–125. IEEE, 2017.

[25] Domenico Mirarchi, Patrizia Vizza, Eugenio Vocaturo, Pietro Hiram Guzzi, and Pierangelo Veltri. A new ict based model for wellness and health care. In *International Workshop on Neural Networks*, pages 243–252. Springer, 2015.

[26] Mahesh Kumar Sharma and Kunwar Singh Vaisla. E-health for rural areas of uttarakhand under e-governance service delivery model. In *2012 1st International Conference on Recent Advances in Information Technology (RAIT)*, pages 622–625. IEEE, 2012.

[27] B Frode Hansen, Espen Bjertness, and Jon Ketil Gronnesby. A socio-ecologic model for periodontal diseases. *Journal of clinical periodontology*, 20(8):584–590, 1993.

[28] Malcolm Clarke, Douglas Bogia, Kai Hassing, Lars Steubesand, Tony Chan, and Deepak Ayyagari. Developing a standard for personal health devices based on 11073. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6174–6176. IEEE, 2007.

[29] Yi Tang, Xiaolian Duan, Ting Fan, Hailing Feng, Wei Jin, and Bozhi Shi. The study of the implementation and the extension of the iso/ieee-11073 standards. In *2017 IEEE Symposium on Computers and Communications (ISCC)*, pages 193–197. IEEE, 2017.

[30] Doo Heon Song, Gwan Hyung Kim, et al. Personal health care management system developed under iso/ieee 11073 with bluetooth hdp. *International Journal of Smart Home*, 8(3):191–196, 2014.

[31] Miguel Galarraga, Luis Serrano, I Martinez, Paula de Toledo, and Melvin Reynolds. Telemonitoring systems interoperability challenge: an updated review of the applicability of iso/ieee 11073 standards for interoperability in telemonitoring. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6161–6165. IEEE, 2007.

[32] EMB/11073 IEEE 11073 Standards Committee. Ieee 11073-20601-2014 - ieee health informatics–personal health device communication - part 20601: Application profile-optimized exchange protocol, 2014. URL https://standards.ieee.org/standard/11073-20601-2014.html. Accessed: 28/12/2018.

[33] ISO. ISO 37120:2018 sustainable cities and communities — indicators for city services and quality of life, 2018. URL https://www.iso.org/standard/68498.html. Accessed: 2020-01-07.

[34] WHO. Noncommunicable diseases report, 2014. URL https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. Accessed: 2019-05-17.

[35] Fedwa Laamarti, Faisal Arafsha, Basim Hafidh, and Abdulmotaleb El Saddik. Automated athlete haptic training system for soccer sprinting. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 303–309. IEEE, 2019.

[36] Mohamad Eid, Umair Saad, and Usama Afzal. A real time vibrotactile biofeedback system for optimizing athlete training. In *2013 IEEE International Symposium on Haptic Audio Visual Environments and Games (HAVE)*, pages 1–6. IEEE, 2013.

[37] Mederic M Hall, Sathish Rajasekaran, Timothy W Thomsen, and Andrew R Peterson. Lactate: friend or foe. *PM&R*, 8:S8–S15, 2016.

[38] L Véronique Billat. Use of blood lactate measurements for prediction of exercise performance and for control of training. *Sports medicine*, 22(3):157–175, 1996.

[39] Oliver Faude, Wilfried Kindermann, and Tim Meyer. Lactate threshold concepts: How valid are they. *Sports Med*, 39(6):469–490, 2009.

[40] George A Brooks. Anaerobic threshold: review of the concept and directions for future research. *Medicine and science in sports and exercise*, 17(1):22–34, 1985.

[41] L. Bruce Gladden. Lactate metabolism: a new paradigm for the third millennium. *The Journal of physiology*, 558(1):5–30, 2004.

[42] Matthew L Goodwin, James E Harris, Andrés Hernández, and L Bruce Gladden. Blood lactate measurements and analysis during exercise: a guide for clinicians. *Journal of diabetes science and technology*, 1(4):558–569, 2007.

[43] W Kindermann, G Simon, and J Keul. The significance of the aerobic-anaerobic transition for the determination of work load intensities during endurance training. *European journal of applied physiology and occupational physiology*, 42(1):25–34, 1979.

[44] H Stegmann and W Kindermann. Comparison of prolonged exercise tests at the individual anaerobic threshold and the fixed anaerobic threshold of 4 mmol· l-1 lactate. *International journal of sports medicine*, 3(02):105–110, 1982.

[45] H Heck, A Mader, Gjetal Hess, S Mücke, R Müller, and W Hollmann. Justification of the 4-mmol/l lactate threshold. *International journal of sports medicine*, 6(03): 117–130, 1985.

[46] Véronique L Billat, Pascal Sirvent, Guillaume Py, Jean-Pierre Koralsztein, and Jacques Mercier. The concept of maximal lactate steady state. *Sports medicine*, 33(6):407–426, 2003.

[47] Karlman Wasserman, Brian J Whipp, SN Koyl, and WL Beaver. Anaerobic threshold and respiratory gas exchange during exercise. *Journal of applied physiology*, 35(2): 236–243, 1973.

[48] Eduardo Marcel Fernandes Nascimento, Maria Augusta Pedutti Dal Molin Kiss, Tony Meireles Santos, Mike Lambert, and Flavio Oliveira Pires. Determination of lactate thresholds in maximal running test by heart rate variability data set. *Asian Journal of Sports Medicine*, 8(3), 2017.

[49] I Llodio, I Garcia-Tabar, L Sánchez-Medina, J Ibáñez, and EM Gorostiaga. Estimation of the maximal lactate steady state in junior soccer players. *International Journal of Sports Medicine*, 36(14):1142–1148, 2015.

[50] Matthew Driller, Nattai Borges, and Daniel Plews. Evaluating a new wearable lactate threshold sensor in recreational to highly trained cyclists. *Sports Engineering*, 19(4): 229–235, 2016.

[51] Benjamin Holfelder, Niklas Brown, and Dieter Bubeck. The influence of sex, stroke and distance on the lactate characteristics in high performance swimming. *PloS one*, 8(10):e77185, 2013.

[52] Ibai Garcia-Tabar, Iñaki Llodio, Luis Sánchez-Medina, Maite Ruesta, Javier Ibañez, and Esteban M Gorostiaga. Heart rate–based prediction of fixed blood lactate thresholds in professional team-sport players. *The Journal of Strength & Conditioning Research*, 29(10):2794–2801, 2015.

[53] LGdM Silva, ME Pacheco, CSG Campbell, V Baldissera, and HG Simões. Comparison between direct and indirect protocols of aerobic fitness evaluation physically active individuals. *Revista Brasileira de Medicina do Esporte*, 11(4):1e–4e, 2005.

[54] A Valizadeh, A Khosravi, and H Azmoon. Fat oxidation rate during and after three exercise intensities in non-athlete young men. *World Appl Sci J*, 15(9):1260, 2011.

[55] Stefan Bircher and Beat Knechtle. Relationship between fat oxidation and lactate threshold in athletes and obese women and men. *Journal of sports science & medicine*, 3(3):174, 2004.

[56] Helmuth Sippel and Matti Möttönen. Combined glucose and lactate values in vitreous humour for postmortem diagnosis of diabetes mellitus. *Forensic science international*, 19(3):217–222, 1982.

[57] Kouichi Kawaji, Yoshikuni Fujita, Yoshitada Yajima, Masuo Shirataka, and Hiroaki Kubo. Usefulness of anaerobic threshold in estimating intensity of exercise for diabetics. *Diabetes research and clinical practice*, 6(4):303–309, 1989.

[58] Ricardo Yukio Asano, Rodrigo Alberto Vieira Browne, Rafael da Costa Sotero, Marcelo Magalhães Sales, José Fernando Vila Nova de Moraes, Carmen Sílvia Grubert Campbell, and Herbert Gustavo Simões. Cycling above rather than below lactate threshold is more effective for nitric oxide release and post-exercise blood pressure reduction in individuals with type-2 diabetes. *Motriz: Revista de Educação Física*, 19(3):633–640, 2013.

[59] EF Coyle, WH Martin, AA Ehsani, JM Hagberg, SA Bloomfield, DR Sinacore, and JO Holloszy. Blood lactate threshold in some well-trained ischemic heart disease patients. *Journal of Applied Physiology*, 54(1):18–23, 1983.

[60] Rodrigo P Simões, Renata G Mendes, Viviane Castello-Simões, Aparecida M Catai, Ross Arena, and Audrey Borghi-Silva. Use of heart rate variability to estimate lactate threshold in coronary artery disease patients during resistance exercise. *Journal of sports science & medicine*, 15(4):649, 2016.

[61] Milena PR Sperling, Rodrigo P Simões, Flávia CR Caruso, Renata G Mendes, Ross Arena, and Audrey Borghi-Silva. Is heart rate variability a feasible method to determine anaerobic threshold in progressive resistance exercise in coronary artery disease? *Brazilian Journal of Physical Therapy*, (AHEAD):0–0, 2016.

[62] Laila CJ Lima, Gabrielle V Assis, Wolysson Hiyane, Wesley S Almeida, Gisela Arsa, Vilmar Baldissera, Carmen SG Campbell, and Herbert G Simões. Hypotensive effects of exercise performed around anaerobic threshold in type 2 diabetic patients. *diabetes research and clinical practice*, 81(2):216–222, 2008.

[63] Rodrigo S Delevatti, Ana Carolina Kanitz, Cristine L Alberton, Elisa Corrêa Marson, Patricia Dias Pantoja, Carolina DertzbocherFeil Pinho, Salime Chedid Lisboa, and Luiz Fernando M Kruel. Glycemic threshold as an alternative method to identify the anaerobic threshold in patients with type 2 diabetes. *Frontiers in physiology*, 9: 1609, 2018.

[64] Herbert Gustavo Simões, Wolysson Carvalho Hiyane, Ronaldo Esch Benford, Bibiano Madrid, Francisco Andriotti Prada, Sérgio Rodrigues Moreira, Fábio Yuzo Nakamura, Ricardo Jacó de Oliveira, and Carmen Sílvia Grubert Campbell. Lactate threshold prediction by blood glucose and rating of perceived exertion in people with type 2 diabetes. *Perceptual and motor skills*, 111(2):365–378, 2010.

[65] Samuel da Silva Aguiar, Caio Victor de Sousa, Marcelo Magalhães Sales, José Morais Souto Filho, Rafael da Costa Sotero, Thiago dos Santos Rosa, and Herbert Gustavo

Simões. Dmax method estimates lactate threshold in individuals with type 2 diabetes. *Sport Sciences for Health*, 12(2):175–181, 2016.

[66] Christian Baumgart, MW Hoppe, and Jürgen Freiwald. Different endurance characteristics of female and male german soccer players. *Biology of sport*, 31(3):227, 2014.

[67] JM Green, TR Crews, AM Bosak, and WW Peveler. Overall and differentiated ratings of perceived exertion at the respiratory compensation threshold: effects of gender and mode. *European journal of applied physiology*, 89(5):445–450, 2003.

[68] M Siahkouhian, S Azizan, and B N Roohi. A new approach for the determination of anaerobic threshold: Methodological survey on the modified Dmax method. *Journal of Human Sport and Exercise*, 7(2):599–607, 2012. doi: 10.4100/jhse.2012.72.23.

[69] Verusca Najara de Carvalho Cunha, Ricardo Moreno Lima, Rafael Rodrigues Da Cunha, Carmen Silvia Grubert Campbell, Sergio Rodrigues Moreira, Eduardo Bodnariuc Fontes, Kaori Celia Sakuma, Herbert Gustavo Simoes, and Guilherme Morais Puga. Individual anaerobic threshold prediction through 1 km and 3 km running performance in young soccer players. *International SportMed Journal*, 15(4):402–414, 2014.

[70] Makoto Ayabe, Takuya Yahiro, Yukari Mori, Kohsaku Takayama, Takuro Tobina, Hiroyuki Higuchi, Kojiro Ishii, Ichiro Sakuma, Yutaka Yoshitake, Hideo Miyazaki, et al. Simple assessment of lactate threshold by means of the bench stepping in older population. *International Journal of Sport and Health Science*, 1(2):207–215, 2003.

[71] Dae-Young Kim and Byoung-Do Seo. Immediate effect of quadriceps kinesio taping on the anaerobic muscle power and anaerobic threshold of healthy college students. *Journal of Physical Therapy Science*, 24(9):919–923, 2012.

[72] Parisa Farzam, Zack Starkweather, and Maria A Franceschini. Validation of a novel wearable, wireless technology to estimate oxygen levels and lactate threshold power in the exercising muscle. *Physiological reports*, 6(7):e13664, 2018.

[73] Nuno Manuel Frade De Sousa, Rodrigo Ferro Magosso, Guilherme Borges Pereira, Richard Diego Leite, Vivian Maria Arakelian, Arlindo Neto Montagnolli, Sérgio Andrade Perez, and Vilmar Baldissera. The measurement of lactate threshold in resistance exercise: a comparison of methods. *Clinical Physiology and Functional Imaging*, 31(5):376–381, 2011.

[74] Jules AAC Heuberger, Pim Gal, Frederik E Stuurman, Wouter AS de Muinck Keizer, Yuri Mejia Miranda, and Adam F Cohen. Repeatability and predictive value of lactate threshold concepts in endurance sports. *PloS one*, 13(11):e0206846, 2018.

[75] Tiago Lazzaretti Fernandes, Rômulo dos Santos Sobreira Nunes, Cesar Cavinato Cal Abad, Andrea Clemente Baptista Silva, Larissa Silva Souza, Paulo Roberto Santos Silva, Cyro Albuquerque, Maria Cláudia Irigoyen, and Arnaldo José Hernandez. Post-analysis methods for lactate threshold depend on training intensity and aerobic capacity in runners. an experimental laboratory study. *Sao Paulo Medical Journal*, 134(3):193–198, 2016.

[76] Arkadiusz Stanula, Tomasz Gabrys, Urszula Szmatlan-Gabrys, Robert Roczniok, Adam Maszczyk, and Przemysław Pietraszewski. Calculating lactate anaerobic thresholds in sports involving different endurance preparation. *Journal of Exercise Science & Fitness*, 11(1):12–18, 2013.

[77] Raffy Dotan. Reverse lactate threshold: a novel single-session approach to reliable high-resolution estimation of the anaerobic threshold. *International journal of sports physiology and performance*, 7(2):141–151, 2012.

[78] Stefan Endler, Christian Secker, and Jörg Bügner. What is the best fitting function? evaluation of lactate curves with common methods from the literature. In *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, pages 185–192. Springer, 2016.

[79] Hugo Cerda-Kohler, Carlos Burgos-Jara, Rodrigo Ramírez-Campillo, Miguel Valdés-Cerda, Eduardo Báez, Daniel Zapata-Gómez, David C Andrade, and Mikel Izquierdo. Analysis of agreement between 4 lactate threshold measurements methods in professional soccer players. *Journal of strength and conditioning research*, 30(10):2864–2870, 2016.

[80] Urtats Etxegarai, Eva Portillo, Jon Irazusta, Ander Arriandiaga, and Itziar Cabanes. Estimation of lactate threshold with machine learning techniques in recreational runners. *Applied Soft Computing*, 63:181–196, 2018.

[81] Silvio R Marques-Neto, Alex S Maior, Geraldo A Maranhão Neto, and Edil L Santos. Analysis of heart rate deflection points to predict the anaerobic threshold by a computerized method. *The Journal of Strength & Conditioning Research*, 26(7): 1967–1974, 2012.

[82] Roohollah Nikooie, Reza Gharakhanlo, Hamid Rajabi, Morteza Bahraminegad, and Ali Ghafari. Noninvasive determination of anaerobic threshold by monitoring the% spo2 changes and respiratory gas exchange. *The Journal of Strength & Conditioning Research*, 23(7):2107–2113, 2009.

[83] Alex Mason, Olga Korostynska, Julien Louis, LE Cordova-Lopez, Badr Abdullah, Jacob Greene, R Connell, and J Hopkins. Noninvasive in-situ measurement of blood lactate using microwave sensors. *IEEE Transactions on Biomedical Engineering*, 65 (3):698–705, 2017.

[84] Huang Shu-Chun, Richard Casaburi, Liao Ming-Feng, Liu Kuo-Cheng, Chen Yu-Jen, Fu Tieh-Cheng, and Su Hong-Ren. Noninvasive prediction of blood lactate through a machine learning-based approach. *Scientific Reports (Nature Publisher Group)*, 9 (1), 2019.

[85] A Erdogan, C Cetin, H Goksu, R Guner, and ML Baydar. Non-invasive detection of the anaerobic threshold by a neural network model of the heart rate—work rate relationship. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 223(3):109–115, 2009.

[86] Hawazin Faiz Badawi, Fedwa Laamarti, Ken Brunet, Ed McNeely, and Abdulmotaleb El Saddik. Non-invasive lactate threshold estimation using machine learning. In *International Conference on Smart Multimedia*, pages 96–104. Springer, 2019.

[87] Zhanlin Ji, Ivan Ganchev, Máirtín O'Droma, Xin Zhang, and Xueji Zhang. A cloud-based x73 ubiquitous mobile healthcare system: design and implementation. *The Scientific World Journal*, 2014, 2014.

[88] Jae-Choong Nam, Won-Kyeong Seo, Jae-Seoung Bae, and You-Ze Cho. Design and development of a u-health system based on the iso/ieee 11073 phd standards. In *The 17th Asia Pacific Conference on Communications*, pages 789–793. IEEE, 2011.

[89] Joon-Ho Lim, Chanyong Park, and Soo-Jun Park. Home healthcare settop-box for senior chronic care using iso/ieee 11073 phd standard. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 216–219. IEEE, 2010.

[90] Aldenor F Martins, Danilo FS Santos, Angelo Perkusich, and Hyggo O Almeida. Upnp and ieee 11073: Integrating personal health devices in home networks. In *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pages 1–6. IEEE, 2014.

[91] Nelia Lasierra, Alvaro Alesanco, and José García. An snmp-based solution to enable remote iso/ieee 11073 technical management. *IEEE Transactions on Information Technology in Biomedicine*, 16(4):709–719, 2012.

[92] Syed Hassan Ahmed, Asanka Sayakkara, Gwanghyeon Kim, and Dongkyun Kim. Self-organized e-health application using ieee 11703: An experimental approach. *Procedia Computer Science*, 32:876–881, 2014.

[93] KSimon Sternly, Kalaiarasi Sonai Muthu Anbananthen, and Seldon Lee. A ubiquitous personal health record (uphr) framework. In *2013 International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013)*. Atlantis Press, 2013.

[94] Wei Li and JongTae Park. Design and implementation of integration architecture of iso 11073 dim with fhir resources using coap. In *2017 International Conference on Information and Communications (ICIC)*, pages 268–273. IEEE, 2017.

[95] Matthias Frohner, Philipp Urbauer, Mathias Forjan, Birgit Pohn, Ferenc Gerbovics, Stefan Sauermann, and Alexander Mense. Development of an android app in compliance with the continua health alliance design guidelines for medical device connectivity in mhealth. *Biomedical Engineering/Biomedizinische Technik*, 57(SI-1-Track-N): 997–999, 2012.

[96] A Fioravanti, G Fico, MT Arredondo, D Salvi, and JL Villalar. Integration of heterogeneous biomedical sensors into an iso/ieee 11073 compliant application. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 1049–1052. IEEE, 2010.

[97] Danilo FS Santos, Kyller C Gorgônio, Angelo Perkusich, and Hyggo O Almeida. A standard-based and context-aware architecture for personal healthcare smart gateways. *Journal of Medical Systems*, 40(10):1–14, 2016.

[98] Alejandro Talaminos, David Naranjo, Gerardo Barbarov, Laura M Roa, and Javier Reina-Tosina. Design and implementation of a standardised framework for the management of a wireless body network in an mobile health environment. *Healthcare Technology Letters*, 4(3):88–92, 2017.

[99] Chanyong Park, Joon-Ho Lim, Ho-youl Jung, and Soojun Park. Iso/ieee 11073 phd standardization of weighting scale using nintendo's wii balance board™ for healthcare services. In *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, pages 195–196. IEEE, 2010.

[100] Merriam-Webster's Dictionary. Proactive, 2020. URL https://www.merriam-webster.com/dictionary/proactive. Accessed: 2018-12-11.

[101] Merriam-Webster's Dictionary. Reactive, 2020. URL https://www.merriam-webster.com/dictionary/reactive. Accessed: 2018-12-11.

[102] Hawazin Badawi and Abdulmotaleb El Saddik. Towards a context-aware biofeedback activity recommendation mobile application for healthy lifestyle. *Procedia Computer Science*, 21:382–389, 2013.

[103] Hawazin Badawi, Mohamad Eid, and Abdulmotaleb El Saddik. Diet advisory system for children using biofeedback sensor. In *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pages 1–4. IEEE, 2012.

[104] Hawazin Badawi, Mohamad Eid, and Abdulmotaleb El Saddik. A real-time biofeedback health advisory system for children care. In *2012 IEEE International Conference on Multimedia and Expo Workshops*, pages 429–434. IEEE, 2012.

[105] KeeHyun Park and SeungHyeon Lim. A multipurpose smart activity monitoring system for personalized health services. *Information Sciences*, 314:240–254, 2015.

[106] Zhuqing Q Xiong, Honghui H Fan, Weizhong Z Wang, Gaosheng S Xie, and Bangyu Y Hwang. Design and development of a 3-lead ecg system based on the iso/ieee 11073-10406 standards. In *International Conference on Health Information Science*, pages 141–147. Springer, 2014.

[107] Jaehyo Jung, Jihwan Lee, Jihoon Lee, and Youn Tae Kim. Development of service network for wearable type acutel myocardial infarction diagnosis system. In *SENSORS, 2013 IEEE*, pages 1–4. IEEE, 2013.

[108] Hyun-Sang Park, Hyun-Young Kim, and Hwa-Sun Kim. Development of standard protocol-based healthcare services for optimized health management. *The Transactions of the Korean Institute of Electrical Engineers*, 67(7):969–975, 2018.

[109] Kazi Masudul Alam and Abdulmotaleb El Saddik. C2ps: A digital twin architecture reference model for the cloud-based cyber-physical systems. *IEEE access*, 5:2050–2062, 2017.

[110] Agusti Solanas, Constantinos Patsakis, Mauro Conti, Ioannis S Vlachos, Victoria Ramos, Francisco Falcone, Octavian Postolache, Pablo A Pérez-Martínez, Roberto Di Pietro, Despina N Perrea, et al. Smart health: A context-aware health paradigm within smart cities. *IEEE Communications Magazine*, 52(8):74–81, 2014.

[111] Jennifer L Kent and Susan Thompson. The three domains of urban planning for health and well-being. *Journal of planning literature*, 29(3):239–256, 2014.

[112] Stéphane Roche and Abbas Rajabifard. Sensing places' life to make city smarter. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 41–46, 2012.

[113] Anind K Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

[114] SQ Dou, HH Zhang, YQ Zhao, AM Wang, YT Xiong, and JM Zuo. Research on construction of spatio-temporal data visualization platform for gis and bim fusion. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:555–563, 2020.

[115] Timo Ruohomäki, Enni Airaksinen, Petteri Huuska, Outi Kesäniemi, Mikko Martikka, and Jarmo Suomisto. Smart city platform enabling digital twin. In *2018 International Conference on Intelligent Systems (IS)*, pages 155–161. IEEE, 2018.

[116] MAXMIND. ISO 3166 Country Codes with Associated Continent, 2021. URL https://dev.maxmind.com/geoip/legacy/codes/country_continent/. Accessed: 2020-08-11.

[117] ISO. ISO 3166, 2021. URL https://www.iso.org/obp/ui/#search/code/. Accessed: 2020-08-11.

[118] Internet Assigned Numbers Authority. iana, 2020. URL https://www.iana.org/. Accessed: 2020-02-10.

[119] Google. Google Maps, 2020. URL https://www.google.com/maps. Accessed: 2020-08-11.

[120] ISO. Iso 37120:2014 sustainable development of communities – indicators for city services and quality of life, 2014. URL https://www.iso.org/standard/62436.html. Accessed: 2017-10-11.

[121] Carlos Mena, Eduardo Fuentes, Yony Ormazábal, Gonzalo Palomo-Vélez, and Iván Palomo. Role of access to parks and markets with anthropometric measurements, biological markers, and a healthy lifestyle. *International Journal of Environmental Health Research*, 25(4):373–383, 2015.

[122] Miltiadis D Lytras and Anna Visvizi. Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. *Sustainability*, 10(6):1998, 2018.

[123] Central Intelligence Agency Internet resource. The World Factbook, 2019. URL https://www.cia.gov/the-world-factbook/about/archives/. Accessed: 2019-12-18.

[124] Marc N Gourevitch, Jessica K Athens, Shoshanna E Levine, Neil Kleiman, and Lorna E Thorpe. City-level measures of health, health determinants, and equity to foster population health improvement: the city health dashboard. *American journal of public health*, 109(4):585–592, 2019.

[125] Technical Working Group on City Health Profiles. *City health profiles: how to report on health in your city*. WHO Regional Office for Europe, 1994. URL https://www.euro.who.int/__data/assets/pdf_file/0009/101061/wa38094ci.pdf. Accessed: 2020-11-11.

[126] Hawazin Faiz Badawi. *A Biofeedback-Based Physical Activity Advisory System*. PhD thesis, Université d'Ottawa/University of Ottawa, 2014.

[127] L Schmitt, T Falck, F Wartena, and D Simons. Novel iso/ieee 11073 standards for personal telehealth systems interoperability. In *2007 Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability (HCMDSS-MDPnP 2007)*, pages 146–148. IEEE, 2007.

[128] James J Campanella, Ledion Bitincka, and John Smalley. Matgat: an application that generates similarity/identity matrices using protein or dna sequences. *BMC bioinformatics*, 4(1):1–4, 2003.

[129] Jonathan Myers and Euan Ashley. Dangerous curves: a perspective on exercise, lactate, and the anaerobic threshold. *Chest*, 111(3):787–795, 1997.

[130] L Véronique Billat. Use of blood lactate measurements for prediction of exercise performance and for control of training. *Sports medicine*, 22(3):157–175, 1996.

[131] Peak Centre for Human Performance, 2019. URL https://www.peakcentre.ca/. Accessed: 2019-08-20.

[132] Fedwa Laamarti, Hawazin Faiz Badawi, Yezhe Ding, Faisal Arafsha, Basim Hafidh, and Abdulmotaleb El Saddik. An iso/ieee 11073 standardized digital twin framework for health and well-being in smart cities. *IEEE Access*, 8:105950–105961, 2020.

[133] Fitness Assessment: Peak Centre, 2020. URL https://www.peakcentre.ca/individual-training/personal-fitness-assessment/. Accessed: 2019-08-20.

[134] Multilayer perceptrons, 2001. URL http://users.ics.aalto.fi/ahonkela/dippa/node41.html. Accessed: 2019-09-12.

[135] World Health Organization. Falls, 2018. URL https://www.who.int/news-room/fact-sheets/detail/falls. Accessed: 2020-11-08.

[136] Joe Verghese, Roee Holtzer, Richard B Lipton, and Cuiling Wang. Quantitative gait markers and incident fall risk in older adults. *The Journals of Gerontology: Series A*, 64(8):896–901, 2009.

[137] Erik Stone, Marjorie Skubic, Marilyn Rantz, Carmen Abbott, and Steve Miller. Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders. *Gait & posture*, 41(1):57–62, 2015.

[138] Denise M Peters, Stacy L Fritz, and Debra E Krotish. Assessing the reliability and validity of a shorter walk test compared with the 10-meter walk test for measurements of gait speed in healthy, older adults. *Journal of geriatric physical therapy*, 36(1): 24–30, 2013.

[139] Sheldon R Simon. Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems. *Journal of biomechanics*, 37(12): 1869–1880, 2004.

[140] Faisal Arafsha, Christina Hanna, Ahmed Aboualmagd, Sarah Fraser, and Abdulmotaleb El Saddik. Instrumented wireless smartinsole system for mobile gait analysis: A validation pilot study with tekscan strideway. *Journal of Sensor and Actuator Networks*, 7(3):36, 2018.

[141] Faisal Arafsha, Fedwa Laamarti, and Abdulmotaleb El Saddik. Development of a wireless cps for gait parameters measurement and analysis. In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5. IEEE, 2018.

[142] EMB/11073 - IEEE 11073 Standards Committee. P11073-10406 - Health Informatics–Personal Health Device Communication Part 10406: Device Specialization–Basic Electrocardiograph (ECG) (1- to 3-lead ECG), 2016. URL https://standards.ieee.org/project/11073-10406.html. Accessed: 2020-11-20.

[143] EMB/11073 IEEE 11073 Standards Committee. IEEE 11073-10404-2008 - IEEE Standard - Health Informatics - Personal Health Device Communication - Part 10404: Device Specialization - Pulse Oximeter, 2010. URL https://standards.ieee.org/standard/11073-10404-2010.html. Accessed: 2020-11-20.

[144] EMB/11073 - IEEE 11073 Standards Committee. IEEE/ISO 11073-10415-2010 - ISO/IEEE International Standard - Health informatics–Personal health device communication–Part 10415: Device specialization–Weighing scale, 2010. URL https://standards.ieee.org/standard/11073-10415-2010.html. Accessed: 2020-11-20.

[145] Talia Salzman, Ahmed Aboualmagd, Hawazin Badawi, Diana Tobón-Vallejo, Hyejun Kim, Lama Dahroug, Fedwa Laamarti, Abdulmotaleb El Saddik, and Sarah Fraser. Prefrontal cortex involvement during dual-task stair climbing in healthy older adults: An fnirs study. *Brain Sciences*, 11(1):71, 2021.

[146] Sarah A Fraser, Olivier Dupuy, Philippe Pouliot, Frédéric Lesage, and Louis Bherer. Comparable cerebral oxygenation patterns in younger and older adults during dual-task walking with increasing load. *Frontiers in aging neuroscience*, 8:240, 2016.

[147] Roy A Wervey, Gerald F Harris, and Jacqueline J Wertsch. Plantar pressure characteristics during stair climbing and descent. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)*, volume 4, pages 1746–1748. IEEE, 1997.

[148] ISO. Sustainable development in communities: City indicators for service delivery and quality of life, 2020. URL https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/en/37120_briefing_note.pdf. Accessed: 28/12/2020.

APPENDICES

Appendix A

Tables

Table A.1: A complete list of top-level domains (TLDs) of all countries developed and maintained by (IANA) [118]

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Afghanistan	.af	Catalonia	.cat	Gaza Strip (Gaza)	.ps
Åland	.ax	Cayman Islands	.ky	Georgia	.ge
Albania	.al	Central African Republic	.cf	Germany	.de
Algeria	.dz	Chad	.td	Ghana	.gh
American Samoa	.as	Chile	.cl	Gibraltar	.gi
Andorra	.ad	China, People's Republic of	.cn	Greece	.gr
Angola	.ao	Christmas Island	.cx	Greenland	.gl
Anguilla	.ai	Cocos (Keeling) Islands	.cc	Grenada	.gd
Antarctica	.aq	Colombia	.co	Guadeloupe	.gp
Antigua and Barbuda	.ag	Comoros	.km	Guam	.gu
Argentina	.ar	Congo	.cd	Guatemala	.gt
Armenia	.am	Congo, Republic of the	.cg	Guernsey	.gg
Aruba	.aw	Cook Islands	.ck	Guinea	.gn
Ascension Island	.ac	Costa Rica	.cr	Guinea-Bissau	.gw
Australia	.au	Côte d'Ivoire (Ivory Coast)	.ci	Guyana	.gy

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Austria	.at	Croatia	.hr	Haiti	.ht
Azerbaijan	.az	Cuba	.cu	Heard Island and McDonald Islands	.hm
Bahamas	.bs	Curaçao	.cw	Honduras	.hn
Bahrain	.bh	Cyprus	.cy	Hong Kong	.hk
Bangladesh	.bd	Czechia (Czech Republic)	.cz	Hungary	.hu
Barbados	.bb	Denmark	.dk	Iceland	.is
Basque Country	.eus	Djibouti	.dj	India	.in
Belarus	.by	Dominica	.dm	Indonesia	.id
Belgium	.be	Dominican Republic	.do	Iran	.ir
Belize	.bz	East Timor (Timor-Leste)	.tl	Iraq	.iq
Benin	.bj	Ecuador	.ec	Ireland	.ie
Bermuda	.bm	Egypt	.eg	Isle of Man	.im
Bhutan	.bt	El Salvador	.sv	Israel	.il
Bolivia	.bo	Equatorial Guinea	.gq	Italy	.it
Bonaire	.bq	Eritrea	.er	Jamaica	.jm
Bosnia and Herzegovina	.ba	Estonia	.ee	Japan	.jp

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Botswana	.bw	Ethiopia	.et	Jersey	.je
Bouvet Island	.bv	European Union	.eu	Jordan	.jo
Brazil	.br	Falkland Islands	.fk	Kazakhstan	.kz
British Indian Ocean Territory	.io	Faeroe Islands	.fo	Kenya	.ke
British Virgin Islands	.vg	Federated States of Micronesia	.fm	Kiribati	.ki
Brunei	.bn	Fiji	.fj	Kuwait	.kw
Bulgaria	.bg	Finland	.fi	Kyrgyzstan	.kg
Burkina Faso	.bf	France	.fr	Laos	.la
Burma	.mm	French Guiana	.gf	Latvia	.lv
Burundi	.bi	French Polynesia	.pf	Lebanon	.lb
Cambodia	.kh	French Southern and Antarctic Lands	.tf	Lesotho	.ls
Cameroon	.cm	Gabon	.ga	Liberia	.lr
Canada	.ca	Galicia	.gal	Libya	.ly
Cape Verde	.cv	Gambia	.gm	Liechtenstein	.li
Lithuania	.lt	Papua New Guinea	.pg	South Sudan	.ss
Luxembourg	.lu	Paraguay	.py	Spain	.es

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Macau	.mo	Peru	.pe	Sri Lanka	.lk
Macedonia, North	.mk	Philippines	.ph	Sudan	.sd
Madagascar	.mg	Pitcairn Islands	.pn	Suriname	.sr
Malawi	.mw	Poland	.pl	Svalbard and Jan Mayen Islands	.sj
Malaysia	.my	Portugal	.pt	Swaziland	.sz
Maldives	.mv	Puerto Rico	.pr	Sweden	.se
Mali	.ml	Qatar	.qa	Switzerland	.ch
Malta	.mt	Romania	.ro	Syria	.sy
Marshall Islands	.mh	Russia	.ru	Taiwan	.tw
Martinique	.mq	Rwanda	.rw	Tajikistan	.tj
Mauritania	.mr	Réunion Island	.re	Tanzania	.tz
Mauritius	.mu	Saba	.bq	Thailand	.th
Mayotte	.yt	Saint Barthélemy	.bl	Togo	.tg
Mexico	.mx	Saint Helena	.sh	Tokelau	.tk
Moldova	.md	Saint Kitts and Nevis	.kn	Tonga	.to
Monaco	.mc	Saint Lucia	.lc	Trinidad & Tobago	.tt
Mongolia	.mn	Saint Martin	.mf	Tunisia	.tn

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Montenegro	.me	Saint-Pierre and Miquelon	.pm	Turkey	.tr
Montserrat	.ms	Saint Vincent and the Grenadines	.vc	Turkmenistan	.tm
Morocco	.ma	Samoa	.ws	Turks and Caicos Islands	.tc
Mozambique	.mz	San Marino	.sm	Tuvalu	.tv
Myanmar	.mm	São Tomé and Príncipe	.st	Uganda	.ug
Namibia	.na	Saudi Arabia	.sa	Ukraine	.ua
Nauru	.nr	Senegal	.sn	United Arab Emirates (UAE)	.ae
Nepal	.np	Serbia	.rs	United Kingdom (UK)	.uk
Netherlands	.nl	Seychelles	.sc	United States of America (USA)	.us
New Caledonia	.nc	Sierra Leone	.sl	United States Virgin Islands	.vi
New Zealand	.nz	Singapore	.sg	Uruguay	.uy
Nicaragua	.ni	Sint Eustatius	.bq	Uzbekistan	.uz
Niger	.ne	Sint Maarten	.sx	Vanuatu	.vu
Nigeria	.ng	Slovakia	.sk	Vatican City	.va

Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)	Country / Territory	Country Code (TLD)
Niue	.nu	Slovenia	.si	Venezuela	.ve
Norfolk Island	.nf	Solomon Islands	.sb	Vietnam	.vn
North Cyprus	.nc	Somalia	.so	Wallis and Futuna	.wf
North Korea	.kp	Somaliland	.so	Western Sahara	.eh
North Macedonia	.mk	South Africa	.za	Yemen	.ye
Northern Mariana	.mp	South Georgia and the South Sandwich Islands	.gs	Zambia	.zm
Norway	.no	South Korea	.kr	Zimbabwe	.zw

Table A.2: Year-code to be used in the G base of the proposed DT-DNA model

Year	Code	Year	Code	Year	Code	Year	Code	Year	Code
2000	YAA	2020	YAX	2040	YBT	2060	YCQ	2080	YDM
2001	YAB	2021	YAY	2041	YBV	2061	YCR	2081	YDN
2002	YAC	2022	YAZ	2042	YBW	2062	YCS	2082	YDP
2003	YAD	2023	YBA	2043	YBX	2063	YCT	2083	YDQ
2004	YAE	2024	YBB	2044	YBY	2064	YCV	2084	YDR
2005	YAF	2025	YBC	2045	YBZ	2065	YCW	2085	YDS
2006	YAG	2026	YBD	2046	YCA	2066	YCX	2086	YDT
2007	YAH	2027	YBE	2047	YCB	2067	YCY	2087	YDV
2008	YAI	2028	YBF	2048	YCC	2068	YCZ	2088	YDW
2009	YAK	2029	YBG	2049	YCD	2069	YDA	2089	YDX
2010	YAL	2030	YBH	2050	YCE	2070	YDB	2090	YDY
2011	YAM	2031	YBI	2051	YCF	2071	YDC	2091	YDZ
2012	YAN	2032	YBK	2052	YCG	2072	YDD	2092	YEA
2013	YAP	2033	YBL	2053	YCH	2073	YDE	2093	YEB
2014	YAQ	2034	YBM	2054	YCI	2074	YDF	2094	YEC
2015	YAR	2035	YBN	2055	YCK	2075	YDG	2095	YED
2016	YAS	2036	YBP	2056	YCL	2076	YDH	2096	YEE
2017	YAT	2037	YBQ	2057	YCM	2077	YDI	2097	YEF
2018	YAV	2038	YBR	2058	YCN	2078	YDK	2098	YEG
2019	YAW	2039	YBS	2059	YCP	2079	YDL	2099	YEH

Table A.3: Government Type to be used in Authority (A) base in the suggested DT-DNA model

Government Type	Code	Government Type	Code
Absolute monarchy	AM	Federal republic	FR
Anarchy	AN	Islamic republic	IR
Authoritarian	AU	Maoism	MA
Commonwealth	CW	Marxism	MX
Communist	CO	Marxism-Leninism	ML
Confederacy (Confederation)	CN	Monarchy	MO
Constitutional	CS	Oligarchy	OL
Constitutional democracy	CD	Parliamentary democracy	PD
Constitutional monarchy	CM	Parliamentary government	PG
Democracy	DM	Presidential	PR
Democratic republic	DR	Republic	RP
Dictatorship	DC	Socialism	SO
Ecclesiastical	EC	Sultanate	SU
Emirate	EM	Theocracy	TC
Federal (Federation)	FD	Totalitarian	TO

Table A.4: Proposed 2-alphabet code for numbers from 0 to 100 to be used in the proposed DT-DNA model

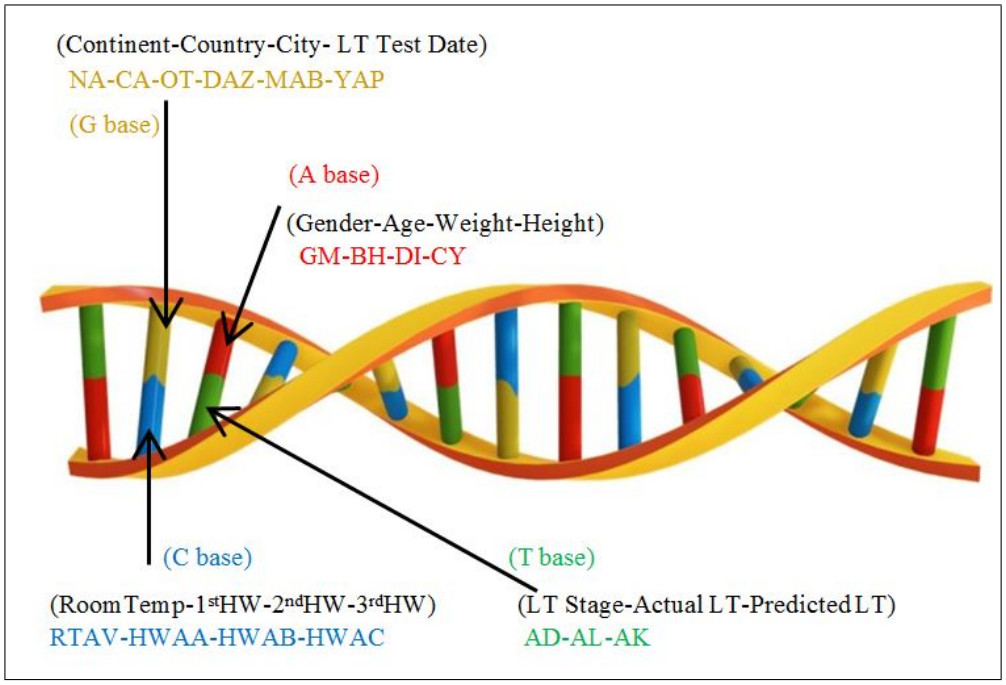
Number	Code	Number	Code	Number	Code	Number	Code	Number	Code
0	ZZ	20	AW	40	BS	60	CP	80	DL
1	AA	21	AX	41	BT	61	CQ	81	DM
2	AB	22	AY	42	BV	62	CR	82	DN
3	AC	23	AZ	43	BW	63	CS	83	DP
4	AD	24	BA	44	BX	64	CT	84	DQ
5	AE	25	BB	45	BY	65	CV	85	DR
6	AF	26	BC	46	BZ	66	CW	86	DS
7	AG	27	BD	47	CA	67	CX	87	DT
8	AH	28	BE	48	CB	68	CY	88	DV
9	AI	29	BF	49	CC	69	CZ	89	DW
10	AK	30	BG	50	CD	70	DA	90	DX
11	AL	31	BH	51	CE	71	DB	91	DY
12	AM	32	BI	52	CF	72	DC	92	DZ
13	AN	33	BK	53	CG	73	DD	93	EA
14	AP	34	BL	54	CH	74	DE	94	EB
15	AQ	35	BM	55	CI	75	DF	95	EC
16	AR	36	BN	56	CK	76	DG	96	ED
17	AS	37	BP	57	CL	77	DH	97	EF
18	AT	38	BQ	58	CM	78	DI	98	EG
19	AV	39	BR	59	CN	79	DK	99	EH
								100	EI

Table A.5: Proposed code for months from January to December to be used in the proposed DT-DNA model

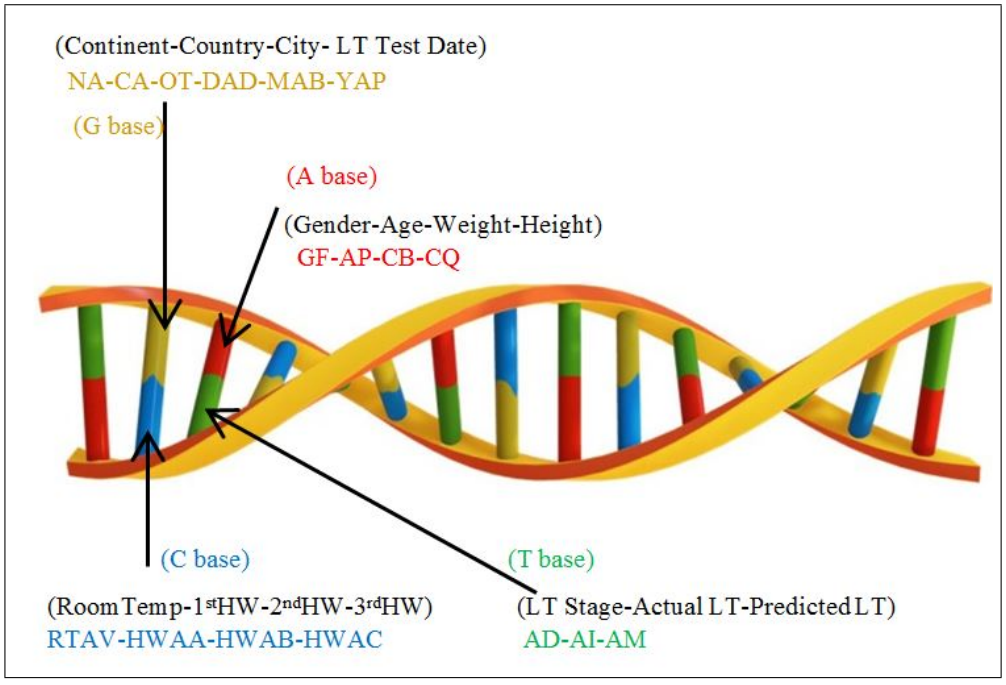
Month	**Code**
January	MAA
February	MAB
March	MAC
April	MAD
May	MAE
June	MAF
July	MAG
August	MAH
September	MAI
October	MAK
November	MAL
December	MAM

Appendix B

Visualization of built DT-DNAs for subjects of Case Study 1 (Section 4.1)

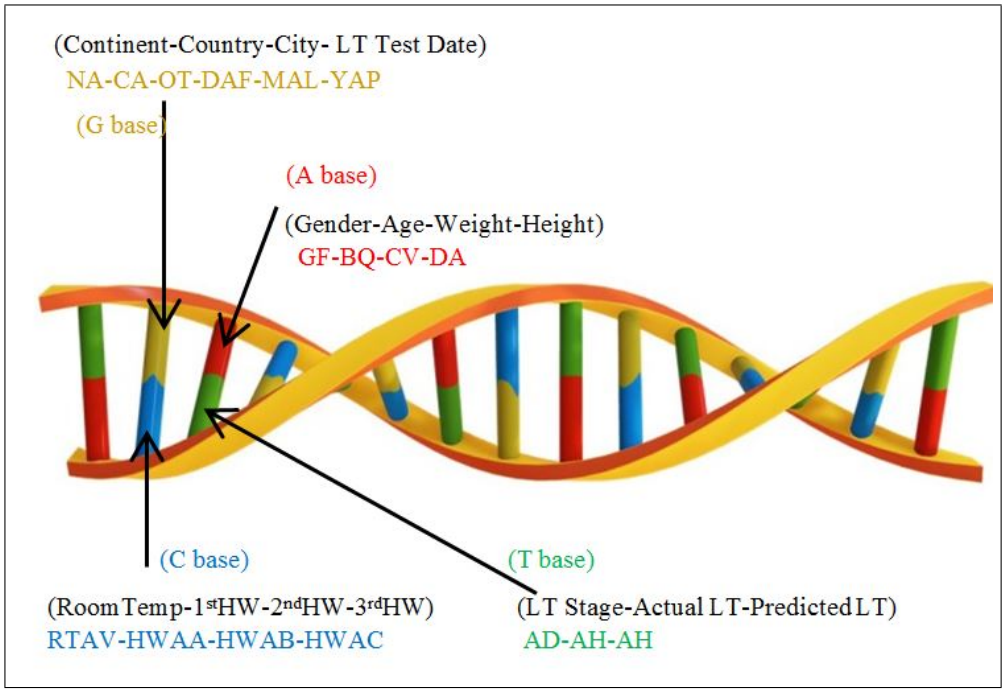


Subject 1

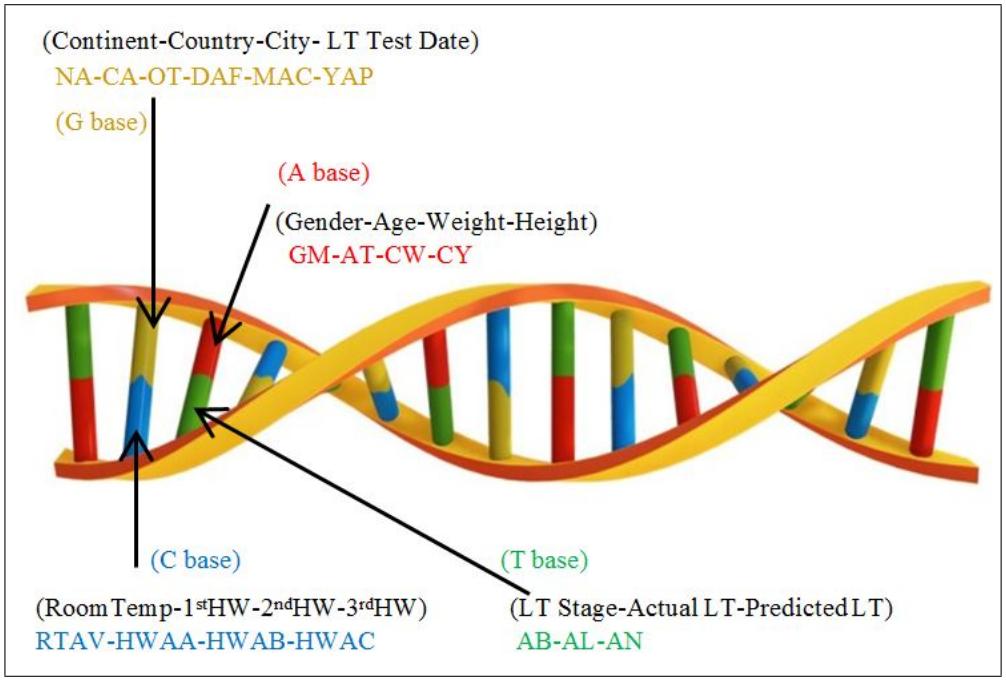


Subject 2

Figure B.1: DT-DNA visualization for Subject 1 and Subject 2 of the experimental sample

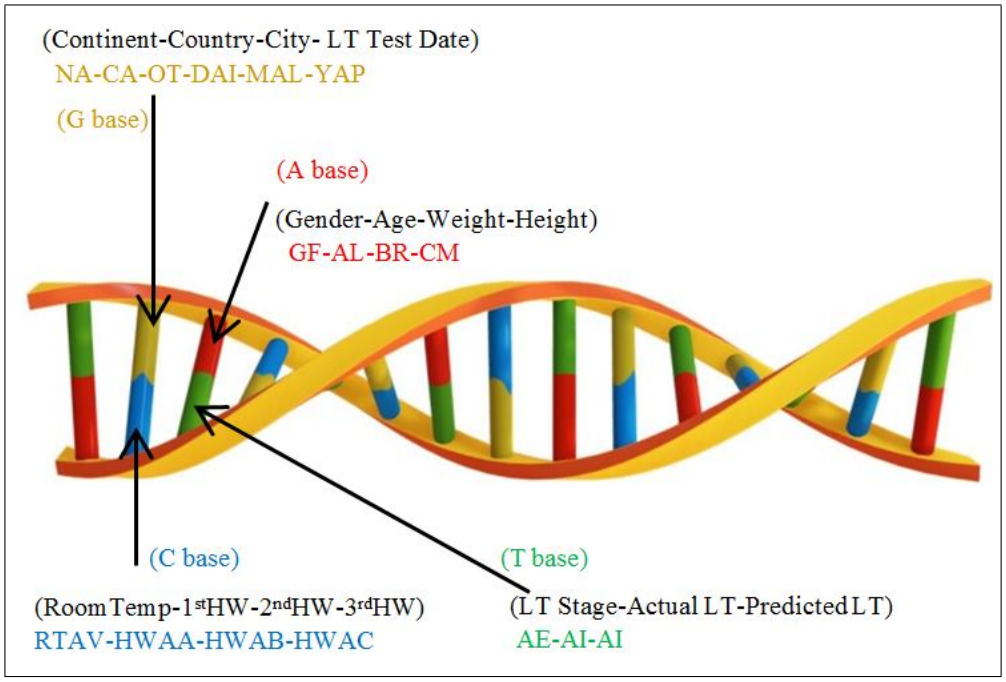


Subject 3

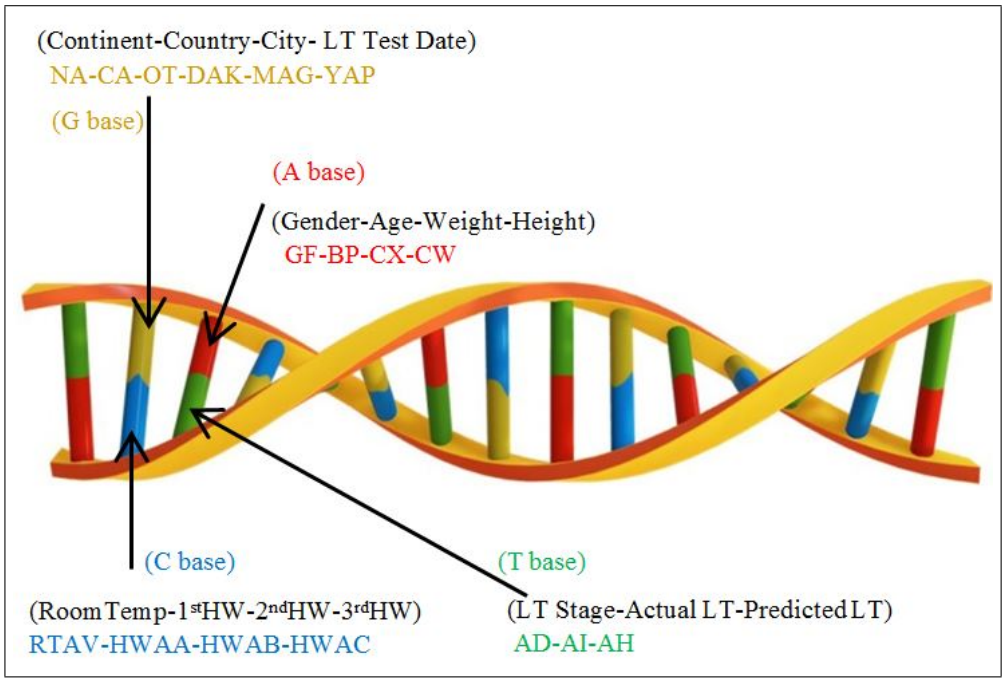


Subject 4

Figure B.2: DT-DNA visualization for Subject 3 and Subject 4 of the experimental sample

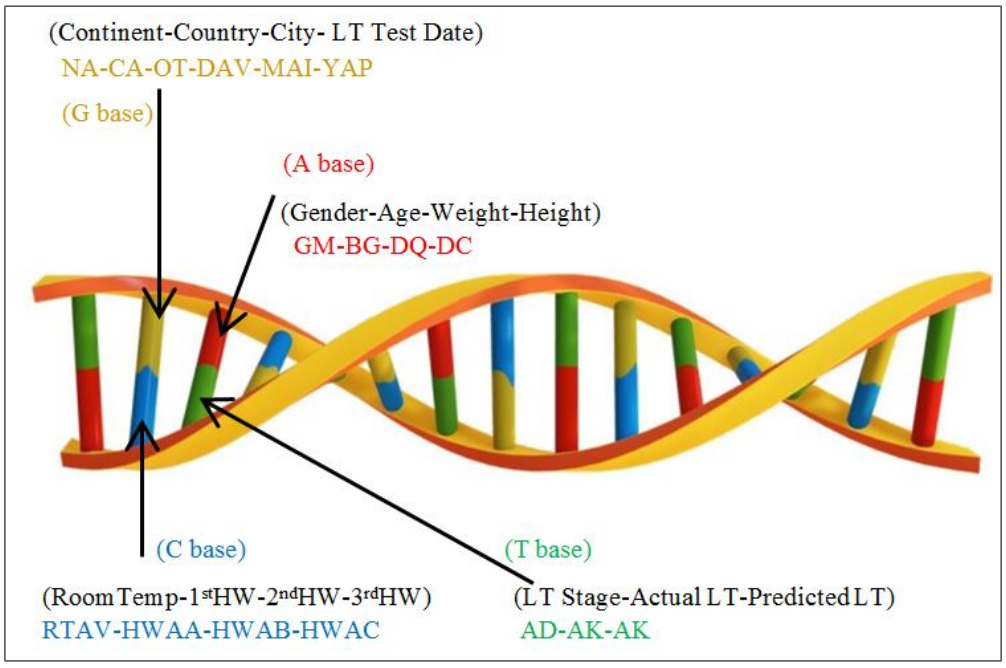


Subject 5

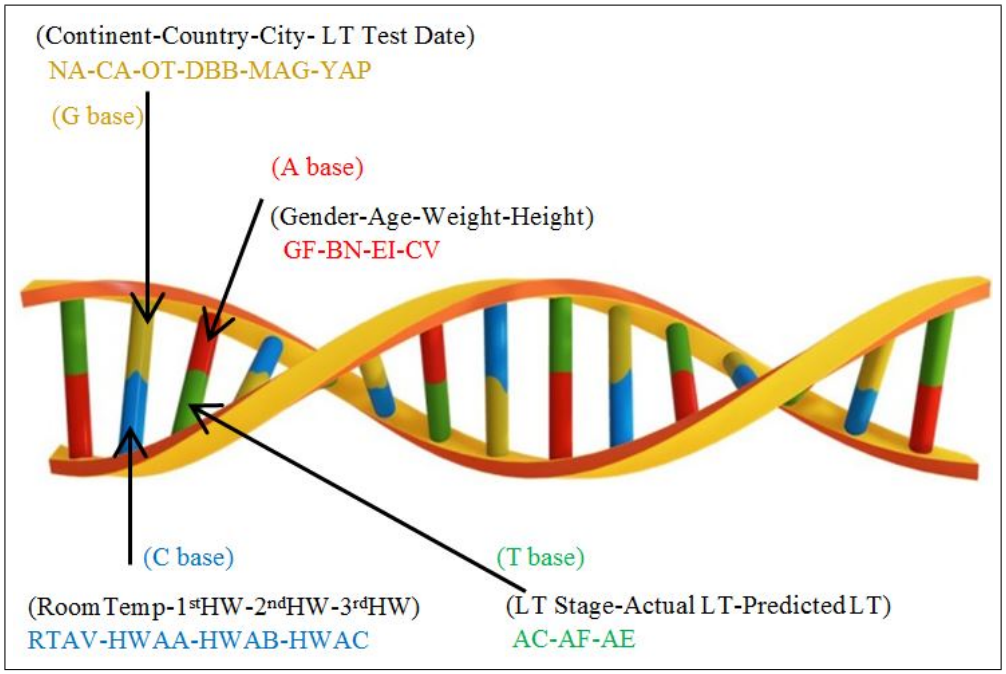


Subject 6

Figure B.3: DT-DNA visualization for Subject 5 and Subject 6 of the experimental sample

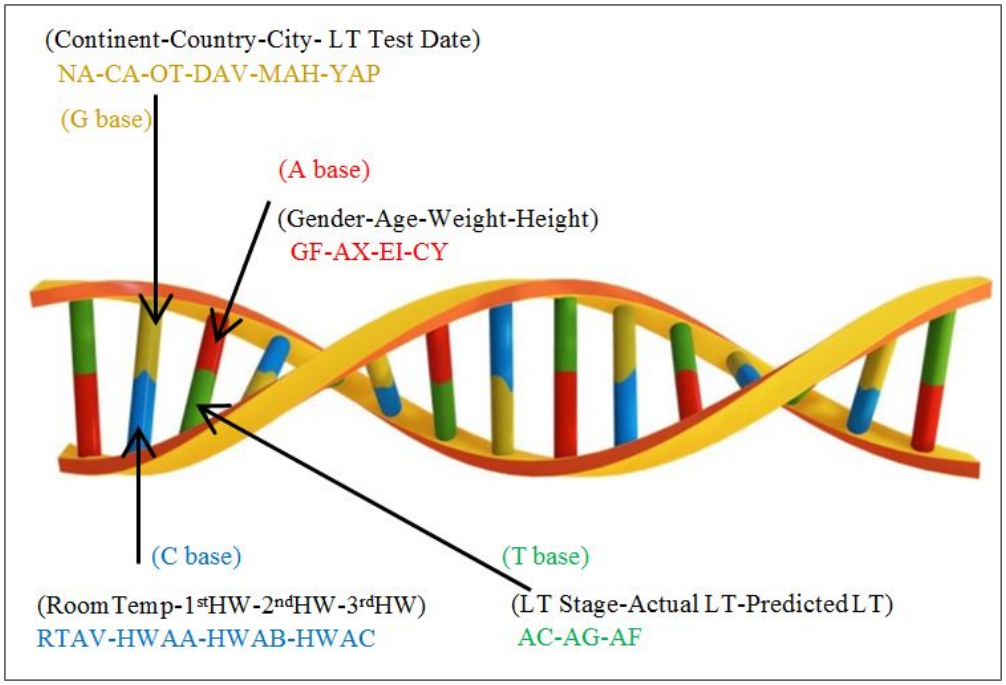


Subject 7

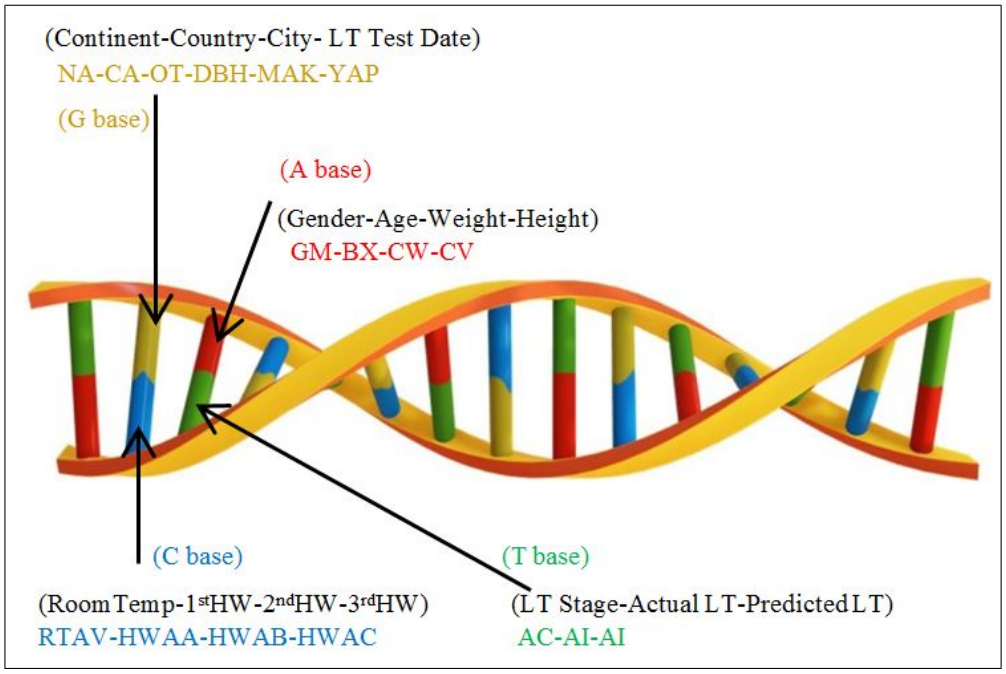


Subject 8

Figure B.4: DT-DNA visualization for Subject 7 and Subject 8 of the experimental sample

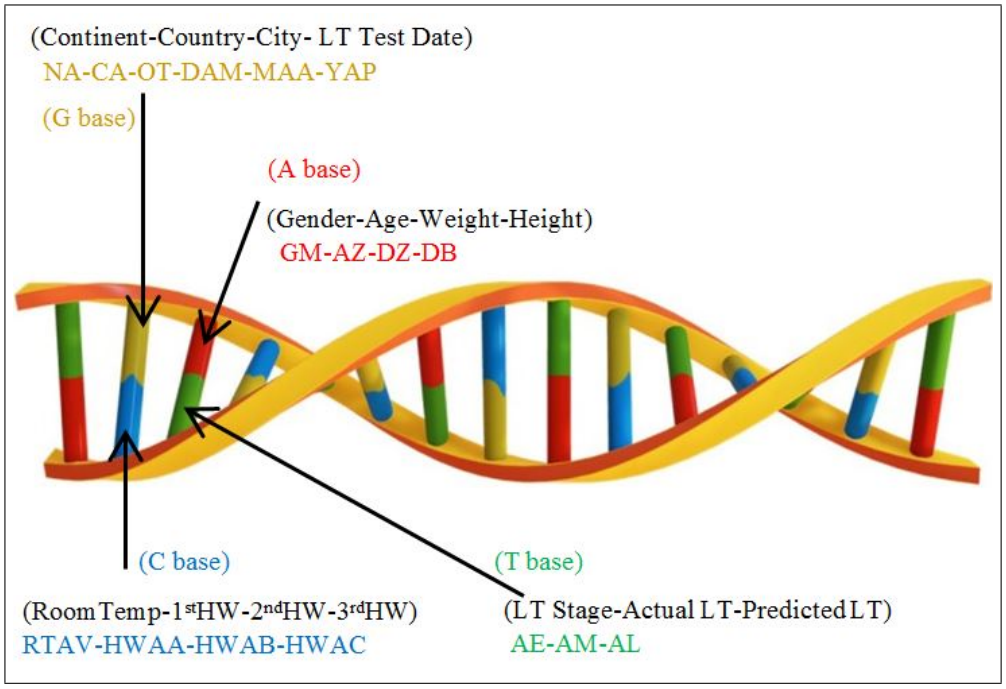


Subject 9

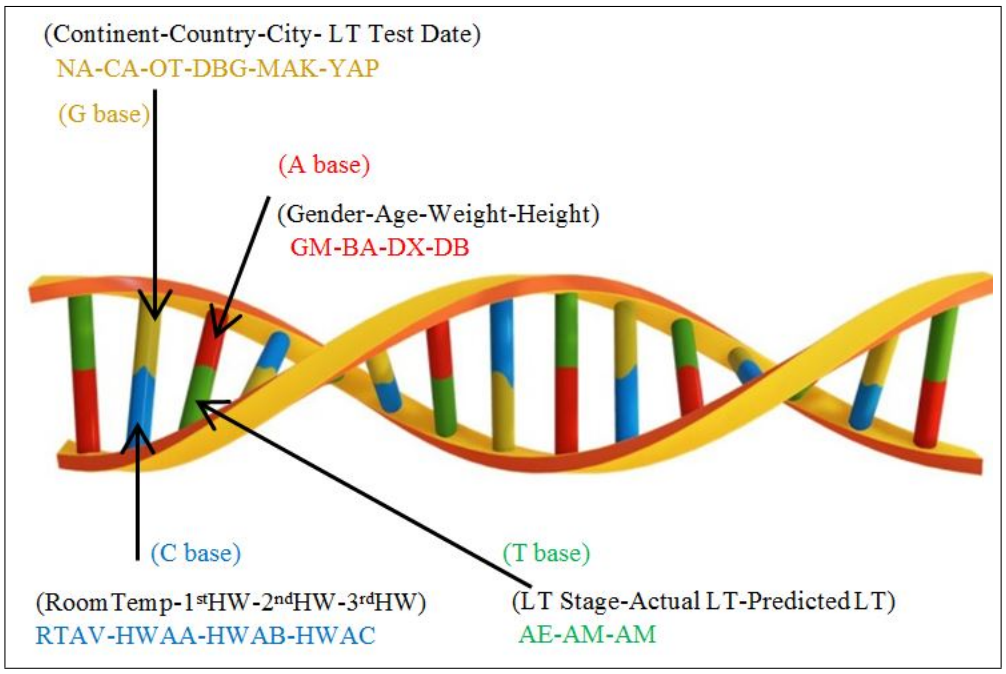


Subject 10

Figure B.5: DT-DNA visualization for Subject 9 and Subject 10 of the experimental sample

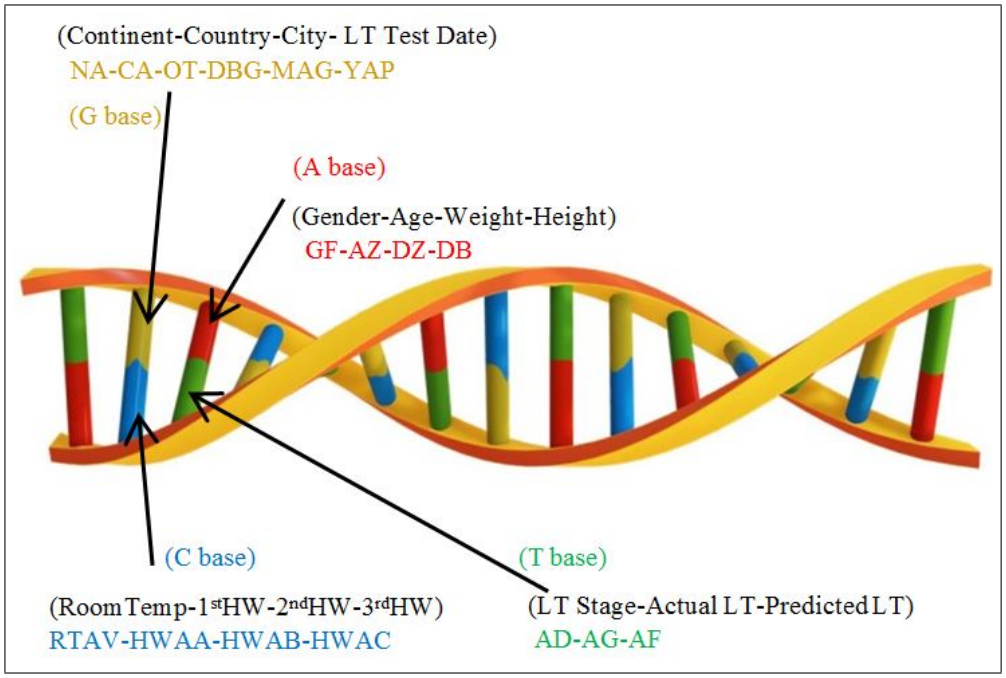


Subject 11a

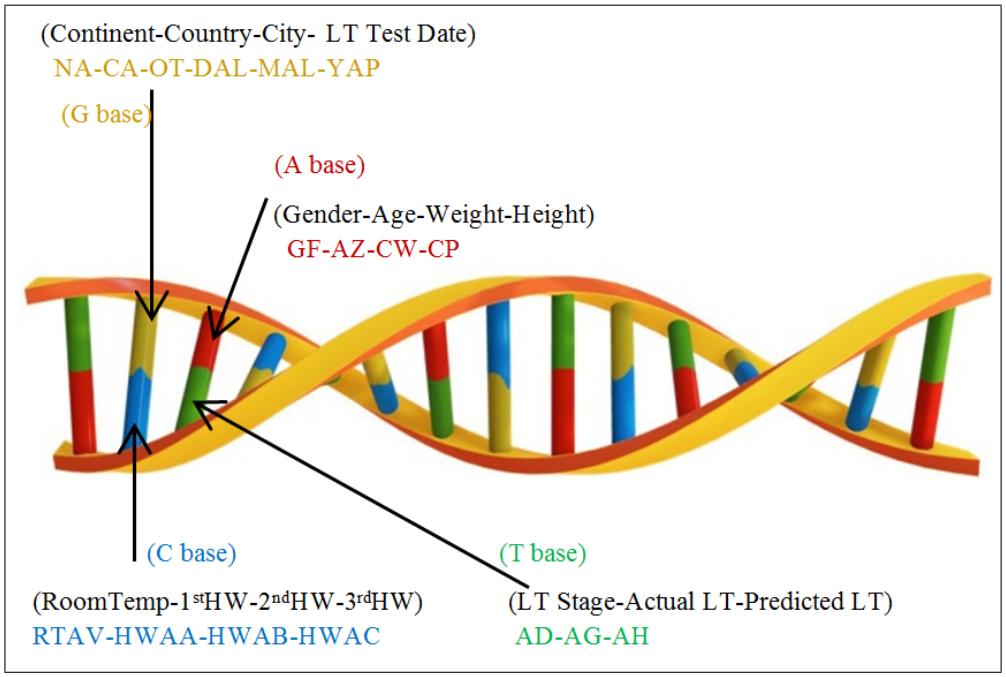


Subject 11b

Figure B.6: DT-DNA visualization for Subject 11a and Subject 11b of the experimental sample

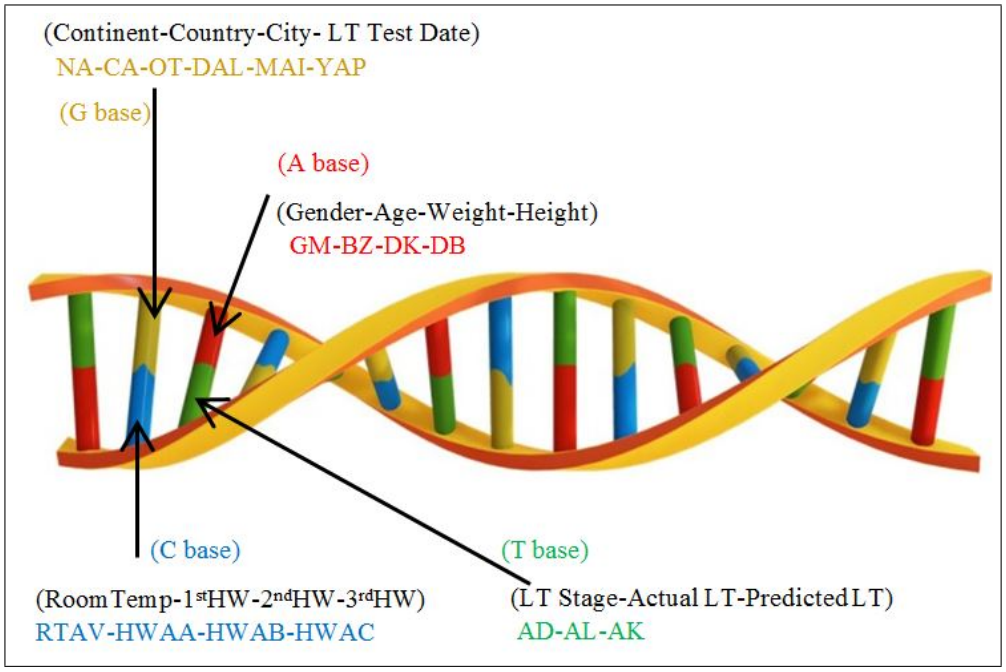


Subject 12a

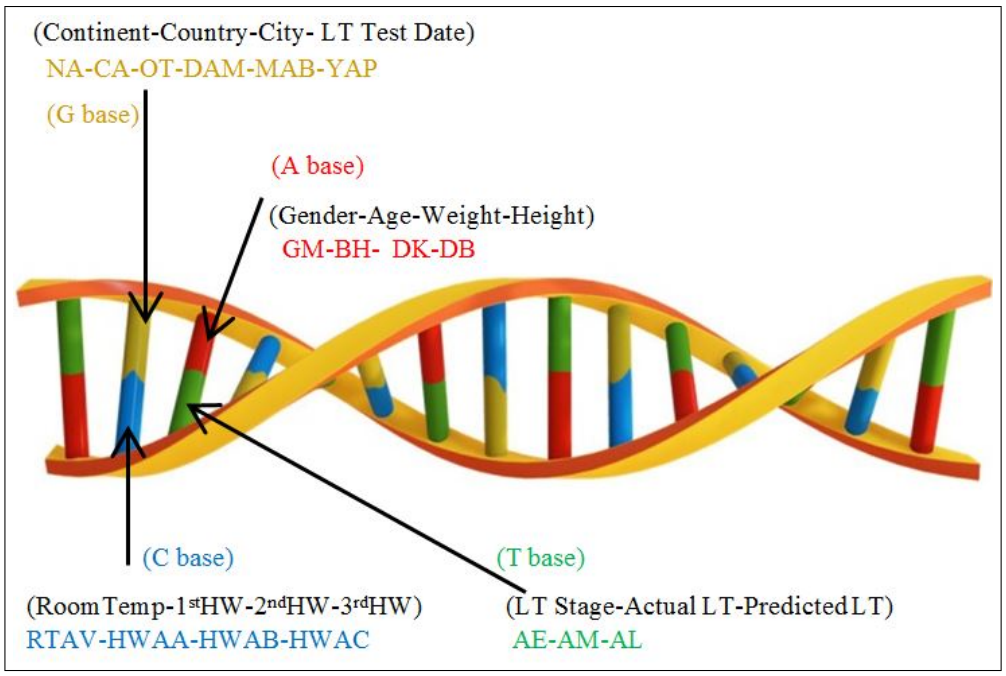


Subject 12b

Figure B.7: DT-DNA visualization for Subject 12a and Subject 12b of the experimental sample

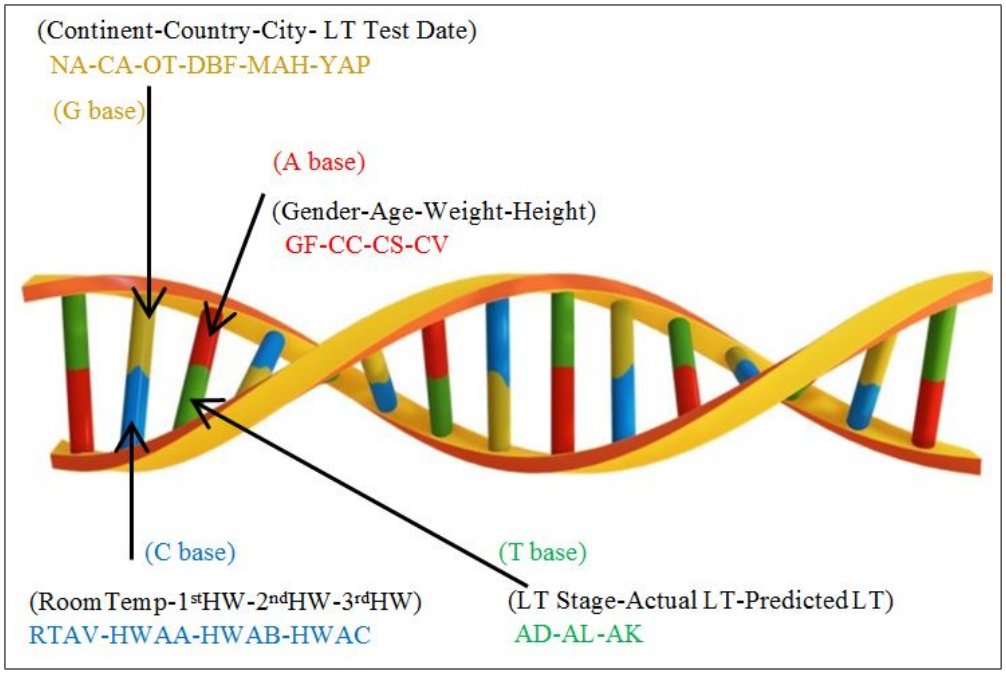


Subject 13a

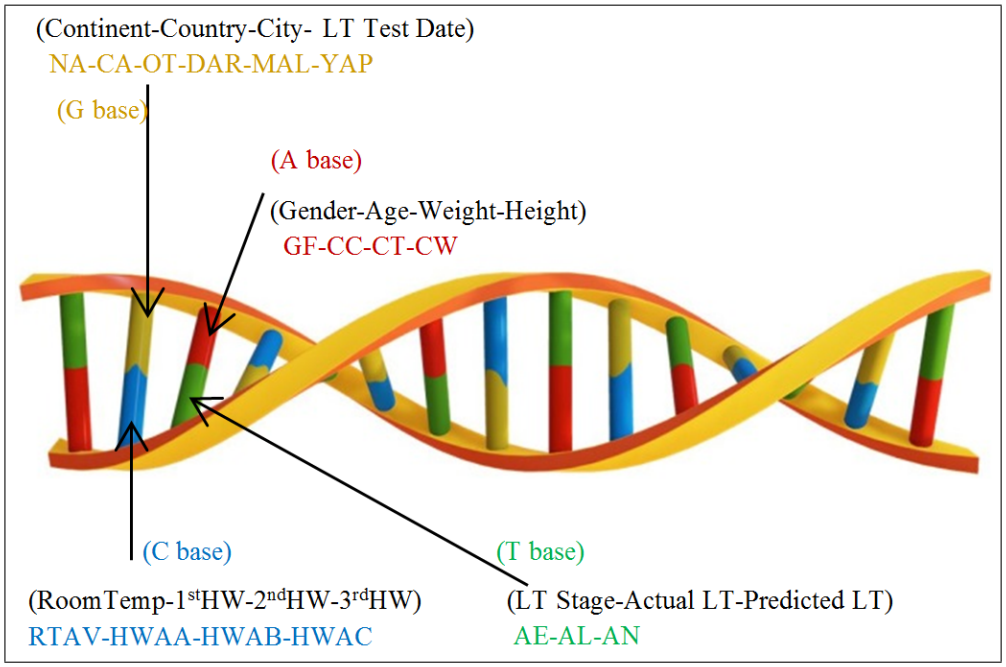


Subject 13b

Figure B.8: DT-DNA visualization for Subject 13a and Subject 13b of the experimental sample

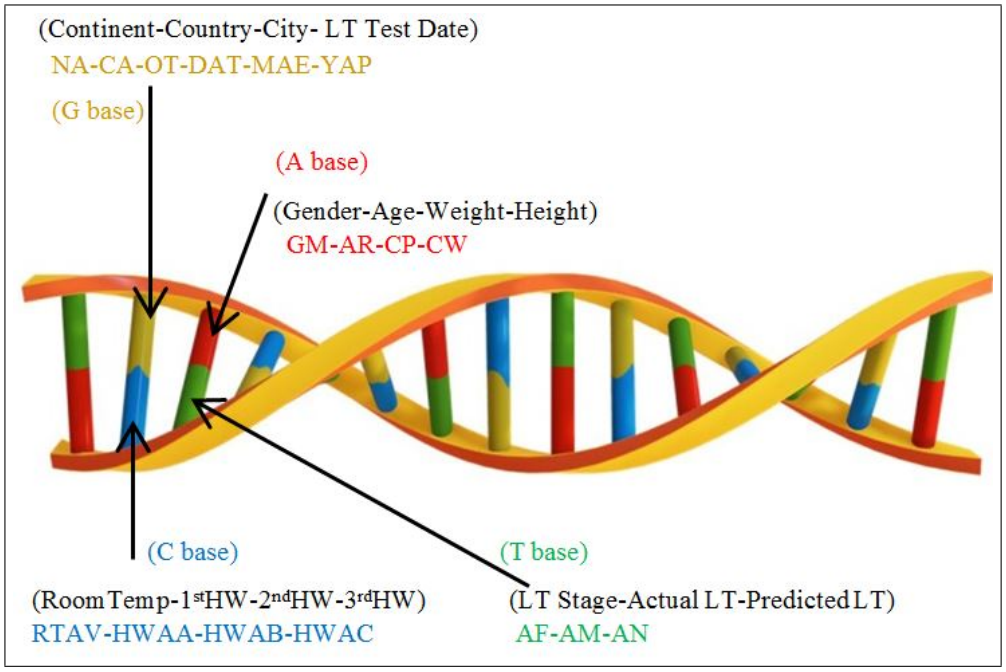


Subject 14a

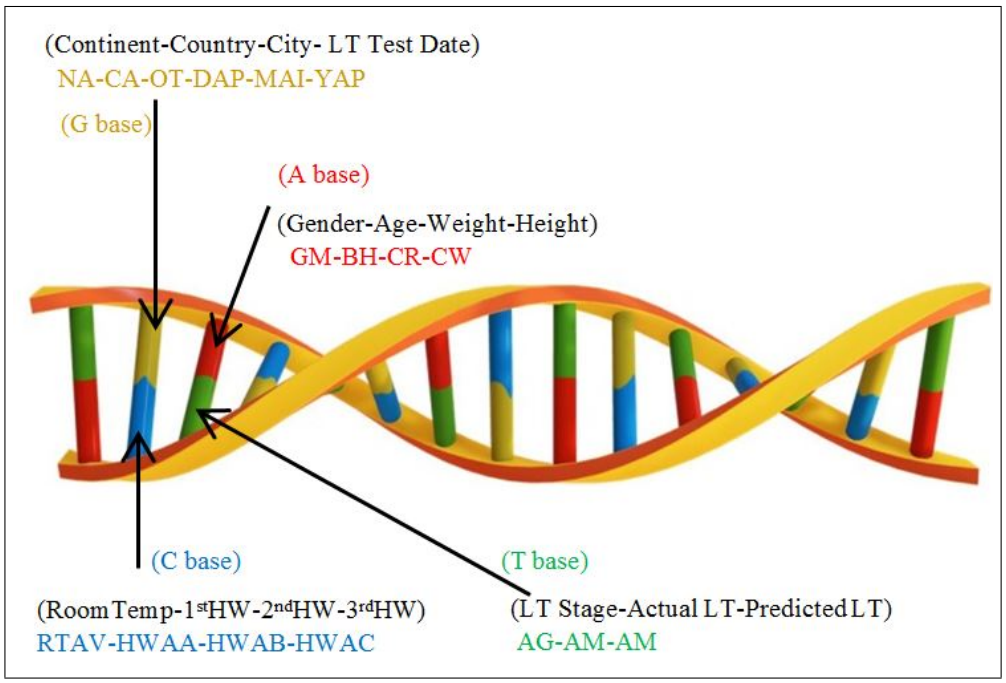


Subject 14b

Figure B.9: DT-DNA visualization for Subject 14a and Subject 14b of the experimental sample



Subject 15a

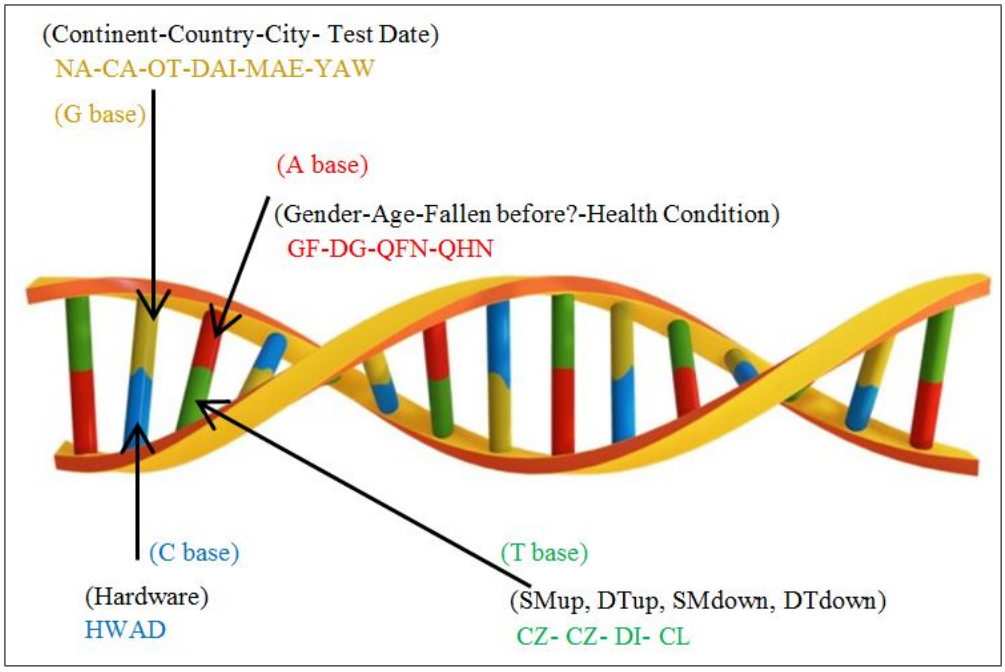


Subject 15b

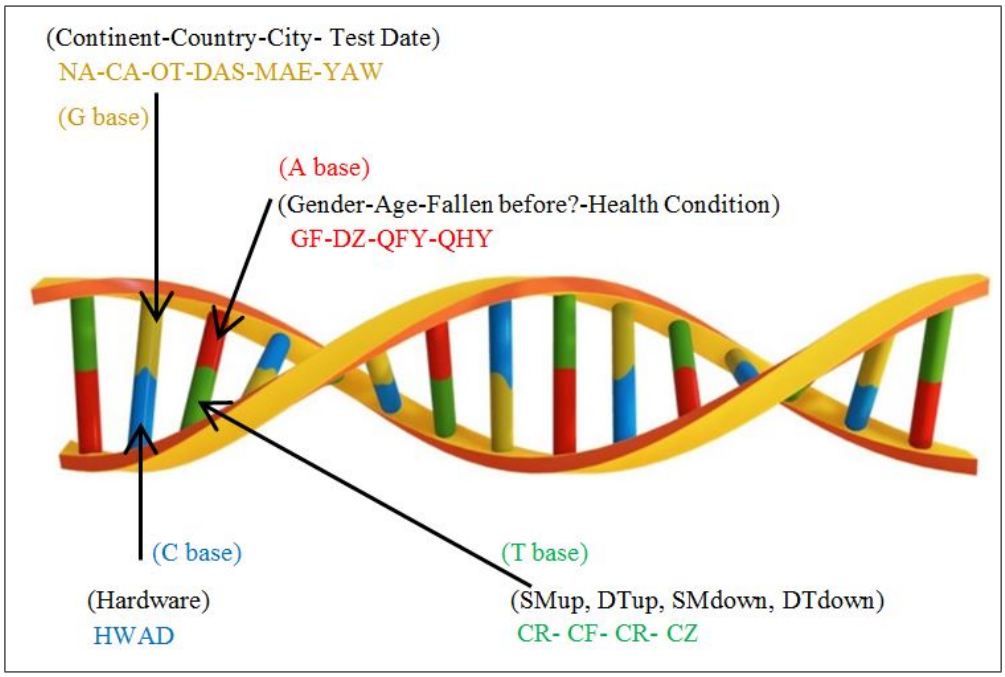
Figure B.10: DT-DNA visualization for Subject 15a and Subject 15b of the experimental sample

Appendix C

Visualization of built DT-DNAs for participants in Case Study 2 (Section 4.2)

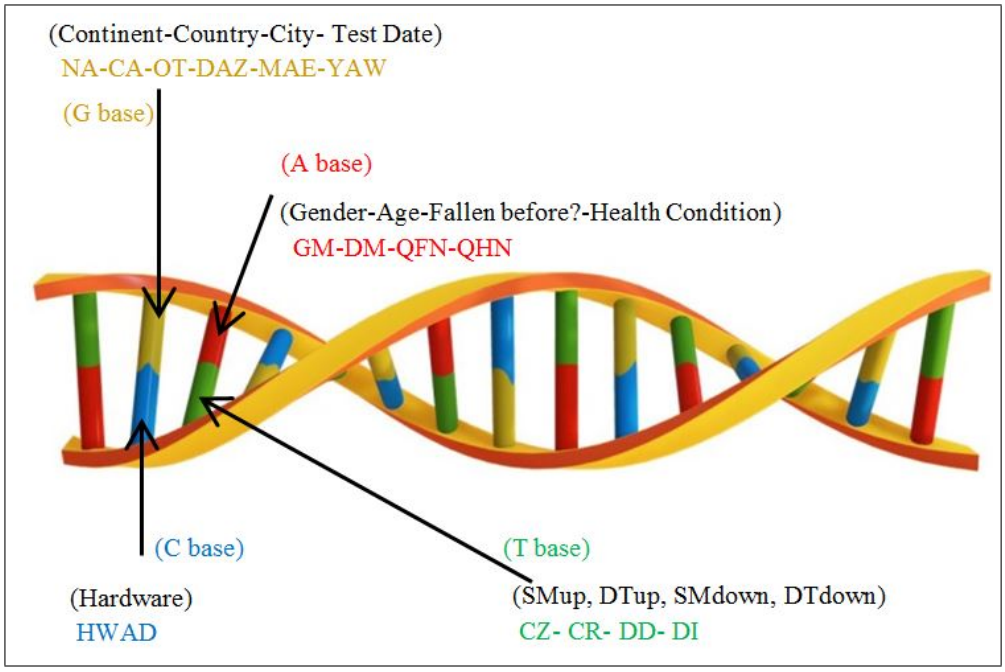


Participant 1

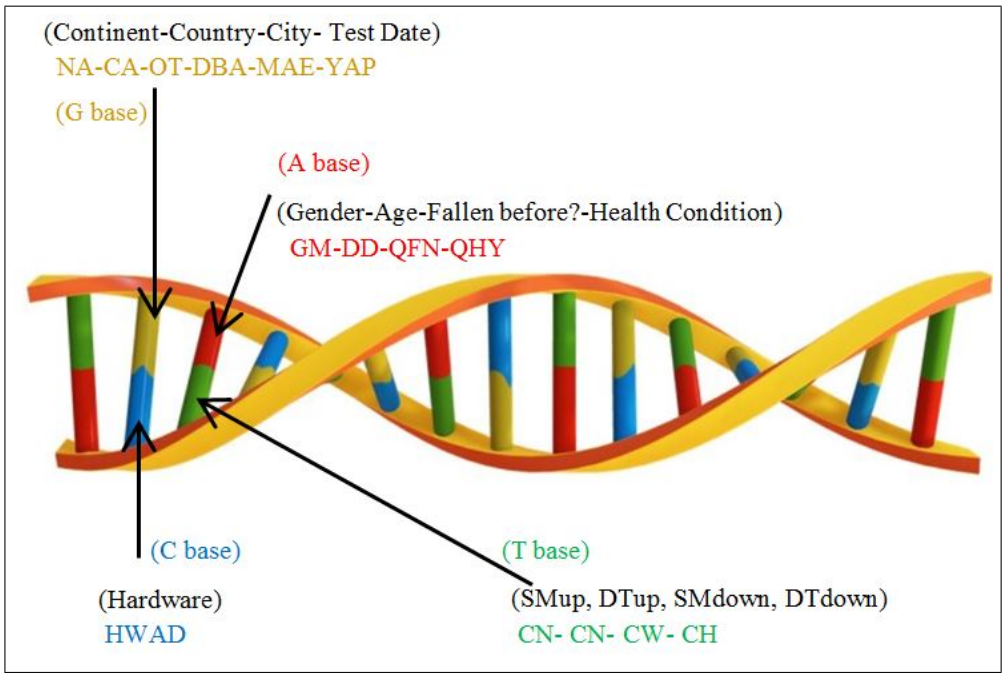


Participant 2

Figure C.1: DT-DNA visualization for Participant 1 and Participant 2 of this study

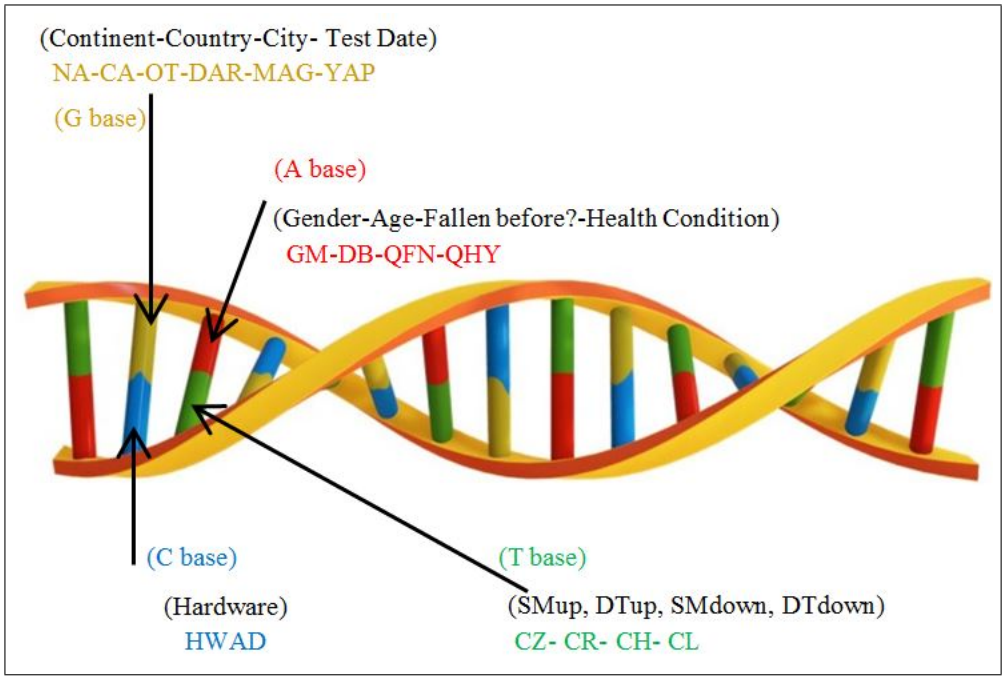


Participant 3

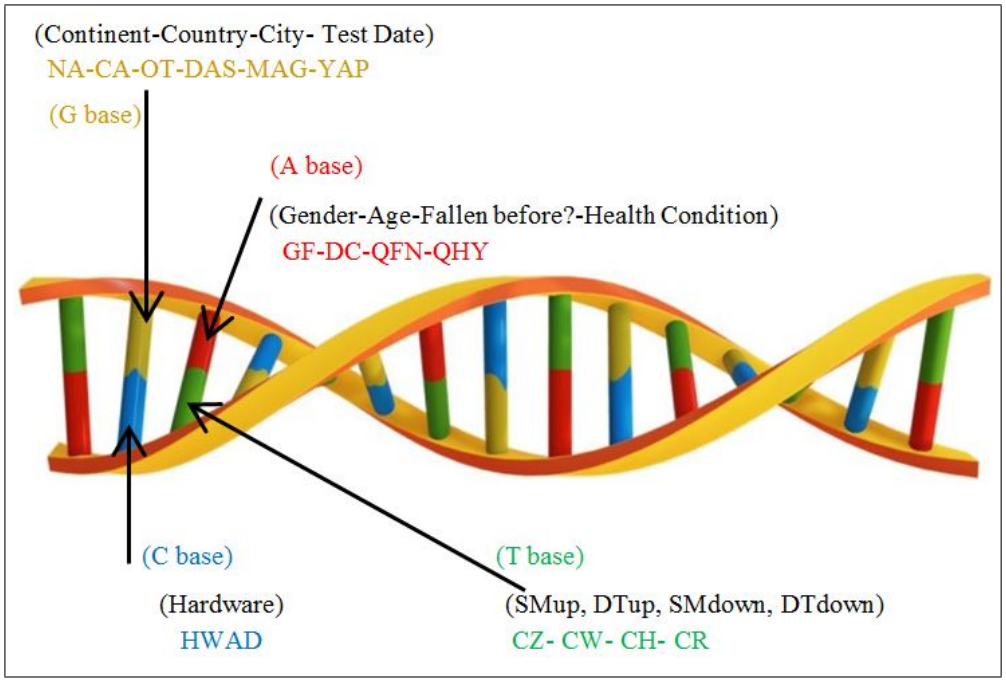


Participant 4

Figure C.2: DT-DNA visualization for Participant 3 and Participant 4 of this study

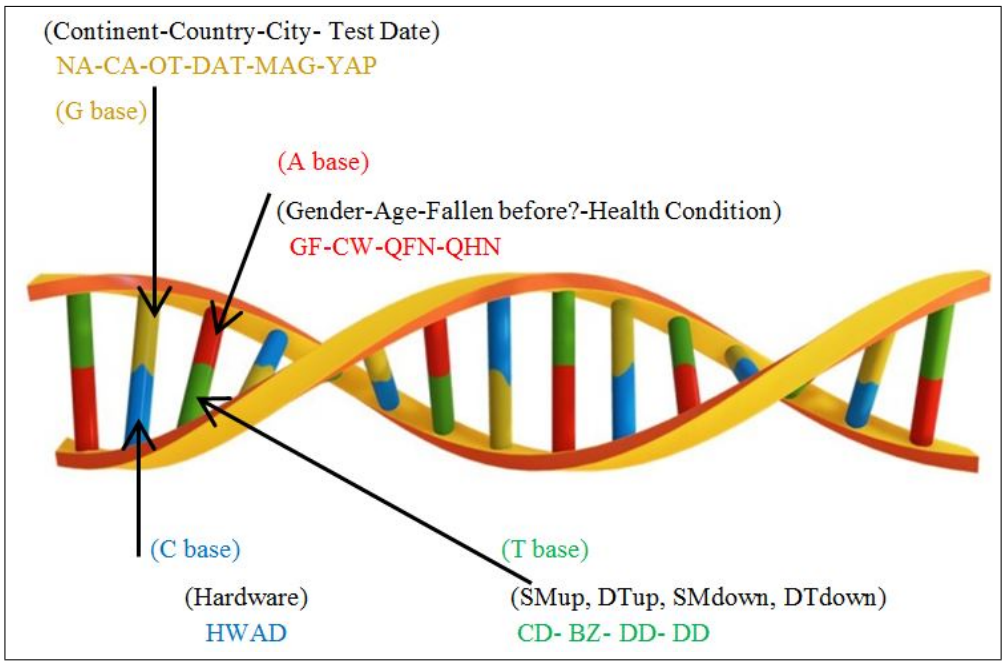


Participant 5

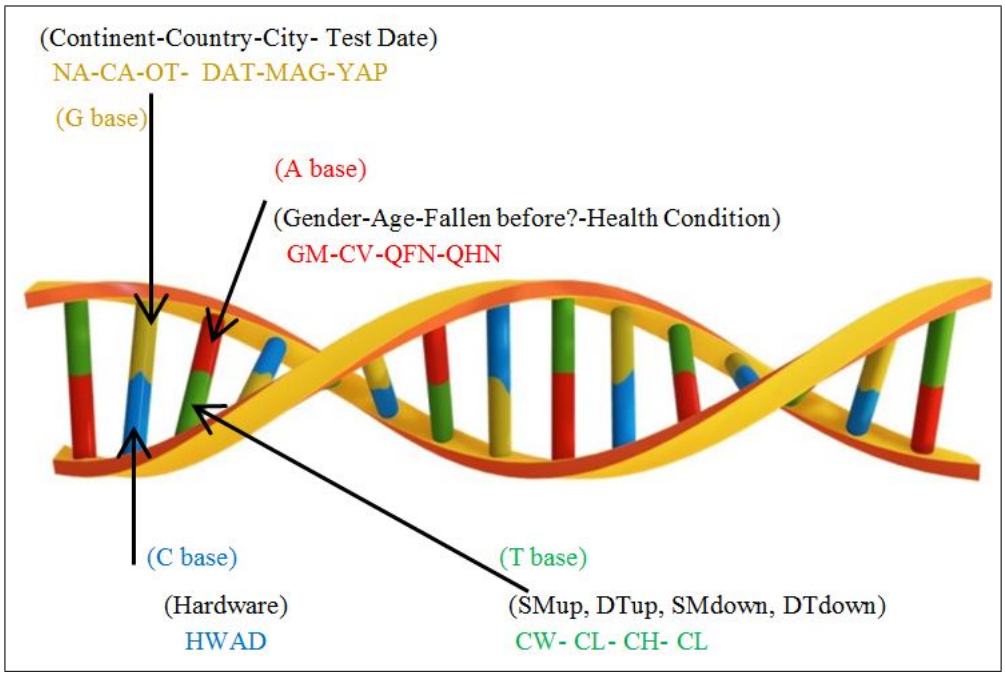


Participant 6

Figure C.3: DT-DNA visualization for Participant 5 and Participant 6 of this study

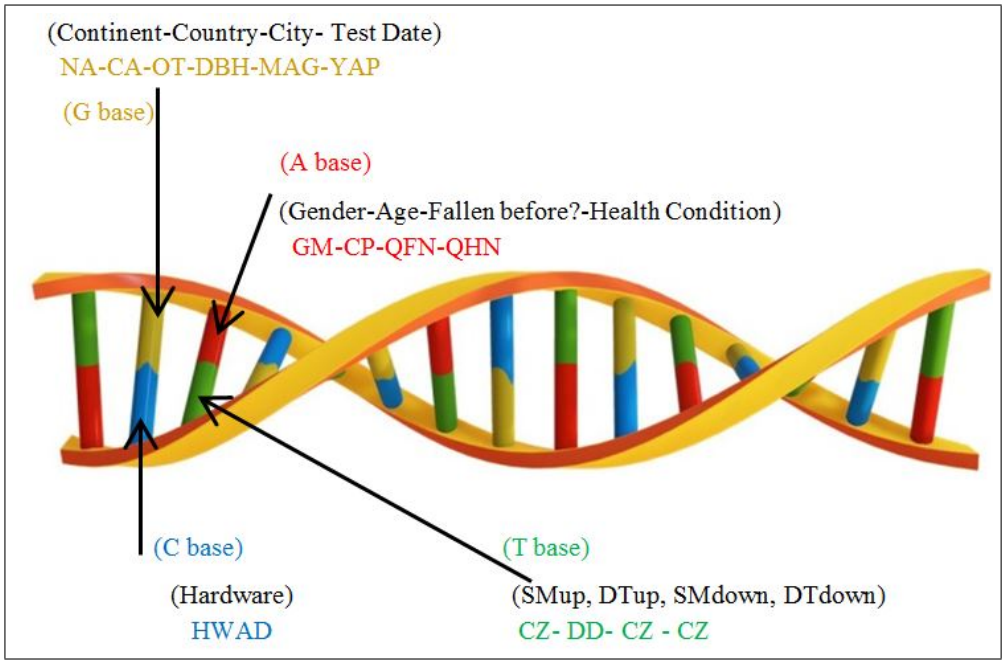


Participant 7

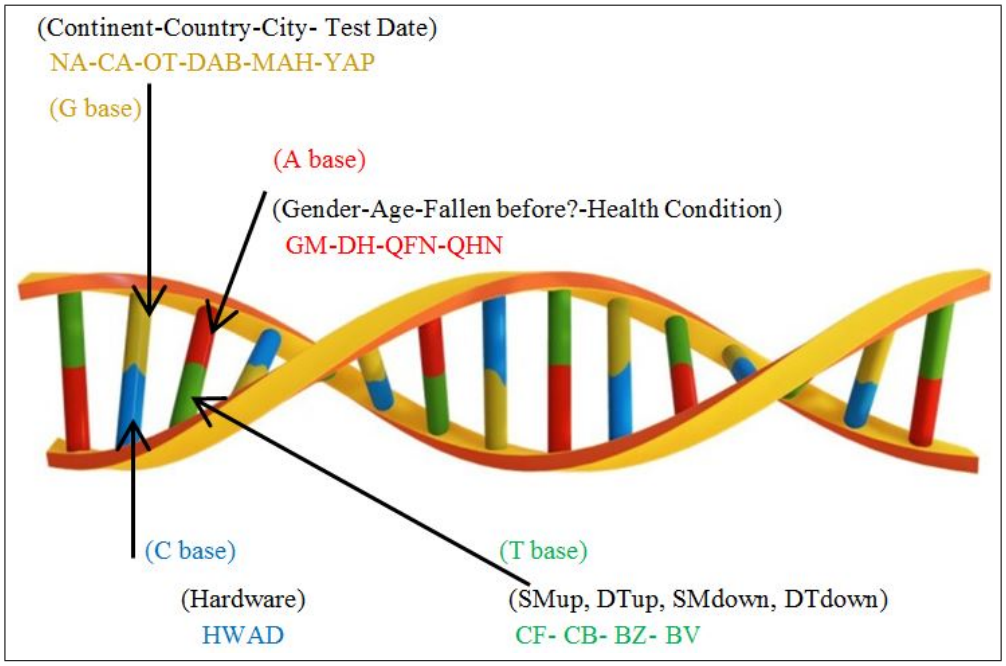


Participant 8

Figure C.4: DT-DNA visualization for Participant 7 and Participant 8 of this study

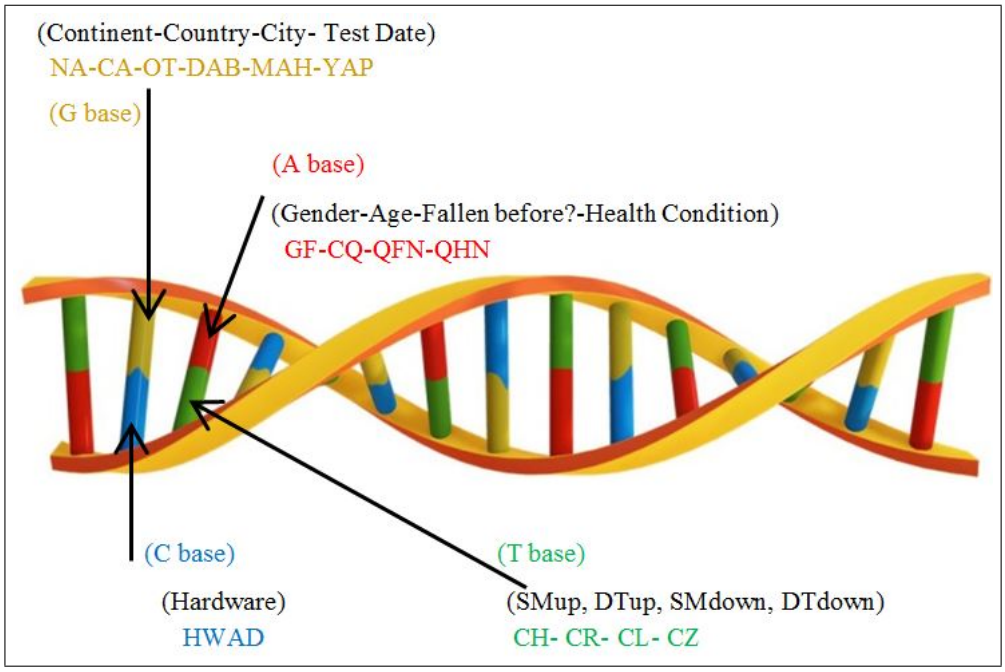


Participant 9

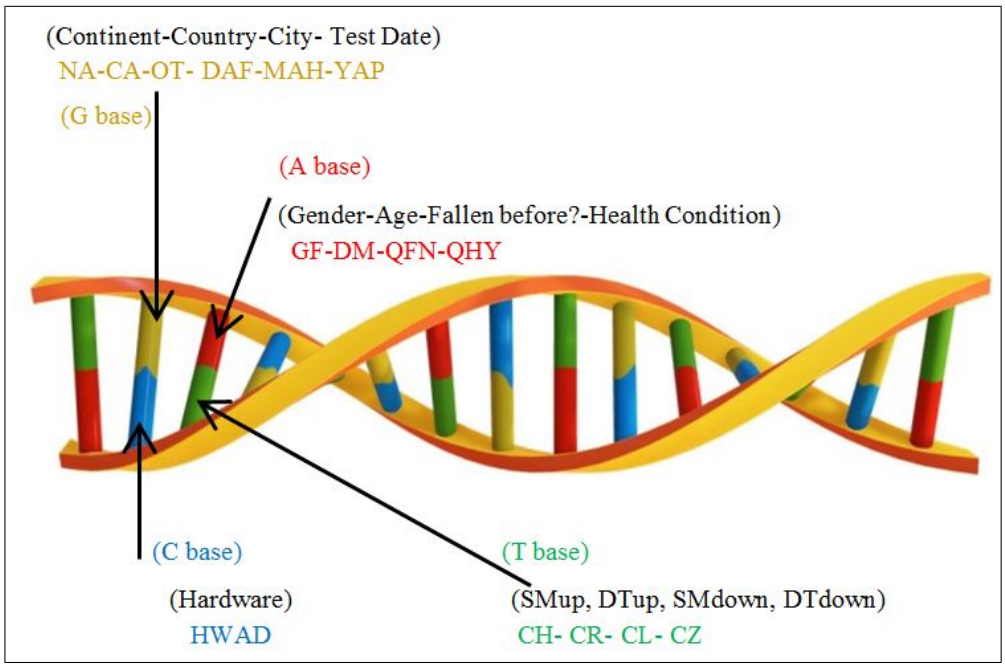


Participant 10

Figure C.5: DT-DNA visualization for Participant 9 and Participant 10 of this study

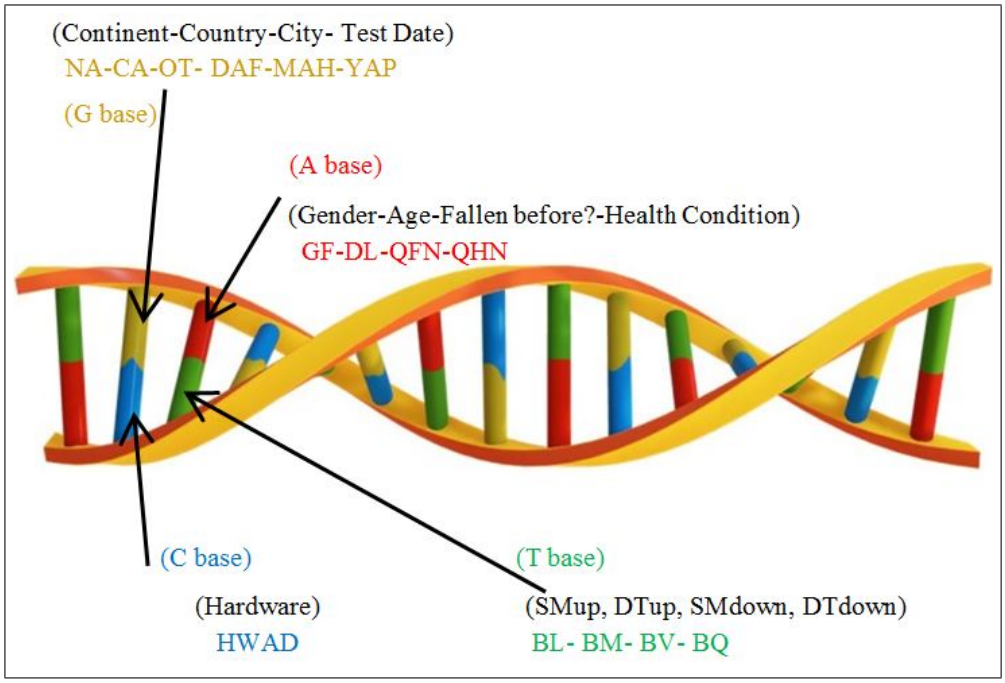


Participant 11

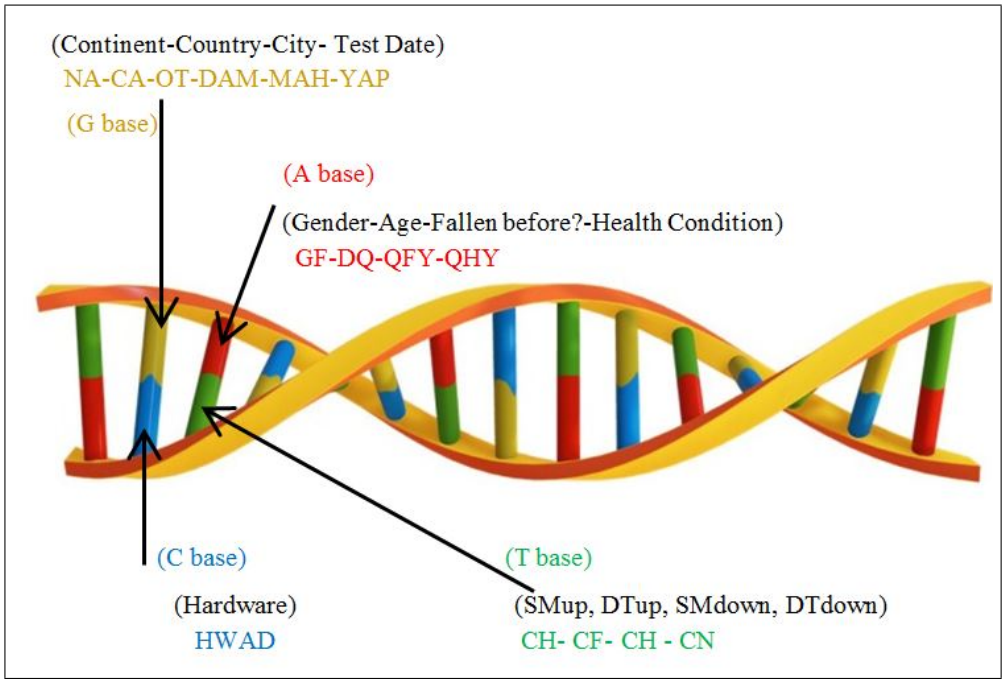


Participant 12

Figure C.6: DT-DNA visualization for Participant 11 and Participant 12 of this study

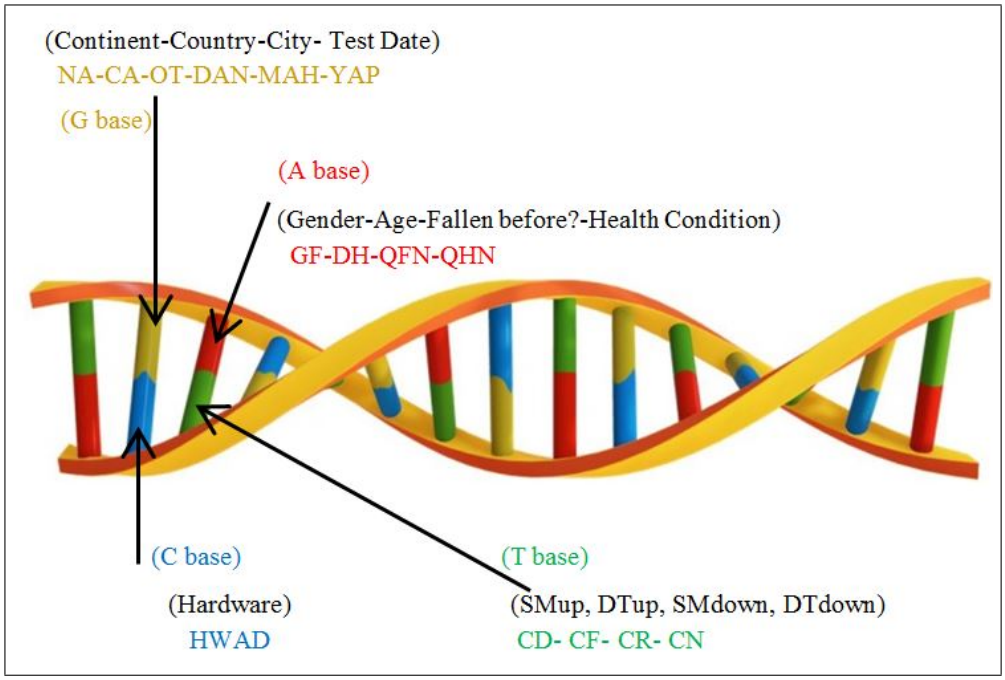


Participant 13

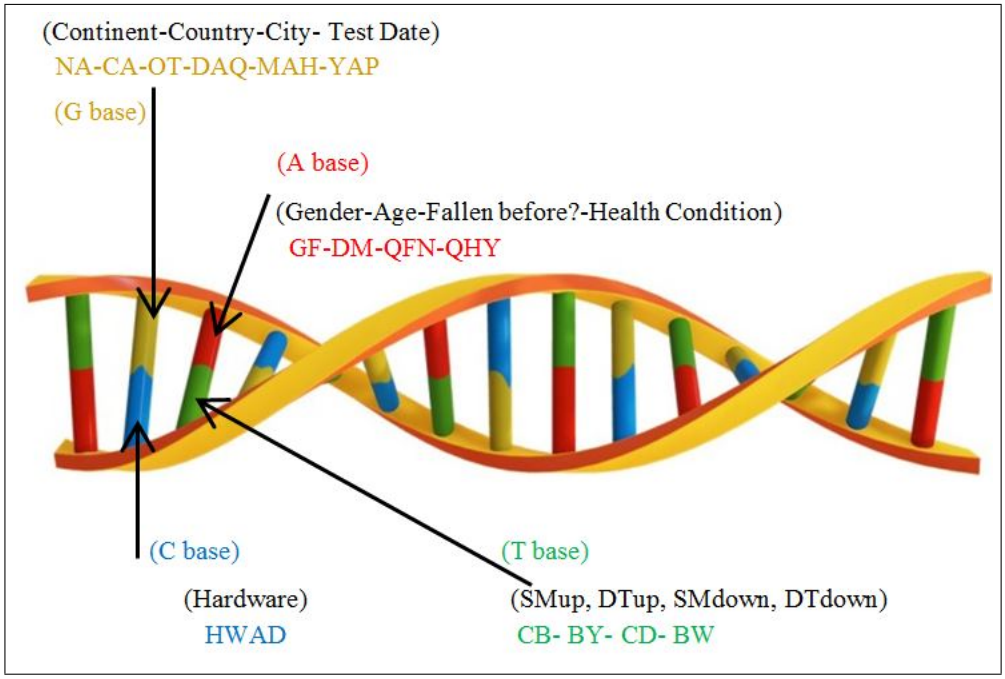


Participant 14

Figure C.7: DT-DNA visualization for Participant 13 and Participant 14 of this study

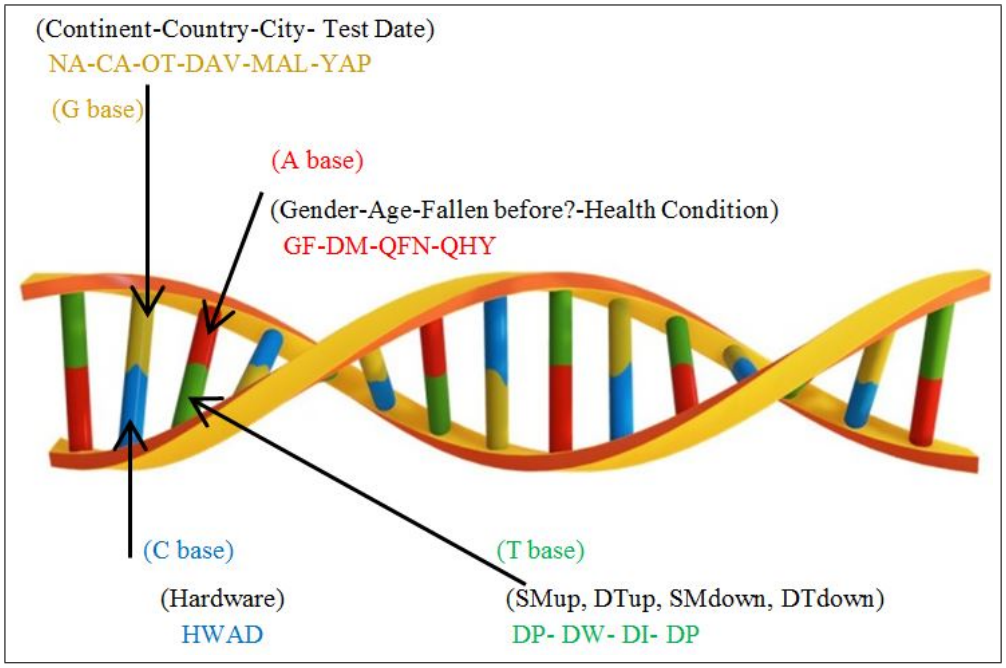


Participant 15

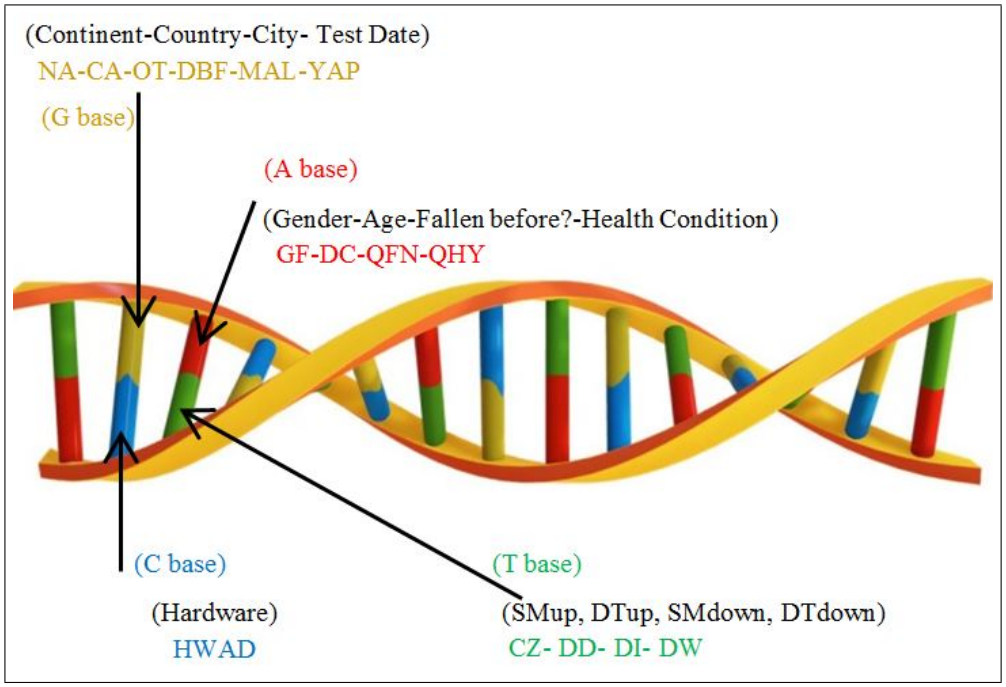


Participant 16

Figure C.8: DT-DNA visualization for Participant 15 and Participant 16 of this study

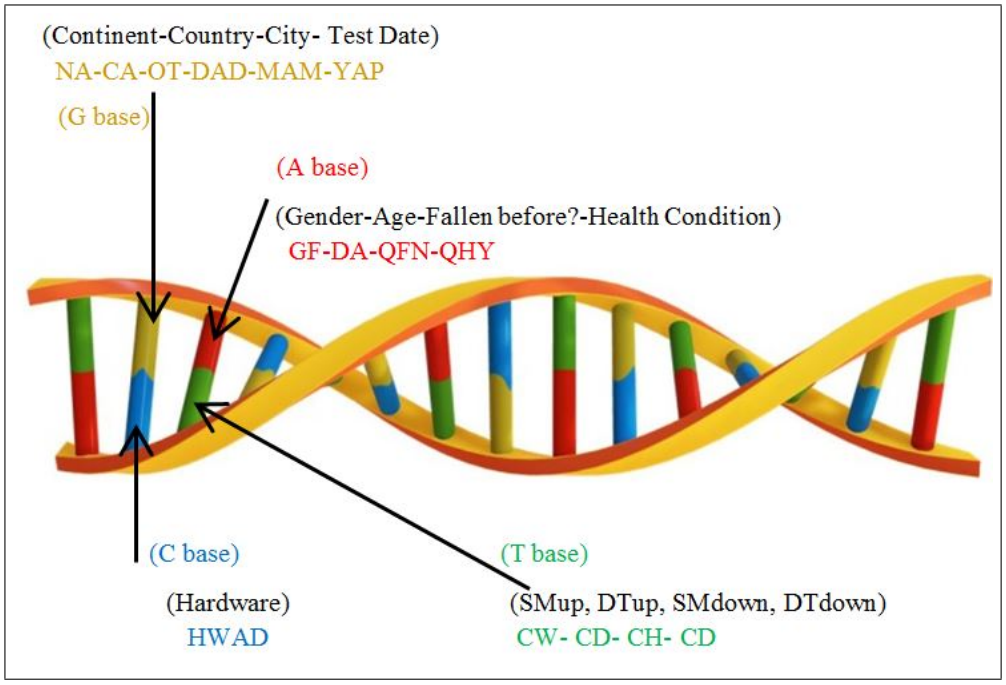


Participant 17

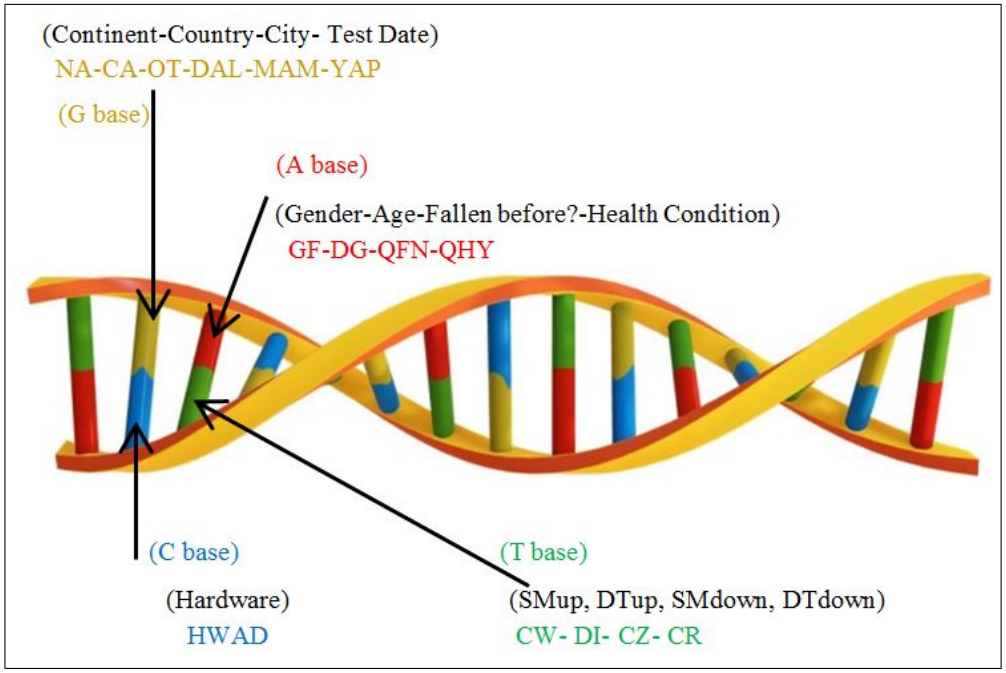


Participant 18

Figure C.9: DT-DNA visualization for Participant 17 and Participant 18 of this study



Participant 19



Participant 20

Figure C.10: DT-DNA visualization for Participant 19 and Participant 20 of this study