Journal of Healthcare Engineering

eHealth Solutions for the Integrated Healthcare

Lead Guest Editor: Giedrius Vanagas
Guest Editors: Rolf Engelbrecht, Robertas Damaševičius, Reima Suomi, and Agusti Solanas

eHealth Solutions for the Integrated Healthcare

Journal of Healthcare Engineering

eHealth Solutions for the Integrated Healthcare

Lead Guest Editor: Giedrius Vanagas
Guest Editors: Rolf Engelbrecht, Robertas Damaševičius, Reima Suomi, and Agusti Solanas

Editorial Board

Contents

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 3846892, 3 pages
https://doi.org/10.1155/2018/3846892

Editorial

eHealth Solutions for the Integrated Healthcare

Giedrius Vanagas,[1] **Rolf Engelbrecht**,[2] **Robertas Damaševičius**,[3] **Reima Suomi**,[4] **and Agusti Solanas**[5]

[1]*Lithuanian University of Health Sciences, Kaunas, Lithuania*
[2]*European Federation for Medical Informatics-Health Information Management Europe ProRec, Berlin, Germany*
[3]*Department of Software Engineering, Kaunas University of Technology, Kaunas, Lithuania*
[4]*Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Turku, Finland*
[5]*Department of Computer Engineering and Mathematics, Smart Health Research Group, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain*

Correspondence should be addressed to Giedrius Vanagas; g.vanagas@gmail.com

Received 21 June 2018; Accepted 21 June 2018; Published 10 July 2018

Information and communication technologies (ICT) bring a whole new dimension into the healthcare arena by introducing electronic media that open the door to the use of new methodologies. Today, ICT support electronic gathering, storage, processing, and exchange of information to treat disease, prevent illness, promote healthy lifestyle, manage patients with chronic illness, and many other applications [1, 2]. Electronic health (eHealth) has the ability to bridge gaps between patients and doctors, patients and relatives, doctors and administrative staff, and so on. Also, eHealth helps to overcome barriers by means that are significantly different from traditional healthcare solutions [3]. It has paved the way for the adoption of sophisticated forms of healthcare provision based on mobile devices and context-awareness, such as mobile health (mHealth) [4] and smart health (sHealth) [5], which require unprecedentedly complex integration strategies.

The advances in healthcare and ICT serve the society as a whole (e.g., patients, healthcare professionals, relatives, and governments) and support the sustainability of healthcare systems. New health ICT systems offer opportunities to improve the effectiveness and efficiency of healthcare services through innovative approaches in clinical service delivery, personalized health, and public health, with a wide impact on the well-being of individuals. Clearly, the study of the integration of healthcare solutions based on ICT is a hot topic that deserves attention.

Despite the well-known potential benefits of adopting eHealth solutions, approaches to ICT implementation used in other industries have had limited success in the healthcare sector. This is not surprising, especially if we take into account the complexity and nature of the healthcare ecosystem. Hence, the implementation of ICT in practice remains difficult and implies many challenges at different levels, involving patients, healthcare providers, and healthcare organizations [6]. Some of the most significant challenges for the adoption of eHealth (and its mHealth and sHealth derivations) are inherent to the complexity of the healthcare system. Take, for example, the *interoperability* problem, which is an old problem inherited from previous non-electronic healthcare infrastructures that is also common in a variety of other domains [7]. Also, ethical issues such as *privacy protection* are a common problem that affects healthcare in general, but it is magnified by the use of ICT [8, 9]. Last but not least, the *integration of new technologies* based on artificial intelligence, smartphones, neural networks, and big data analysis (to name a few) into the healthcare sector poses magnificent challenges for the current healthcare system's limitations, which the scientific community is struggling to overcome. In this sense, the feasibility of many applications, policies, and data concerning the costs, effects, and effectiveness of eHealth and telemedicine are at stake. There is a lack of research-based, empirically-tested models to progress towards large-scale use of ICT in the health sector.

In this special issue, we have undertaken the task of collecting a set of articles that address some of the aforementioned challenges faced by current eHealth infrastructure and applications.

Monitoring long-distance fast walking in daily activities is challenging due to the lack of specific scientific equipment. In the literature, most gait experiments are performed by walking on treadmills. Unfortunately, gait data acquired on treadmills are quite different from the real ones gotten on the ground. The article by W.-F. Wang et al. "*Study on Tripping Risks in Fast Walking through Cadence-Controlled Gait Analysis*" approaches this problem by performing all walking tests intended to reveal important clues for tripping risks in fast walking in a flat and straight pathway. The results show that fast walking with bigger strides and lower cadence is the best way to keep safety on ordinary ground.

Similarly, the monitoring of vital signs is a very active research area. In the article "*Analysis of a Pulse Rate Variability Measurement Using a Smartphone Camera*," A. Bánhalmi et al. approach the problem of using off-the-shelf technology (e.g., smartphones) to gather biometric data efficiently and reliably. More specifically, the authors analyse the possibility of using a smartphone camera to measure the pulse rate variability of patients. Their experiments show that photoplethysmography has high accuracy and does not differ more from ECG than ECG channels themselves. This study opens the door to simplified remote monitoring of the heart function.

In a related topic, haemovigilance is attracting attention. It is the set of surveillance procedures covering the entire blood transfusion chain, from the donation and processing of blood and its components, through to their provision and transfusion to patients, and including their follow-up [10]. In the article "*The Evolving Role of Information Technology in Haemovigilance Systems*," A. Ramoa et al. study 23 haemovigilance organizations in their use of information systems. They find an increasing number of these organizations choosing web-based solutions to take care of haemovigilance. There are still some nonelectronic notification systems, but they lack in data completeness and consistency. The authors support the development of electronic haemovigilance systems and conclude that national haemovigilance systems could benefit from international guidelines for their implementation and maintenance.

There are a number of approaches to develop functionalities for medical decision support systems, which involve some extra efforts from users, thus limiting the spread of such systems in practice. V. L. Malykh and S. V. Rudetskiy provide a review of general approaches to decision support systems development based on nonreduced big clinical data in their article "*Approaches to Medical Decision-Making Based on Big Clinical Data*." The article discusses different approaches to building a medical decision support system based on big data. The authors sought to abstain from any data reduction and apply universal teaching and big data processing methods independent of disease classification standards. The paper assesses and compares the accuracy of recommendations among three options: case-based reasoning, simple single-layer neural network, and probabilistic neural network. Further, the paper substantiates the assumption regarding the most efficient approach to solving the specified problem.

In a similar research line, knowledge-based systems can notably improve the management of distributed eHealth systems, where communication and understanding between medical professionals and different tasks become essential. Usually, medical texts are full of references to medical entities, which could be utilized by knowledge-based medical decision support systems. However, the diverse and ambiguous nature of linguistic ME forms are challenging and hinder automatic medical entities recognition and linking, hence requiring tedious work to annotate data and define features. Xu et al. propose an unsupervised framework for recognizing and linking medical entities mentioned in Chinese online medical texts. Their solution is the first complete unsupervised solution for Chinese medical text with both medical entity recognition and linking, which has considerable value in many applications such as medical knowledge-based construction and expansion, semantic comprehension of medical text, and medical Q&A systems.

Interacting with healthcare systems and applications is an emerging topic that deserves special attention when dealing with disabled people. The article "*Projection mapping user interface for disabled people*" by J. Gelšvartas et al. seeks to help improve user interfaces for people with motoric or speech disabilities through projection mapping. This technique makes possible to create a natural augmented reality information presentation environment. The authors provide a detailed description of a camera-projector system calibration procedure. The described system performs tabletop object detection and automatic projection mapping. The proposed system was tested with real users and, overall, the interface was evaluated positively by the system users, which in most cases were able to learn how to use the system very quickly. The article clearly opens ways for new ideas to produce support systems for motoric or speech disabled people.

Last but not least, the constant aging of the population raises a new set of problems for eHealth systems. In coming years, age-related and degenerative diseases will become the main burden for public health systems. Lauraitis et al. address the problem of identifying Huntington's disease (HD) at its early stage, so that elder patients could benefit from future medical interventions that may help delaying the progress of the disease. They created a computerized behavioural model, which allows predicting an impaired reaction condition for HD patients. The model is embodied in a mobile application available on smartphones and tablet PCs, which allows predicting the functional capacity level of subjects by performing an on-screen touch-based test, thus providing a low-cost alternative to the currently used HD symptom assessment procedures.

The integration of eHealth solutions is far from simple, and it must capture the attention of the healthcare research community for the next years. The adoption of new mobile and context-aware technologies will multiply the benefits for patients and doctors but, at the same time, it will increase the complexity of the eHealth systems and it will be associated to major challenges. In this special issue, we have gathered some relevant examples of research studies that have put their efforts towards this direction. Notwithstanding, there is a lot of unused potential in information systems to support eHealth and its mHealth and sHealth derivations.

It has become clear that technology is not the limiting factor for future progress in healthcare. On the contrary, the main limitations come from the lack of innovativeness of ICT use and the lack of incentives from users and communities to adopt new technology-based solutions for the integrated healthcare. We hope that these limitations will be solved in the next years and that the articles collected in this special issue will contribute with their grain of sand to the overall improvement of healthcare systems worldwide.

Giedrius Vanagas
Rolf Engelbrecht
Robertas Damaševičius
Reima Suomi
Agusti Solanas

References

[1] R. L. Bashshur, G. W. Shannon, E. A. Krupinski et al., "National telemedicine initiatives: essential to healthcare reform," *Telemedicine and e-Health*, vol. 15, no. 6, pp. 600–610, 2009.

[2] M.-P. Gagnon, M. Desmartis, M. Labrecque et al., "Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals," *Journal of Medical Systems*, vol. 36, no. 1, pp. 241–277, 2012.

[3] A. Carbo, M. Gupta, L. Tamariz et al., "Mobile technologies for managing heart failure: a systematic review and meta-analysis," *Telemedicine and e-Health*, 2018, In press.

[4] A. Solanas, A. Martinez-Balleste, P. A. Perez-Martinez, A. F. de la Pena, and J. Ramos, "m-carer: privacy-aware monitoring for people with mild cognitive impairment and dementia," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 19–27, 2013.

[5] A. Solanas, C. Patsakis, M. Conti et al., "Smart health: a context-aware health paradigm within smart cities," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 74–81, 2014.

[6] F. Mair, C. May, C. O'Donnell, T. Finch, F. Sullivan, and E. Murray, "Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review," *Bulletin of the World Health Organization*, vol. 90, no. 5, pp. 357–364, 2012.

[7] C. Patsakis, P. Laird, M. Clear, M. Bouroche, and A. Solanas, "Interoperable privacy-aware e-participation within smart cities," *Computer*, vol. 48, no. 1, pp. 52–58, 2015.

[8] A. Martinez-Balleste, P. A. Pérez-Martínez, and A. Solanas, "The pursuit of citizens' privacy: a privacy-aware smart city is possible," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 136–141, 2013.

[9] A. Solanas, A. Martinez-Balleste, and J. M. Mateo-Sanz, "Distributed architecture with double-phase microaggregation for the private sharing of biomedical data in mobile health," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 901–910, 2013.

[10] World Health Organization, "Haemovigilance," WHO, Geneva, Switzerland, 2016, http://www.who.int/bloodsafety/haemovigilance/en/.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 3917659, 10 pages
https://doi.org/10.1155/2018/3917659

Research Article

Approaches to Medical Decision-Making Based on Big Clinical Data

V. L. Malykh and S. V. Rudetskiy

Medical Informatics Research Center, Ailamazyan Program Systems Institute of RAS, Pereslavl-Zalessky, Russia

Correspondence should be addressed to V. L. Malykh; mvl@interin.ru

Received 21 September 2017; Revised 14 February 2018; Accepted 30 April 2018; Published 4 June 2018

Academic Editor: Giedrius Vanagas

The paper discusses different approaches to building a medical decision support system based on big data. The authors sought to abstain from any data reduction and apply universal teaching and big data processing methods independent of disease classification standards. The paper assesses and compares the accuracy of recommendations among three options: case-based reasoning, simple single-layer neural network, and probabilistic neural network. Further, the paper substantiates the assumption regarding the most efficient approach to solving the specified problem.

1. Introduction

Providing support to medical decision-making is one of the most urgent issues in healthcare automation. It has been repeatedly noted in different articles, reports, and forum discussions [1] both in Russia and abroad that MIS introduction requires a considerable extra effort from users—doctors in the first place—to enter primary data into the system. Naturally, doctors expect practical intelligent outcomes from big clinical data accumulated by modern MISs. Handler et al. [2] present the operating paradigm of 5th generation MISs, referred to as "MIS as Mentor." Malykh et al. [3] adds one more qualitative characteristic to the above paradigm—"MIS as automated mentor." "It is advisable to abandon the practice of active user dialogs typical of expert systems, involving requests for data that the system considers missing from the user, and substitute the dialog with an automated nonintrusive algorithm that draws its own logical conclusions and generates recommendations in a completely automated manner based on available data, without involving the user in the process. The user may either accept or ignore the system's prompts and recommendations; however, they will not provoke rejection in users if delivered automatically without requiring a dialog with the system." To provide a brief qualitative description of this increasing subjectivity of MISs, we have proposed a new term "active MIS" that emphasizes a certain degree of independence from users or subjectivity of the cyber system. Kohane [4] presents the most "balanced" definition of personalized medicine, "personalized medicine is the practice of clinical decision-making such that the decisions made maximize the outcomes that the patient most cares about and minimize those that the patient fears the most, on the basis of as much knowledge about the individual's state as is available." This perception of personal medicine is focused on clinical decision-making and once again exhibits the urgency and importance of scientific research in the area. Therefore, building an automated active mentor-type system that provides recommendations regarding treatment and diagnostic activities to the doctor is an urgent practical task.

Butko and Olshansky [5] and Kotov [6] provide a retrospective overview of approaches to building medical decision support systems. The applied approaches were restricted in many respects by the abilities of computers at that time. Accordingly, there was no such problem as processing big medical data. Technologies have evolved to the point when big medical data (both on individuals and the population in general) collection and accumulation is finally feasible. At the same time, big data processing and intelligent system learning methods were evolving as well. Along with "deep learning," the

term "deep patient" [7] was coined, meaning the opportunity to extract increasingly more complete, deep, and valuable information about patients from big clinical data using deep learning methods.

Malykh et al. [8] mention the possibility of creating national-scale clinical data banks. Herrett et al. [9] provide an example of a database (DB) containing anonymous medical records on primary healthcare services provided. This DB was created by a joint effort of 674 general practitioners and covers over 11.3 mn patients in Great Britain.

Decision-making in hospitals has evolved from being opinion-based to being based on sound scientific evidence. This decision-making is recognized as evidence-based practice. Perpetual publication of new evidence combined with the demands of everyday practice makes it difficult for health professionals to keep up to date [10].

A large number of publications are devoted to medical decision support systems (DSSs), including publications in specialized scientific journals (*Artificial Intelligence in Medicine, BMC Medical Informatics and Decision Making, International Journal of Medical Informatics, Medical Decision Making*, etc.). The work does not aim to give an overview of different approaches to making of decision support systems, referring readers to the original reviews [11–13]. We can give a few definitions for decision support system from Wikipedia: "Clinical Decision Support systems link health observations with health knowledge to influence health choices by clinicians for improved health care" and "active knowledge systems, which use two or more items of patient data to generate case-specific advice." No one doubts the feasibility of such systems and that they have a positive impact on professional practice, patient outcomes, length of hospital stay, and hospital costs. The main problem is to find effective approaches to building such systems.

A number of contemporary approaches to medical decision support system development are listed by Malykh et al. [14].

The first one of these approaches involves provision of relevant data sources to doctors, helping them make decisions independently. The system does not recommend any final solutions—instead, it suggests data sources to study and find answers to current questions (Evidence-Based Clinical Decision Support at the Point of Care | UpToDate URL: http://www.uptodate.com/home).

The second approach is to use clinical pathways. Clinical pathways represent prescriptive models of the standard healthcare procedures that need to be undertaken for a specific patient population. Instances of the clinical pathways (also known as cases) describe the actual diagnostic-therapeutic cycle of an individual patient [15]. But even in the case of the use of clinical pathways, the process of clinical decision-making has high complexity. While the medical knowledge used in the decision process comes partially from published research contributions and widespread medical guidelines (with various kinds of evidence levels), it is generally accepted that the decision process is profoundly influenced by the expertise and experiences of the involved medical experts [15].

The third approach involves development of a large number of individual narrow-focused decision support systems. This approach helps achieve top quality when solving isolated problems [6, 12]; however, it is almost impossible to apply it to big clinical data.

The fourth approach that claims to have a global scope of application is focused on building a cognitive system capable of self-learning and knowledge digestion directly from nonformalized text sources (IBM Watson http://www.ibm.com/smarterplanet/us/en/ibmwatson/).

None of the reviewed approaches is immaculate. All of them require efforts of experts and regular updates of knowledge bases. Moreover, each of the approaches is in fact tailored to specific purposes.

The latest Russian-language review [12] noted that clinical decision support systems have not become widespread in Russia. This is due to the complexity of the development of such systems, the specific character of the systems already developed, and the need to involve high-class experts in the development.

In this paper, we will review general approaches to decision support system development based on nonreduced big clinical data. The main expectations related to application of general approaches ensue from the case-based nature of decision-making in healthcare, and the assumption that big clinical data already contain enough knowledge for efficient decision-making.

There are two other factors that draw attention to systems based on machine learning or precedent approach.

First of them is that there are trends in the development of our civilization which include an explosive development of information technologies (among them M2M, Big Data, and IoT), their strong need for formalized knowledge, and practical absence of qualified experts who could formalize that knowledge. The chief editor of the Rational Enterprise Management (REM) magazine (Russia) holds regular discussions on a wide range of problems including the above-mentioned ones. Results of the discussions are published in the REM editor's column. The guests of a recent discussion [16] included Igor Rudym (Intel), Dmitriy Tameev (PTC), Alexander Belotserkovskiy (Microsoft), Igor Girkin (Cisco), and Igor Kulinitchev (IBM). All the participants agreed that, nowadays, the key challenge of IT development is not associated with hardware or software, but it needs breakthrough approaches to data analysis.

As for the second factor, it is obvious that, nowadays, there are no qualified experts in the field of knowledge even in key branches. The actual situation is even more critical as the experts who are able to solve at least a part of these problems are not able to cope with ever increasing information flow. From this point of view, precedent-based DSSs practically need no experts. Experts may be needed for enhancing or optimizing existing medical data bases and knowledge bases [14].

2. Model and Methods

We regard the diagnostic and treatment process (DTP) as a discrete controlled process with a memory. The model was first introduced by Malykh et al. [17, 18] written in Russian. In English, the model is described by Malykh et al. [8, 19]. To

ensure further understanding of the essence of the problem, let us provide an extract from the source.

Modern medical information systems store electronic medical records and contain descriptions of millions of various clinical cases. The degree of formalization of clinical data stored in MISs varies. MISs model the diagnostic and treatment process as a sequence of controlling events reflecting diagnostic and treatment activities, and a sequence of monitoring events describing the condition of the patient. Controlling events are well formalized; medical organizations keep statistical and business records of such events, plan them, and allocate required resources. Medical data related to monitoring of patients' condition are less formalized and may be partly available in the form of plain text medical documents.

Previous studies provide evidence that is possible to model the DTP using controlled stochastic Markov processes [18]. The model is based on the assumption that the DTP is a discrete controlled process. The model introduces the notions of control U and state X. Controls are diagnostic and treatment decisions made and executed in future. Controls are different diagnostic and treatment activities prescribed by doctors, including diagnostic tests, medicines, surgical interventions, various procedures, and manipulations. The choice of diagnostic and treatment activities is based on the accumulated medical knowledge and the doctor's individual experience. The scope of potential diagnostic and treatment activities comprises previously applied measures with proven efficiency. Controls are essentially precedent dependent.

The choice of control (X_i, U_i) depends not only on the current state (X_i) but also on the overall background of the process as well as controls applied at earlier DTP stages $\{i, i-1, i-2, \ldots\}$. This is due to the specific features and nature of the treatment process. To take the process memory effect into account, it is proposed to include the integral property of the relevant control in the extended state of the discrete process. Each control in the DTP can be associated with some integral property of such control. For example, such integral properties include full dose of medicine taken by the patient at this stage of the DTP or full dose of radiation the patient was exposed to in the course of radiotherapy. The frequency (number) of application of different control elements is also regarded as an integral property (e.g., the number of assigned ECGs).

DTP modeling based on the Markov process appears sufficiently substantiated [17, 18, 20], especially in cases involving DTP description for inpatients with strictly regular monitoring and medical decision-making.

Thus, in the model, the DTP is represented by a sequence of vectors of equal length and structure V split into two components—control U and monitored properties X. Control components have non-negative numerical values. A zero value of control at this stage of the process means that this kind of control has never been applied before, starting from the beginning of the process and up until this step inclusively. Components of monitored properties are of different nature. They can be dimensional physical values or non-numerical, for example, assignment of a property's value to a specific class. Since it is almost impossible to monitor all the properties at the same time, certain components of properties may be unknown to us. When applying different methods to the model, we may need to digitize non-numerical values of components and identify missing values of monitored properties.

2.1. Definition of the Objective. We will review several methods that can be applied to build a cybernetic taught system. The input into the system will be a sequence of vectors describing a discrete DTP in accordance with the presented model. The output will consist of recommendations proposing diagnostic and treatment options (choice of controls) for this particular state of the process. A diagram of the system is presented in Figure 1.

Let us define the objective more accurately and assume that each DTP model is considered in the context of an already available predominant diagnosis. For each model, we have an array of earlier observed DTP implementations. Such implementations are sources of knowledge about treatment of a particular nosology, and they are used to teach a cybernetic recommender system to operate in the given context. Based on available DTP implementations, we defined a glossary of controls and monitored properties for each model. Issues related to normalization of primary data, outlier testing and exclusion, and approaches to data generalization based on assignment of monitored properties to generic classes are beyond the scope of this paper [18]. It might also be necessary to extract data directly from the text of medical documents. Once this enormous and useful effort is completed, we will have a bank of clinical data containing sets of DTPs with homogeneous descriptions for each nosology present in the bank. We would like to emphasize that no primary data reduction is envisaged, such as focusing solely on properties meaningful in the context of the relevant nosology. Data are extracted from the MIS "as is"—exactly as there were entered in the MIS by doctors, assuming such data will most likely contain significant and meaningful information for the relevant nosology.

Finally, let us provide examples of typical properties of nonreduced primary data. We believe that a process ensemble in a data bank may reach 10^3–10^6 processes for an individual nosology. The dimension of a vector describing one step of a discrete DTP exceeds 10^3. The dimension of a control (output of the cybernetic system) may also exceed 10^3.

The case-based approach, including its application to medical decision support, has been described in sufficient detail in multiple sources [6, 14, 19]. The main idea of the case-based approach is quite simple—find a clinical case in the DB similar to the one in focus and use it for medical decision support purposes. Additionally, clinical cases used as precedents during the search can be filtered taking into account such factors as reputation of medical organizations that such cases originate from, reputation of doctors who created such cases, or relevance of the cases in view of contemporary medical technologies. To ensure successful application of the case-based approach, it is necessary to have representative DBs of clinical cases.

Malykh et al. [14] present assessment results with respect to the accuracy of diagnostic and treatment activities

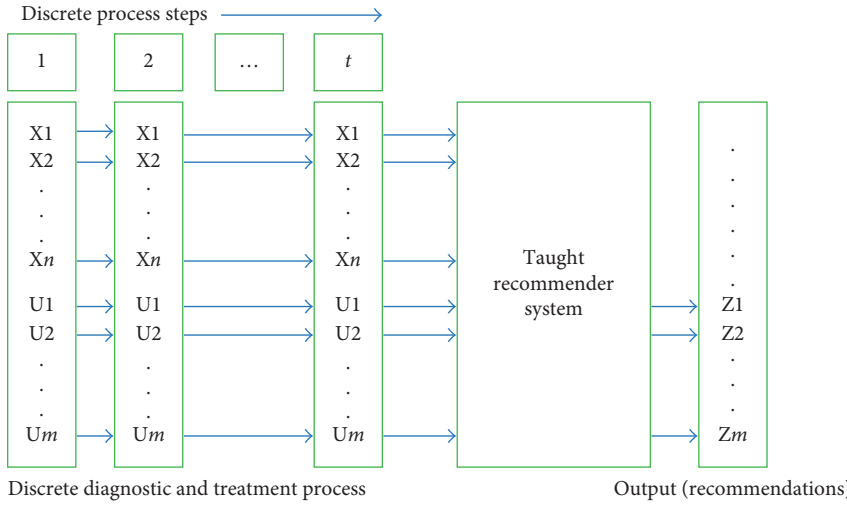


Figure 1: Recommender system.

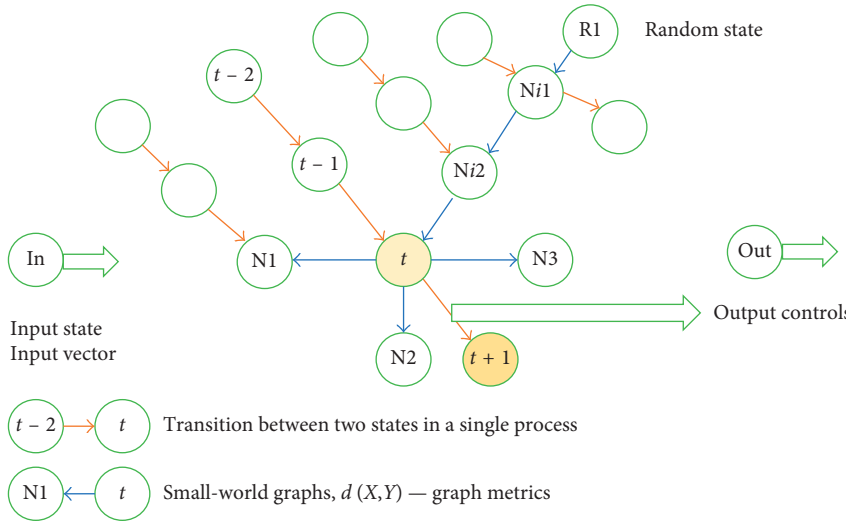


Figure 2: Structure of a case-based system.

recommended using case-based reasoning. The structure of the cybernetic system chosen for the approach in focus is presented in Figure 2.

We have a network and each node in it is presented by a single DTP state. Each individual DTP represents a specific route within the network (routes are marked in Figure 2 by orange arrows). In the model, each state is represented by vector V. A metric or distance $d(X, Y)$ is defined for each state. Based on the defined metric or distance, a small-world graph is plotted [21]. For each node in the small-world graph, n (graph parameter) closest neighbors are identified. In Figure 2, closest neighbors are marked with pointing blue arrows; four closest neighbors are specified for node t—N1, N2, N3, and Ni2.

Here is how the recommender system operates. The input into the system is a current state of the DTP: The situation when the input contains the entire implemented sequence of process states is beyond the scope of this paper. Several nodes are randomly selected on the small-world graph (R1 in the example presented in Figure 2). From original nodes towards their closest neighbors, we go down to the graph node minimizing locally the distance between the node (R1 → Ni1 → Ni2 → t in Figure 2) and the input state. The best of all the identified local minimums is selected. It will be regarded as the closest neighbor of input state In. At this point, the recommended control can be calculated as the difference between integral properties of control components of two vectors. In Figure 2, these are state vectors $(t+1)$ and (t). The recommended control is $U = U(t+1) - U(t)$.

It is easy to assess the scale of the network in focus. In the example with 1,000 processes for one main nosology with the average duration of the process equal to ten days, we will need 10,000 network nodes. Each node will store a vector with the dimension 1,000 or higher. Computational experiments show that 0.5–1% of the total number of nodes is sufficient as random initial network nodes. In case with 10,000 nodes, the number of initial nodes will be 50–100. The

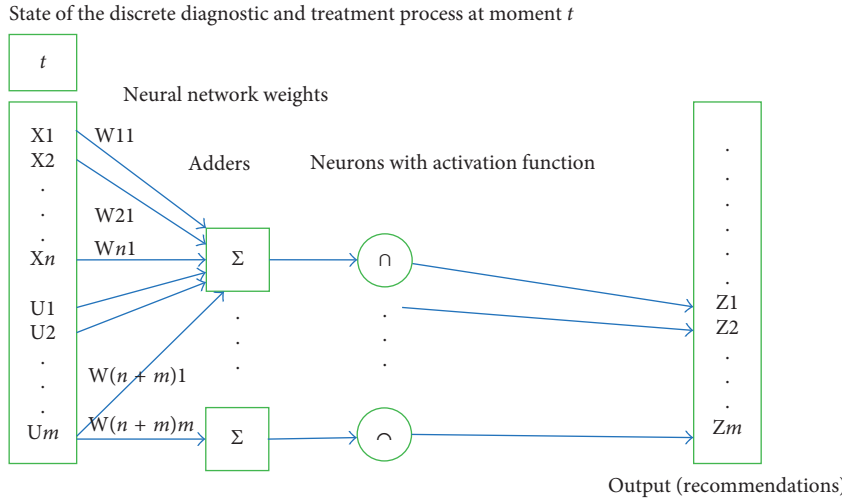


Figure 3: Neural network.

descent along the small-world graph was quick, and the routes did not exceed 10 steps on average. The number of edges originating from each node in the small-world graph was equal to 8. The top-down assessment of the number of metric calculations in this case equals to $100*10*8$. It is possible to accelerate the calculations by splitting the small-world graph into layers corresponding to specific DTP lengths and searching for closest neighbors within the layer corresponding to the input state. In the above example, we would have layers consisting of 1,000 states, and we would search for closest neighbors starting from 5 to 10 randomly selected nodes. This is fully acceptable in view of the computational requirements: computational experiments show that, in this case, computations can be performed almost real-time.

Let us review the network teaching process. Teaching means adding new DTP implementations to the network. The number of metric calculations d when adding k states of a new process to the network containing m states equals to $k*m$. This is absolutely acceptable in view of the computational requirements As a result, new knowledge will be added to the network, and it will be extended by k new nodes and $(k-1+k*n)$ edges. It is essential to emphasize the network's sensitivity to new knowledge. Apparently, any newly added DTP implementation may have a significant impact on the decision recommended by the system if the closest neighbor is selected from the added implementation. It may be asserted that the network digests new knowledge and starts applying it immediately. We will not see this in approaches described below.

As an alternative approach, let us consider a basic neural network with a single layer. The structure of the network is outlined in Figure 3.

Current DTP state is used as input to a basic one-layer neural network. The network contains m adders and m neurons in accordance with the dimension of control component U. In the output, each neuron has either one of the values {0,1}. Output 1 of neuron i means the system recommends control U_i for this state. Output 0 of neuron i means the system refuses to recommend control U_i for this state.

Let us refer to the network scale as an example. Let the dimension of input vector be 1,000 and that of the control component 500. In such case the teaching process will involve definition of $1{,}000*500$ weights. Let us remark that no major reduction of the neural network is possible to solve the above problem. The reason is that the dimension of the control component is the number of diagnostic and treatment activities that can be prescribed for this nosology, including coexisting illnesses. And this number is enormous. Adding new layers to the neural network will only make matters worse by increasing the number of taught parameters.

Let us examine the network teaching process. Initially, a certain set of DTPs is selected and used for network teaching purposes, including calculation of weights. New DTP implementations emerge. How should we use this new knowledge? If a sufficiently large volume of DTP implementations was used to teach the network (1,000 to 10,000) and new implementations constitute an insignificant share of the teaching sample (e.g., 100 new implementations versus 10,000 is merely 1%), it can be asserted that network reteaching will not result in any noticeable changes in teaching parameters, and consequently, any major variations in the network's output. This kind of network is rough and conservative; it can "digest" new knowledge only when the volume of such is sufficient. In this respect, neural networks are not as good as networks applying the case-based approach.

As another alternative approach, let us consider a probabilistic neural network. The structure of the network is outlined in Figure 4. For each state (state vector V), there is one kernel function f(V) common for all the states. In our case, we used a multivariate Gaussian distribution function with a diagonal covariance matrix. The kernel function includes parameter σ affecting the function's width. Each state is classified into $2m$ classes, where m is the dimension of the control component. If a doctor applies control L to state t, then t belongs to class KL1; otherwise, it belongs to class KL0.

Figure 5 shows the impact of control parameter σ on the type of distribution.

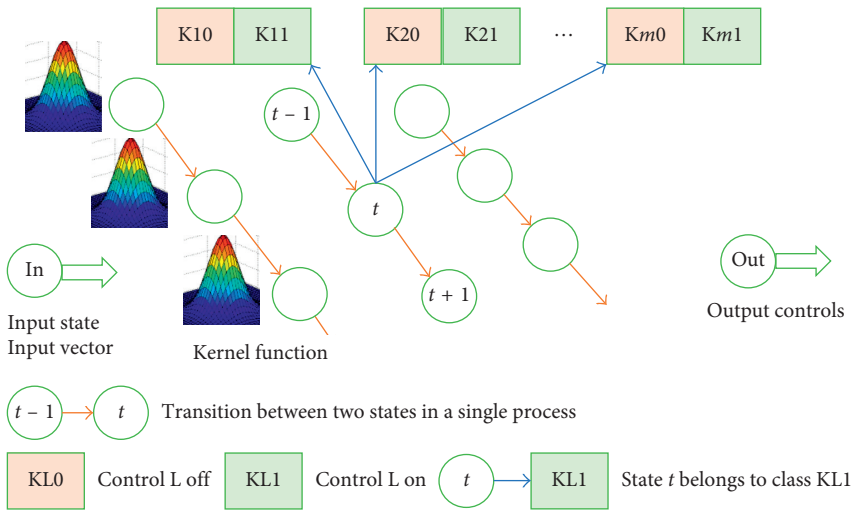


Figure 4: Probabilistic neural network.

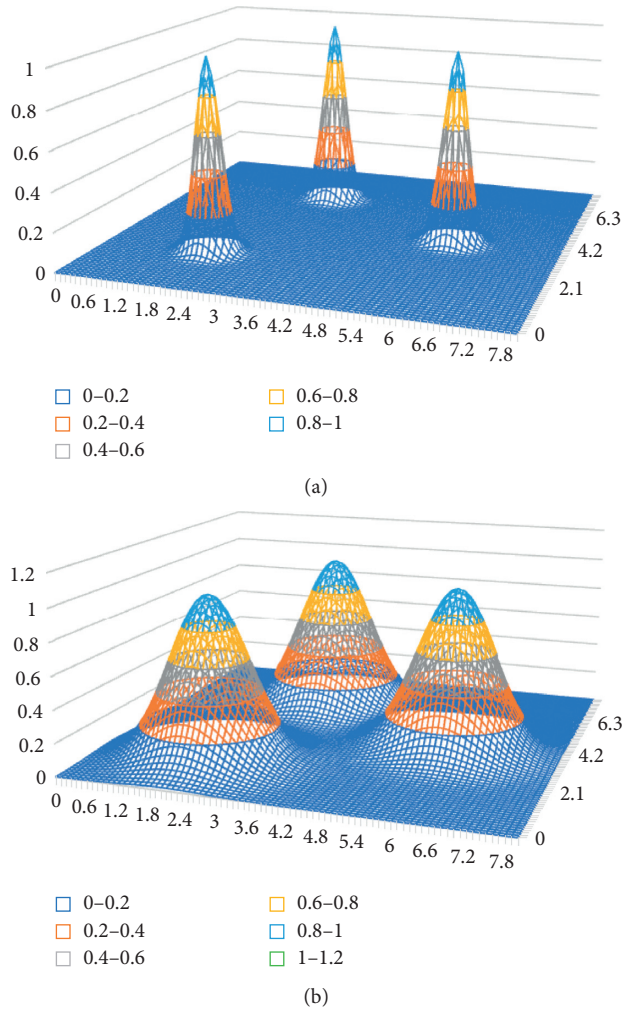


Figure 5: Impact of control parameter σ on kernel functions and type of distribution.

Now, a probability density function can be "restored" for each class. For input vector In, we apply Bayes' formula to calculate the posterior probability of belonging to each class and generate recommendations regarding the choice of diagnostic and treatment activities for this state.

Let us refer to the network scale as an example. Let the dimension of the input vector be 1,000, the dimension of the control component be 500, and the teaching sample contain 1,000 processes with 10 states in each. We will need to calculate 10,000 kernel functions and then calculate 1,000 posterior probabilities of the input vector belonging to each class for various distributions of kernel function supports for 500∗2 different classes.

Let us examine the network teaching process. The teaching process is focused on adding new DTP implementations to the network, including assignment of states to different classes. If the number of new implementations is a small share of the teaching sample used earlier, it can be asserted that adding new implementations will have no major impact on the network's output. The probabilistic neural network proves to be rough and conservative; it can "digest" new knowledge only when the volume of such is sufficient. In this respect, probabilistic neural networks are not as good as networks applying the case-based approach.

3. Results

We performed computational experiments for a network built using the case-based approach in 2015-2016. The results were published in Malykh et al. [14]. To compare different approaches to the problem, we will present the results of paper [14] in a slightly modified format.

Table 1 shows that the number of correct recommendations (TP True Positive) varies from 58.7 to 94.9% depending on the type of nosology. The majority of recommendations match the doctor's actions.

In the matter of neural networks, computational experiments for all nosologies listed in Table 1 required quite a lot of time and computing power. The practical value of such full-scale experiments was unclear. Therefore, it was decided to limit computational experiments to estimations for nosology J13. Table 2 contains general information about the experiment with a single-layer neural network.

Table 1: Accuracy assessment of recommended diagnostic and treatment activities for seven nosologies using the case-based approach.

MKB-10 code/nosology	*Total number of clinical precedents/number of control precedents*	Number of correct recommendations among control precedents	Number of recommendations with a different control level among control precedents	Number of diagnostic and treatment activities the decision support system was unable to provide recommendations for among control precedents
	Number of states/number of controlled variables	Absolute value/share in the total number of diagnostic and treatment activities	Absolute value/share in the total number of diagnostic and treatment activities	Absolute value/share in the total number of diagnostic and treatment activities
J13/pneumonia due to *Streptococcus pneumoniae*	*166/11* 2938/118	6788/81.6%	3923/47.2%	1530/18.4%
K80.1/calculus of gallbladder with other cholecystitis	*1018/128* 12853/931	34468/76.7%	18390/40.9%	10490/23.3%
H25.1/age-related nuclear cataract	*1205/121* 5509/293%	3522/94.9%	539/14.5%	189/5.1%
H26.2/complicated cataract	*1255/126* 5778/249%	4362/91.4%	1617/33.9%	408/8.6%
I67.4/hypertensive encephalopathy	*1336/134* 23165/1431	65678/72.4%	37563/41.4%	25060/27.6%
I67.9/cerebrovascular disease, unspecified	*1403/141* 24875/1518	58649/75.4%	32447/41.7%	19117/24.6%
N20.1/calculus of ureter	*1632/164* 15922/205	17489/58.7%	9948/58.7%	12291/41.3%

Table 2: Accuracy assessment of recommended diagnostic and treatment activities for nosology J13 based on a single-layer neural network.

MKB-10 code/nosology	*Total number of clinical precedents/number of control precedents*	Number of correct positive recommendations among control precedents	Number of incorrect positive recommendations among control precedents	*Total number of negative recommendations/total number of positive recommendations* Share of correct negative recommendations/share of correct positive recommendations
	Number of neural network inputs/number of neural network outputs (number of controlled variables)	Absolute value/share in the total number of positive recommendations	Absolute value/share in the total number of positive recommendations	Absolute value/percent
J13/pneumonia due to *Streptococcus pneumoniae*	*266/11* 224/222	339/40.31%	502/59.69%	*35567/841* 98.55%/40.31%

Let us emphasize that the volume of statistics on this illness stored in the DB has increased compared to an earlier experiment involving the same nosology—from 166 to 266 completed clinical processes. Controls included all types of drug prescriptions (222 different pharmaceutical products in our case). Data normalization involved adjustment of prescribed dosages of pharmaceutical products to unified dose units. The only monitored variable was "inpatient days." Inputs also included bias. 49,728 weights had to be determined. The optimized target function was a quadratic residual between neural network output and control components monitored in control samples, adjusted to (0, 1). We used a nonstandard neurons activation bell curve (Gaussian function). This choice of activation function was based on the fact that integral values of many controls had apparent limits stipulated by Russian federal healthcare standards (standards of the Russian Ministry of Health). Different insurance programs also limit integral values of controls. Healthcare providers will not exceed these limits unless they find it necessary. Formally, with respect to the model, it means that once an integral property of a control reaches a certain limit, it stops growing further or such growth is highly unlikely. The gradient of the target function with respect to weights was calculated explicitly, and the steepest descent method was applied. Teaching included 1,006 descent steps. Criteria reflecting the accuracy of the neural network are presented in Table 3.

The relevant receiver operating characteristic (ROC) error curve is shown in Figure 6.

Results of the experiment based on a probabilistic neural network are presented in Table 4. The state vector dimension was equal to 639. The control component included 125

Table 3: Accuracy of recommended diagnostic and treatment activities for nosology J13 based on a single-layer neural network with an activation threshold equal to 0.1.

Absolute values (neuron activation threshold equal to 0.1)			
TP	339	502	FP
TN	35,052	515	FN
Percent (neuron activation threshold equal to 0.1)			
TP	40.31%	59.69%	FP
TN	98.55%	1.45%	FN

TP, true positive; FP, false positive; TN, true negative; FN, false negative.

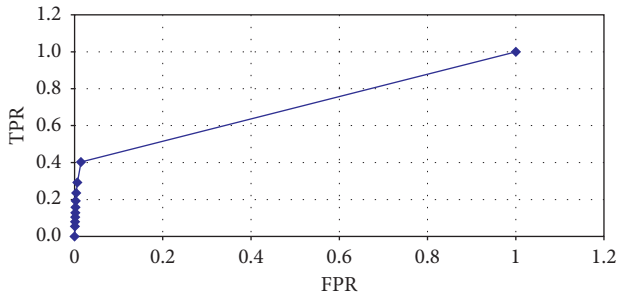


Figure 6: ROC error curve.

diagnostic tests, 200 laboratory tests (different kinds), 222 different pharmaceuticals, 87 medical treatments, and 4 controls classified as "others." The only monitored property was "inpatient days." The number of kernels (states) in the teaching sample of 266 processes was 4,361. The dimension of the state vector in the probabilistic neural network was almost three times the dimension of the state vector in the single-layer neural network (639 versus 223). To make the results of both networks comparable, the output of the probabilistic neural network was considered to be the same as for the first neural network. The output was a vector with a dimension of 222, related to prescription of different pharmaceuticals. Both neural networks generated 36,408 positive and negative recommendations for the control sample. The experiment involved one control parameter σ, a multiplier for a diagonal covariance matrix used in the kernel function (multivariate Gaussian distribution of independent random variables). A value grid was predetermined for the parameter σ, and the best value of the parameter was chosen based on experimental calculation results [22]. Calculations were performed for the following values of σ: (0.1, 0.5, 1, and 2.5). The best results were obtained for $\sigma = 2.5$. They are presented in Table 4. Let us emphasize that standard deviation values of the state vector components calculated for the teaching sample were significant and often exceeded average values. The multiplier equal to 2.5 yields "wide" kernel functions (see the rightmost distribution in Figure 5). With "sharp" kernel functions ($\sigma = 0.1$), the results were obviously worse.

4. Summary

The focus of this paper was how to build a medical decision support system based on big clinical data. The authors review general approaches to the problem that do not involve individual models for specific nosologies and neither do they require engagement of experts in the relevant subject area to

Table 4: Accuracy of recommended diagnostic and treatment activities for nosology J13 based on a probabilistic neural network with $\sigma = 2.5$.

Absolute values ($\sigma = 2.5$)			
TP	233	191	FP
TN	35,376	608	FN
Percent ($\sigma = 2.5$)			
TP	55.0%	45.0	FP
TN	98.31%	1.69%	FN

such modeling or knowledge extraction from data. Data are extracted from the MIS without reduction, "as is." It is assumed that the data contain significant information reflecting medical knowledge and contemporary medical treatment technologies. Three different approaches to big clinical data processing were examined: (1) case-based reasoning for decision-making; (2) decision-making based on a single-layer neural network; and (3) decision-making based on a probabilistic neural network. Experimental calculations were performed to assess the accuracy of recommendations generated using different approaches.

Drawbacks of the above neural networks with respect to the given problem were identified. The overall accuracy of provided recommendations was rather high. Moreover, the accuracy of negative recommendations that the neural networks learned to provide was very high (98–99%). However, the accuracy of positive recommendations provided by the neural networks was not so high (40–55%, which is obviously insufficient for successful practical application). Another disadvantage of neural networks is their rough and conservative nature, particularly when digesting isolated portions of new data with the volume insignificant compared to previously available data.

The case-based approach to decision-making yielded more accurate recommendations (59–95%), which is sufficient for its successful practical application. Another advantage of the case-based approach is its sensitivity to new data. With respect to calculations, the case-based approach is also more efficient compared to other options under consideration as it ensures a high operating speed of the decision support system, thus making it acceptable for practical application. These are the key findings of the study conducted.

This offers encouraging prospects for designing and developing decision support systems for physicians based on empirical components of medical knowledge. This approach also corresponds to existing case-based character of management and decision-making in medical practice. So far, the results indicate that precedent-based approach has a high effectiveness and could naturally enhance other approaches to supporting physicians' decision-making, particularly knowledge-based ones. The obvious practical value of this approach lies in the fact that it can be complementary to other knowledge-based approaches (clinical pathways, Evidence-Based Clinical Decision Support, expert systems, Watson, etc.). The doctor will be able to make decisions based on the best examples of medical practice, finding precedents of clinical cases close to the given case.

The constraints of precedent-based approach include the need for a representative database of verified precedents

excluding medical errors. From another perspective, precedents with corrected errors are of particular interest to physicians training and further prevention of such errors. The information about the results of these errors and possible ways of correcting them is also valuable. Thus, precedent-based approach could be widely spread as an educational tool. On the other hand, the precedent-based approach does not imply formalization of medical knowledge, which entails poor cognitive justification of generated recommendations. Consequently, justifications only describe how other patients were treated in similar clinical cases. There are also problems with optimization of provided metrics, compression of state descriptions, and construction of training procedures. These problems are connected with high dimensionality of the space of state characteristics and samples of clinical precedents. However, discussion of these issues and possible ways of addressing them has been left outside of this research [14].

In further studies, we are going to focus on detailed application of the case-based approach, analyze metrics, and distances not only for pairs of vectors but also for pairs of vector sequences, and examine issues concerned with intelligent normalization of primary data and data extraction from plain texts of medical documents.

Disclosure

UDC 007.52 (Automatically operated systems without any humans among system links, robots, and automated machines).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Professor V. M. Khachumov, a doctor of engineering sciences, and V. P. Fralenko, a candidate of engineering sciences, for the consulting support on neural networks and teaching methods, as well as Professor N. N. Nepevoda, a doctor of physical and mathematical sciences, for the discussion and assessment of the outcomes. Some of the outcomes presented in the paper were achieved earlier under the support of the Ministry of Education and Science of the Russian Federation (Project RFMEFI60714X0089) and in the context of Grant 13-07-12012 provided by the Russian Foundation for Basic Research.

References

[1] MedSoft, "Medical software," in *Conference Presentations of the 12th International Forum MedSoft-2016, in Russian*, Moscow, Russia, 2016, http://www.armit.ru/medsoft/2016/conference/prog.

[2] B. S. Handler, J. Thomas, and B. R. Hieb, *Gartner's 2007 Criteria for the Enterprise CPR*, 2014, http://rsept.wikispaces.com/file/view/Gartner_Criteria_for_the_Enterprise_CPR_2007.pdf.

[3] V. L. Malykh, S. V. Rudetskiy, and M. I. Khatkevich, *Active MIS. Information Technologies for the Physician*, Publisher House "Public Health Manager", Moscow, Russia, no. 6, pp. 16–24, 2016, in Russian.

[4] I. S. Kohane, "The twin questions of personalized medicine: who are you and whom do you most resemble?," *Genome Medicine*, vol. 1, no. 1, p. 4, 2009.

[5] S. N. Butko and V. K. Olshansky, *New Decision Support Systems in Foreign Healthcare. Automation and Remote Control*, Springer, Berlin, Germany, 1990.

[6] Y. B. Kotov, *New Mathematical Approaches to Medical Diagnostics*, Editorial URSS, Moscow, Russia, 2004.

[7] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records," *Scientific Reports*, vol. 6, no. 1, p. 26094, 2016.

[8] V. L. Malykh and D. V. Belyshev, "Case-based reasoning in clinical processes using clinical data banks," in *Proceedings of 2015 International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON)*, Technopark of Novosibirsk Akademgorodok, pp. 211–216, Novosibirsk, Russia, 2015.

[9] E. Herrett, A. M. Gallagher, K. Bhaskaran et al., "Data resource profile: clinical practice research datalink (CPRD)," *International Journal of Epidemiology*, vol. 44, no. 3, pp. 827–836, 2015.

[10] T. Rotter, L. Kinsman, E. James et al., "Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs," *Cochrane Database System Review*, vol. 3, p. CD006632, 2010.

[11] AHRQ, *Clinical Decision Support Systems: State of the Art*-Agency for Healthcare Research and Quality U.S, AHRQ Publication No. 09-0069-EF, Rockville, MD, USA, 2009.

[12] I. V. Efimenko and V. F. Khoroshevsky, "Intelligent decision support systems in medicine: state of the art and beyond in Russian," in *Proceedings of OSTIS*, Open Semantic Technologies for Intelligent Systems, Minsk, Belarus, 2017, http://proc.ostis.net/proc/Proceedings%20OSTIS-2017.pdf.

[13] Wikipedia, *Clinical Decision Support System*, http://en.academic.ru/dic.nsf/enwiki/1126747, 2017.

[14] V. L. Malykh, I. N. Kononenko, and S. V. Rudetskiy, "Estimation of accuracy of recommended diagnostic and treatment actions based on precedent approach," in *Proceedings of the International Conference e-Health 2016*, pp. 52–58, Madeira, Portugal, July 2016.

[15] F. Caron, J. Vanthienen, and B. Baesens, "Healthcare analytics: examine the diagnosis-treatment cycle," *Procedia Technology*, vol. 9, pp. 996–1004, 2013.

[16] E. Vasilyeva, *Industrial Internet of Things (IoT)*, Rational Enterprise Management, Athens, GA, USA, 2015, in Russian.

[17] V. L. Malykh and Y. I. Guliev, "Controlled stochastic precedent process with memory as a mathematical model of the diagnostic and treatment process," *Information Technologies and Computational Systems*, vol. 2, pp. 62–72, 2014, in Russian.

[18] V. L. Malykh, Y. I. Guliev, A. V. Eremin, and S. V. Rudetskyi, "Management and decision making in clinical processes," in *Proceedings of XII All-Russian Conference on Problems of Management of VSPU-2014*, pp. 6518, Moscow, Russia, 2014, in Russian.

[19] V. L. Malykh and Y. I. Guliev, "Precedent approach to decisionaking in clinical processes,", in *MEDINFO 2015: Health-nabled Health, Studies in Health Technology and Informatics,* Vol. 216, Ed., IMIA and IOS Press, Amsterdam, Netherlands, 2015.

[20] C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach," *Artificial Intelligence in Medicine*, vol. 57, no. 1, pp. 9–19, 2013.

[21] Y. Malkov, A. Ponomarenko, A. Logvinov, and V. Krylov, "Approximate nearest neighbor algorithm based on navigable small world graphs," *Information Systems*, vol. 45, pp. 61–68, 2014.

[22] R. N. Kvetniy, V. V. Kabachiy, and O. O. Chumachenko, *Probabilistic Neural Networks in Time Series Identification*, Information Technologies and Computers, Scientific works of Vinnytsia National Technical University, Vinnytsia, Ukraine, 2010, in Russian.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 2723178, 11 pages
https://doi.org/10.1155/2018/2723178

Research Article

Study on Tripping Risks in Fast Walking through Cadence-Controlled Gait Analysis

Wen-Fong Wang,[1] Wei-Chih Lien,[2] Che-Yu Liu,[1] and Ching-Yu Yang[3]

[1]*Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu, Yunlin 640, Taiwan*
[2]*Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 600, Taiwan*
[3]*Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, Magong, Penghu 880, Taiwan*

Correspondence should be addressed to Wei-Chih Lien; lwclwhab@ms8.hinet.net

Received 20 October 2017; Accepted 5 April 2018; Published 24 May 2018

Academic Editor: Giedrius Vanagas

Fast walking is a common exercise for most people to promote health. However, a higher cadence due to fast walking on ordinary or uneven ground raises the risk of tripping. To investigate the tripping issue, research to observe the gait in fast walking is needed. To explore the relationship between fast gait and the risk of tripping, a gait recording system with a specific synchronization mechanism was developed in this work. The system can acquire gait signals from wearable sensors and action cameras at different cadences. Meanwhile, algorithms for gait cycle segmentation and characteristic extraction were proposed for analyzing a fast gait. In the gait analysis, the correlations of low, moderate, and high cadence in cueing and no cueing gaits were computed, and two results were obtained. First, the higher the cadence is, the larger the motion strength in the terminal foot swing will be and the smaller the motion strength at the starting foot swing. Second, the decreased distance of foot clearance becomes more conspicuous as the cadence increased, especially if one is walking more than 120 beats. The results indicate that fast walking with bigger strides and lower cadence is the best way to maintain safety in moving over ordinary ground.

1. Introduction

Walking is an important daily activity, and many people do that to maintain healthy condition. Movements in walking can be affected by the rhythmic sounds, which make the movements smoother and more stable [1, 2]. In walking, many people use the music tempo to regulate their cadence and breathing rate to keep their exercises more lasting. Additionally, they also conduct fast walking to increase their exercise intensity for calorie consumption. In terms of medical treatment, many patients that have Parkinson's disease [3], children with cerebral palsy [4], and many patients that have strokes [5] and gait rehabilitation [6] use metronomes to help their exercises.

In the literature, rhythm on gait has two research topics, which are in health promotion and medical applications. For the health promotion, some studies reveal that the music tempo affects gait movements and benefits human health conditions. In [7], Mendonca et al. studied the music played at the treadmill to regulate the walking pace and analyze the gait parameters. Their experiment included walking with and without music cueing synchronization, and the results showed that when there was music cueing synchronization, people would change their walking step with the music tempo. When there was no music cueing synchronization, people would not change their walking step with the music tempo obviously. In [8], Eikema et al. found that old people's gait parameters, such as the stride and step, reduced dramatically, and their walking speed could be changed using the voice cue to regulate their pace on treadmills. After the training on walking in cue, their gait and strides were improved obviously to be closer to young people.

In medical applications, acoustic tempos are frequently used in gait rehabilitation, with positive instantaneous and prolonged transfer effects on various gait characteristics. In [9], rhythmic auditory cues including music and metronome beats were used to improve disordered gait in clinical conditions. Their study indicated that music and metronome cues may not be used interchangeably, and the cue type and frequency (cadence) should be considered when evaluating effects of rhythmic auditory cues on gait. In [10], the gait pattern in various rhythmic auditory stimulation tempos was investigated in stroke patients. The results showed that as the tempo increased, the spatiotemporal gait parameters of the stroke patients changed significantly. After gait training with rhythmic auditory stimulation, gait parameters of the affected and unaffected sides improved significantly compared to baseline.

In the previous studies, treadmills were often used for step and pace control, and thus, the conducted experiments were often restricted in space [5, 7, 8, 10]. Therefore, the research results were biased due to few data from daily walk. In recent years, the advancement of the wearable technology has made sensors miniaturized and widely applied in many research and applications [11]. As the wearable sensors overcome space restrictions and improve the degree of freedom in use, these devices can be used to replace fixed and expensive measuring apparatuses [12]. In this study, we developed a gait recording system, which equipped with a set of wearable triaxial acceleration sensors and action cameras. The sensors are in miniature size and can be easily worn by subjects to acquire gait signals from daily walk. The action cameras can be used to record gait movements. With this system, gait signals for daily walking can be acquired without space restrictions.

In [13], Rubenstein described that the risks of falls and trips are due to some identifiable factors such as weakness, unsteady gait, confusion, and certain medications and, simultaneously, indicated that attention to the factors can greatly reduce the rates of falling. He also described that one of the most effective fall reduction programs involve environmental inspection and hazard reduction. In [14], Lord et al. indicated that the most challenging area to incorporate into fall aetiology is that of environmental risk factors since the research base is smaller than that for other types of risk factors. At the same time, research in this area has typically been less rigorous or has been troubled by methodological limitations. Since a gait recording system with the wearable technology has been developed in this work to improve the methodological limitations, we are interested in exploring the tripping risks in fast walking through an uneven ground, which may exist with a few environmental risk factors, by a cadence-controlled strategy. With the application of action cameras and a specific synchronization mechanism between the signals from the wearable sensors and the images from the cameras, we realize the gait analysis of fast walking without space restrictions.

This investigation is organized as follows: Section 2 presents research materials, including gait description, gait signal acquisition, cadence control, and a signal synchronization mechanism. Section 3 describes research methods, including the experimental protocol, signal preprocessing, signal segmentation and characteristic extraction, and correlation analysis. Section 4 shows the results, which include correlation analysis in cadence versus gait events and periods with and without cueing. In addition, analyses of gait phase variation and gait periods versus walking types are also done for exploring the tripping risks. Concluding remarks and future work are given in Section 5.

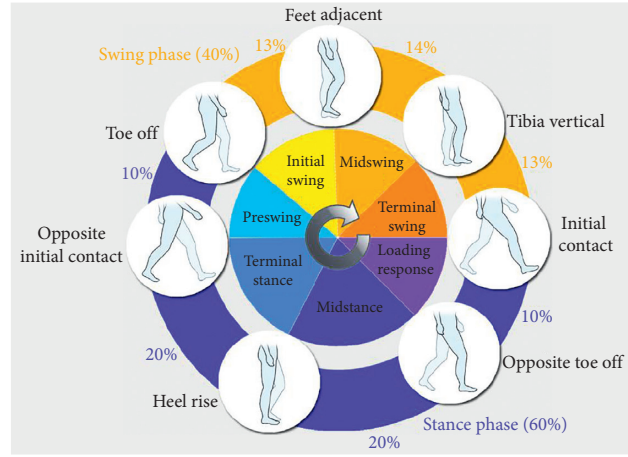


Figure 1: Events, periods, and cycles of human gait [15].

2. Research Materials

2.1. Gait Description. Human walking consists of a number of cyclically sequential gait events, and one complete gait cycle (Figure 1) can be measured from one toe off event to the next one [15]. As shown in Figure 1, a full gait cycle includes a stance phase and a swing phase, where 60% of the cycle is spent in stance and 40% in swing [16].

The gait cycle has seven gait events, that is, initial contact (IC), opposite toe off (OTO), heel rise (HR), opposite initial contact (OIC), toe off (TO), feet adjacent (FA), and tibia vertical (TV) [17]. It can also be subdivided into seven periods, that is, loading response, midstance, terminal stance, preswing, initial swing, midswing, and terminal swing. While finely subdividing gait events and periods for every cycle, different gait parameters such as the cycle duration, stance time, swing time, stride length, step length, cadence, and walking speed can be analyzed to learn about personal gait features.

2.2. System of Signal Acquisition. In this study, a wearable sensor module with a triaxial accelerometer shown in Figure 2 was developed to acquire motion signals related to human gait. The size of this sensor is about $50 \times 30\,\text{mm}^2$, and it includes a microprocessor (TI MSP430™ series [18]), a triaxial accelerometer (ST LIS3DH [19]), a 8 GB micro-SD flash memory card, and a 3.3 V lithium battery. The acceleration measuring range of the sensor is dynamically user selectable full scales of ±2 g/ ±4 g/ ±8 g/ ±16 g, and the sampling data rate is programmable from 1 Hz to 5 kHz. Here, the acceleration measuring range is set to ±16 g and the sampling data rate to 120 Hz. For the acquired signal data, they would be stored in the micro-SD memory card temporarily,

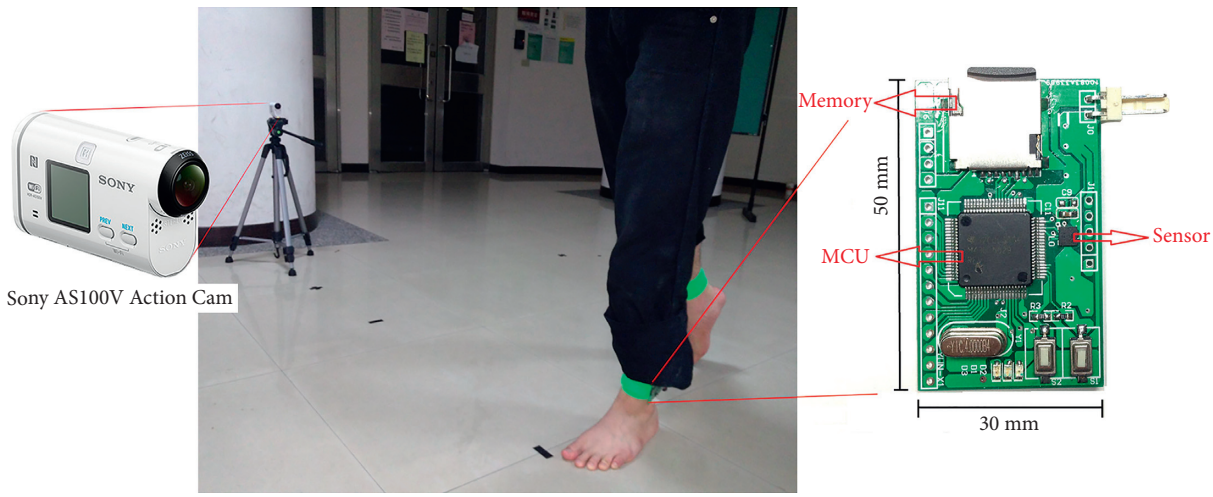


Figure 2: Wearable sensor module with a triaxial accelerometer and a high-speed camera for gait signal acquisition.

and later, the data are uploaded to server computers for signal processing. In the procedure of motion signal acquisition, the sensor would be attached steadily on one foot of a subject to a position 2~3 cm above his/her ankle.

To record the video of gait movements, a camera (Sony AS100V Action Cam [20]) shown in Figure 2 was applied. The highest image capture rate of the camera is 240 FPS (frames per second), and each captured image has the resolution of 1280×720 pixels. Thus, the camera satisfies the requirement of our experiment for recording and watching the actions of gait events. For the sake of recording synchronization between the sensors and cameras, the image capture rate was set to 120 FPS to meet the requirement of synchronizing gait signals and foot locomotion images.

2.3. Control of Cadence. In this study, the music tempo was employed to adjust the cadence of subjects for observing the gait variability in fast walking. Normally, the music tempo is calculated by beats per minute (BPM), and the number of beats for walking means the number of steps to go in a minute. Generally, the cadence of a normal person in walking is about 80~120 BPM, and for fast walking, it is about 120~140 BPM. For the purpose of comparing the gait patterns in slow, normal, and fast walking, we focused on the tempo of 80, 100, and 120 BPM for the controlled cadence.

Usually, music has many different styles such as pop music like hip-hop, rock and roll, and jazz, which have a strong tempo to ease the regulation of the walking rhythm. Here, the jazz's tempo was considered to regulate the walking rhythm [21]. This idea has three considerations. Firstly, it usually possesses the drum sounds of low frequency for beats so that the subjects can identify them clearly while walking. Secondly, the jazz's tempo is almost fixed and good for gait control. Lastly, it can make subjects feel relaxed and walk with ease. Although a metronome seems a little better than the jazz's music to provide a clear tempo, we still preferred the jazz's tempo due to the above considerations.

2.4. Synchronization between Motion Signals and Captured Images from Video. In the aforementioned signal acquisition system, the sampling rate of the wearable sensors was set to 120 Hz, and the video frame rate of the high-speed cameras was also set to 120 FPS. Therefore, a synchronization mechanism is needed to relate the image of one gait event recorded by the camera to one motion signal acquired by the wearable sensor sequentially for identifying the gait characteristics. To complete the synchronization, a leading clearance matching synchronization mechanism was proposed, and we asked all participants to stand still for 10 seconds after the setup of the sensors and cameras and before each walking test. In this way, a long and clear section of the temporal gap can be created for identifying the synchronized origin of the gait images and motion signals.

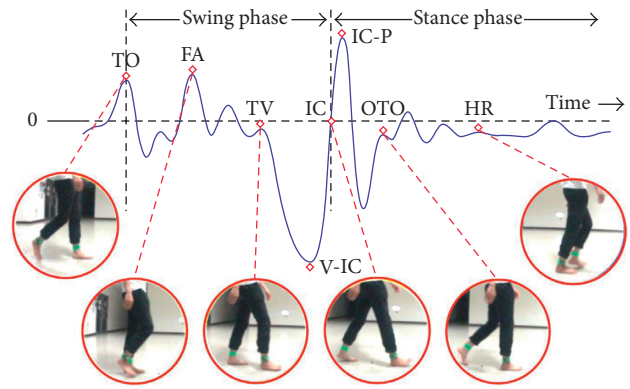


Figure 3: Matching gait characteristics for the gait events of the right foot through image marking in one gait cycle.

Under the supervision of one medical expert in our team, we discovered the gait events continuously in the recorded video through a frame-by-frame video player in slow motion. At the same time, we marked the temporal position of each found image and used that position to extract the corresponding accelerometrical signal of each gait event in the acquired signal set. In this study, we call the corresponding accelerometrical signal of a gait event as a gait characteristic. As shown in Figure 3, a sample curve of gait cycles acquired from the right foot of one participant is depicted with a significant valley (labeled by "V-IC"),

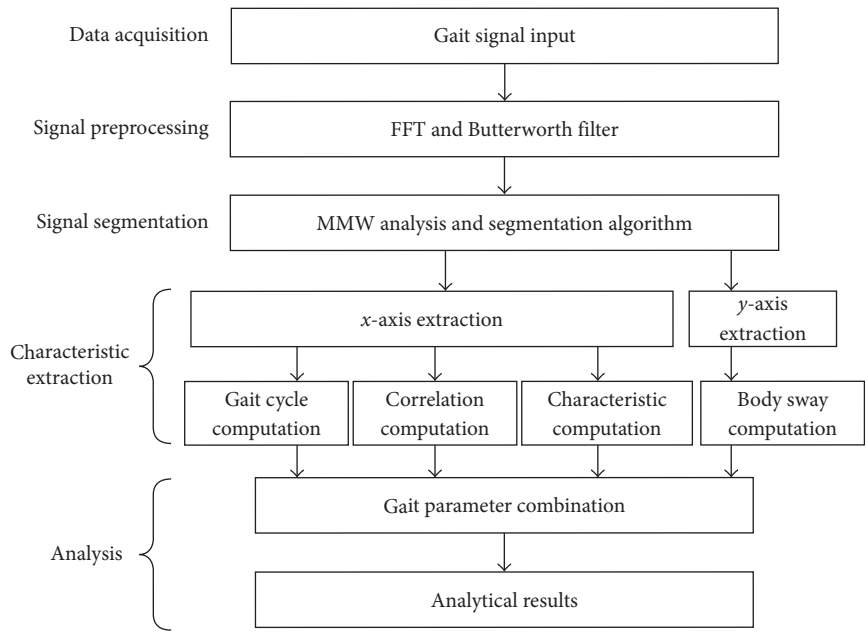


Figure 4: The flow diagram of gait signal processing.

a significant peak ("IC-P"), and six gait events. For the gait characteristics of the left foot, they can be acquired in the same way as above. By locating several sets of the gait cycles as a reference and using specific computer algorithms, it becomes easy and quick to find out the gait events in the recorded video and the corresponding gait characteristics.

Through the synchronization mechanism, the appeared gait characteristics can be verified exactly by the actions of the gait events as shown in Figures 1 and 3. From the temporal positions of the gait events and characteristics, relevant gait parameters can be calculated accordingly.

3. Research Methods

To obtain useful human gait information, there are five stages, that is, data acquisition, signal preprocessing, signal segmentation, characteristic extraction, and analysis, for the signal processing of gait characteristics (Figure 4). For the gait signal acquisition, two wearable triaxial acceleration sensors were used, where each foot was put on one sensor. Since there were many other unrelated signals involved in the raw gait signals from the sensors, the fast Fourier transform (FFT) [22] was used to analyze the noise and signal spectra and a Butterworth filter [23] to remove the noise. Subsequently, the filtered signals on x- and y-axes were segmented according to the gait cycles with the assistance of using the modulus maxima wavelet (MMW) analysis [24] to identify the temporal separation positions. A number of gait characteristics were then extracted to compute the essential gait parameters. Finally, the gait parameters were fused together for the gait analysis of tripping risks in free and fast walking under various cadences regulated by different jazz's tempos.

3.1. Experimental Protocol. In this study, twenty-five healthy participants were recruited for the experiment of walking under different cadences, and each of them had completed

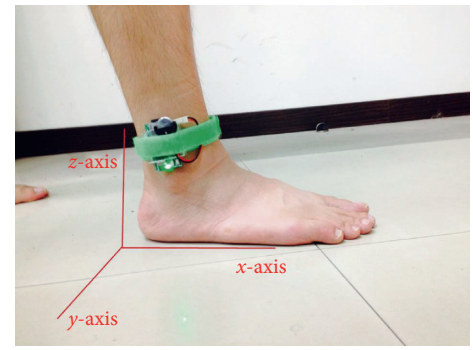


Figure 5: The 3D coordinate system for signals from the wearable triaxial acceleration sensors.

an auditory test to ensure that their hearing and tempo recognition were normal. In terms of performing all gait movements, all participants had been verified to have no medical restrictions on their activities, tempo recognition, and walking. Before the walking tests, all participants only knew that they would take part in the experiment of normal walking, and they did not know that the experiment involved musical rhythm. The data obtained from the cases of without and with playing music tempos would be compared and then be analyzed for the effects of walking on playing tempos for all gait events and periods.

In the walking tests, the sensors were fixed on the feet of the participants at the position 2~3 cm above their ankle of each foot. The sensors measure the gait signals in a 3D coordinate system, where the longitudinal, lateral, and vertical axes are represented as x-, y-, and z-axes as shown in Figure 5. It is worthy to note that the signal component of the longitudinal axis (or x-axis) has the most obvious features of gait while walking and shows mainly the movement characteristics in the acceleration aspect. Therefore, the

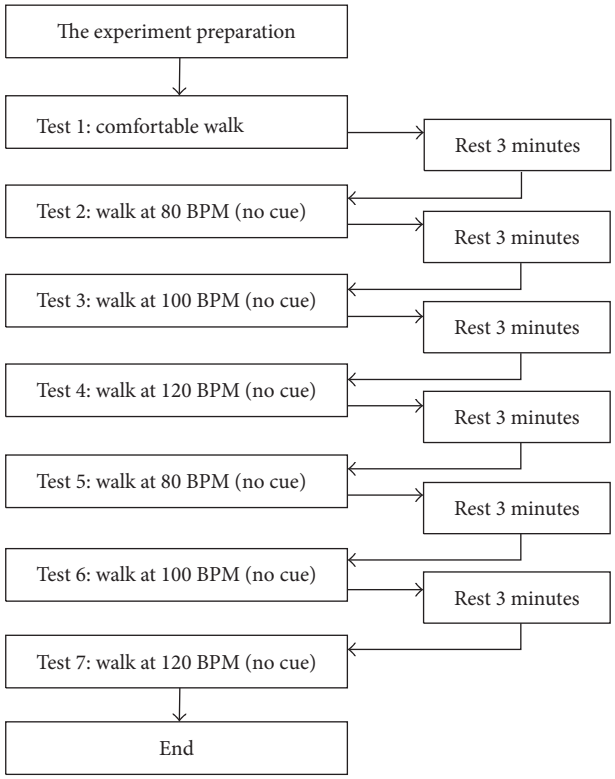


Figure 6: The flow diagram of the walking tests.

x-axial component of the gait signals in the accelerometer was mainly investigated and analyzed to extract the gait parameters in our study. For the signal's lateral axis (or y-axis), the corresponding signal components reveal the left and right accelerations of gait movements, and they help to understand the walking balance of participants. Thus, the y-axial signal components of gait movements would be incorporated into our study of gait analysis. In the observations of the signal's vertical axis (or z-axis), it was found that the signal's variations were greatly with personal habit in walking, and it might be suitable to investigate personal walking habit. Therefore, the signal components of x- and y-axes were considered as the basis to study and analyze the gait characteristics of tripping risks in this study.

After installing the sensors, the participants were asked to walk barefoot to and fro the experimental pathway for 3 minutes so that they could get accustomed to walking naturally wearing the sensors. In the tests, all participants walked barefoot on a 30-meter straight pathway of a ceramic tile floor in 7 different walking tests, which followed the procedure as shown in Figure 6. For the purpose of synchronizing the motion signals and captured images from video, one camera installed at the starting point of the pathway was employed. In addition, there were six cameras set on both sides of the pathway at the positions of 10, 15, and 20 meters from the starting point. These cameras were used to record the gait movements in the tests.

For the first test, the participants were asked to walk in their most natural and comfortable way, which would be treated as the normal gait (Test 1). After that, they were asked to take a 3-minute rest, and meanwhile, jazz music at

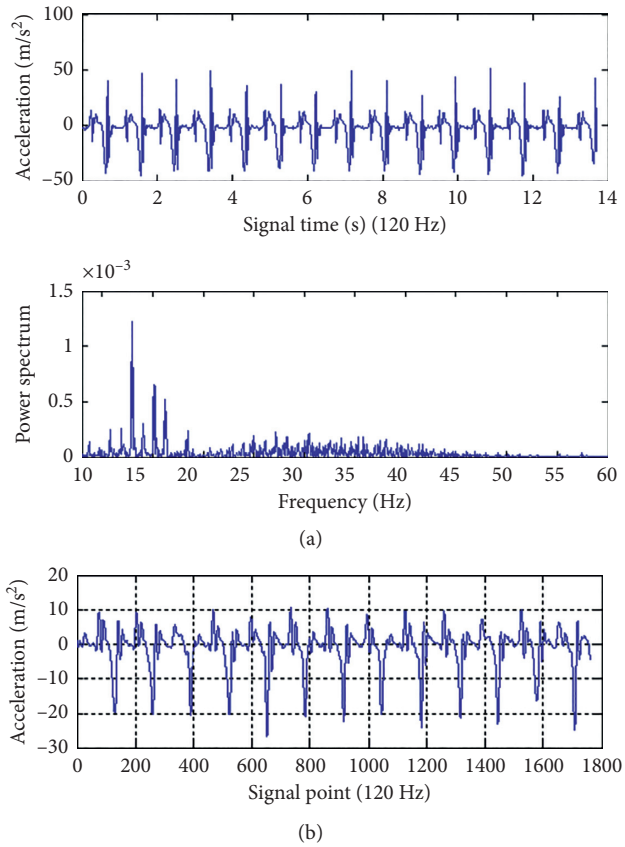


Figure 7: Gait noise removal: (a) the raw gait signals and their FFT spectrum and (b) the filtered gait signals.

80 BPM was played to prepare the participants for the next test. The volume of the music that the participants listened to was around 65~70 db, since the volume of conversation between normal people is usually about 60 db. After the rest, the participants were asked to walk while the jazz at 80 BPM was playing (Test 2), and then, they were asked to take the second 3-minute rest with the jazz at 100 BPM being played. After that, they were asked to walk while the jazz at 100 BPM was playing (Test 3), and then, they took the third 3-minute rest with the jazz at 120 BPM being played. After that, they walked again while the jazz at 120 BPM was playing (Test 4), and then, they took the fourth 3-minute rest with the jazz at 80 BPM being played. Up to this point, the participants completed the tests with no cue.

For the subsequent tests, the participants were asked to listen to the played jazz's tempo and synchronize their walking cadence with the tempo. The fourth 3-minute rest could enable them to adapt their walking cadence to follow the music tempo. The next test made the participants walk to the beats of the jazz at 80 BPM (Test 5). After that, they took the fifth 3-minute rest with the jazz at 100 BPM being played. After the rest, they were asked to walk to the beats of the jazz at 100 BPM (Test 6). Then, after the sixth 3-minute rest with the jazz at 120 BPM being played, they were asked to walk to the beats of the jazz at 120 BPM (Test 7).

3.2. Signal Preprocessing. Since the sensors have very high sensitivity, slight environmental vibrations may cause

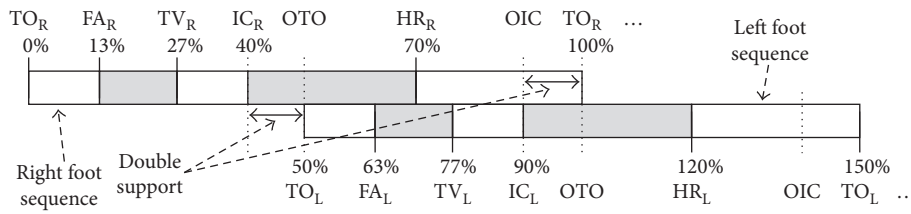


Figure 8: Gait signal synthesis of the left and right feet: the cyclic sequence of gait events.

Step 1. *Signal extrema depiction*: apply an MMW algorithm to the accelerometrical signals (Figure 9(a)) to produce signals' peaks and valleys (Figure 9(b)).

Step 2. *"V-IC" extraction*: apply a threshold $\mathrm{TH}_{\mathrm{V-IC}}$ to filter out those illegal extrema as shown in Figure 9(c). For extracting "V-IC",

(1) retain v_i with $|v_i| < \mathrm{TH}_{\mathrm{V-IC}}$ for the V-IC of the ith cycle, where v_i is the signal of the ith V-IC;
(2) for all consecutive v_i and v_{i+1}, compute temporal_length = $(t(v_{i+1}) - t(v_i))$ in the unit of signal points, where $t(v_i)$ indicates the temporal position of v_i;
(3) if $51 \leq$ temporal_length ≤ 120 signal points (since the normal cadence in walking is about 60~140 BPM), retain both of the v_i and v_{i+1};
(4) if not, discard the bigger one.

Step 3. *"IC-P" extraction*: do the same way as the last step to "V-IC" and then produce p_i.

Step 4. *IC extraction*: for the ith IC, search a x_i point between $t(v_i)$ and $t(p_i)$ with $|x_i| = 0$; mark x_i as the IC characteristic as shown in Figure 9(d).

Step 5. *Gait cycle segmentation*: for each neighboring pair $(\mathrm{IC}_{i-1}, \mathrm{IC}_i)$ in the gait signals of one foot, compute $\mathrm{Gait\_Cycle\_length}_i = (t(\mathrm{IC}_i) - t(\mathrm{IC}_{i-1}))$ in the unit of signal points.

Algorithm 1: Gait cycle segmentation (GCS).

significant and abrupt changes in the acquired signals. As the gait signals were captured and recorded by the sensors, the original signals obtained were often mixed up with many other irrelevant signals, which are collectively called noises. As a consequence, the FFT was employed to find out the major spectrum of the gait movements of common people as shown in Figure 7(a). It can be seen that the frequency distribution is roughly from 0 Hz to 50 Hz. Since the gait movements of common people are usually at a lower frequency band, the Butterworth filter [25], which is a low-pass filter and has high computational efficiency, was used to remove the frequency higher than 20 Hz for the gait analysis in this study. After removing the noise signals, the filtered gait signals can be seen as shown in Figure 7(b).

3.3. Gait Cycle Segmentation and Characteristic Extraction. To investigate gait cycles, the gait characteristics of both left and right feet were synthesized as shown in Figure 8 based on the aforementioned leading clearance matching synchronization mechanism. By observing Figures 1, 3, and 8, a gait cycle of the right foot starts from a TO event ($\mathrm{TO_R}$) and then passes through the characteristics $\mathrm{FA_R}$, $\mathrm{TV_R}$, and $\mathrm{IC_R}$, to complete the swing phase of the right foot. Subsequently, the right foot turns into the stance phase, which has the characteristic sequence of an OTO event (i.e., $\mathrm{TO_L}$), $\mathrm{HR_R}$, and an OIC event ($\mathrm{IC_L}$). For a gait cycle of the left foot, the sequence of gait characteristics is the same as that of the right foot.

In Figure 3, the easiest temporal separation positions to be recognized for the two gait phases are the TO event (a signal peak) and the IC event (a zero-crossing point) between the V-IC point (a signal valley) and the IC-P point (a signal peak). Therefore, if the signal characteristics of TO and IC were identified exactly, then gait phases and cycles could be segmented accordingly. The temporal positions of the above two characteristics could then be used to compute the durations of a stance phase, a swing phase, and the associated gait cycle.

Since it is a tedious work to find out all gait events manually from the recorded video, computer algorithms should be considered to extract the gait characteristics from the triaxial accelerometrical signals. Since the TO event and the zero-crossing point for IC are the most important characteristics to segment and extract all gait characteristics, the detailed working steps for the gait segmentation and characteristic extraction are described in Algorithm 1.

In the GCS algorithm, there are five major steps to locate the temporal positions of V-IC, IC-P, and IC, and then, a gait cycle can be determined. To extract other gait characteristics, a gait events' neighboring relationship shown in Figures 1, 3, and 8 can be found and exploited. For example, the temporal interval between the IC and TO points is about 60% of a gait cycle length; thus, the IC temporal position to the TO should normally be 60% $\pm$ 5%. This neighboring relationship can be utilized to locate the TO temporal position from the IC. The rest characteristics can also be located in the same way as the working steps described in Algorithm 2.

In Algorithms 1 and 2, no working procedure for extracting the HR characteristic was described. This is because no significant signal feature around the HR gait event can be

Step 1. *"TO" extraction*: extract the "TO" of the ith gait cycle by finding the first TO peak, which appears in the interval $[t(\mathrm{IC}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (60\%-5\%), t(\mathrm{IC}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (60\% + 5\%)]$ as shown in Figure 9(e).
Step 2. *"FA" extraction*: extract the "FA" of the ith gait cycle by finding the first FA peak, which appears in the interval $[t(\mathrm{TO}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (13\%-3\%), t(\mathrm{TO}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (13\% + 3\%)]$.
Step 3. *"TV" extraction*: extract the "TV" of the ith gait cycle by finding the first TV peak, which appears in the interval $[t(\mathrm{TO}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (27\% - 3\%), t(\mathrm{TO}_i) + \mathrm{Gait\_Cycle\_length}_{i+1} * (27\% + 3\%)]$.

Algorithm 2: Gait characteristic extraction (GCE).

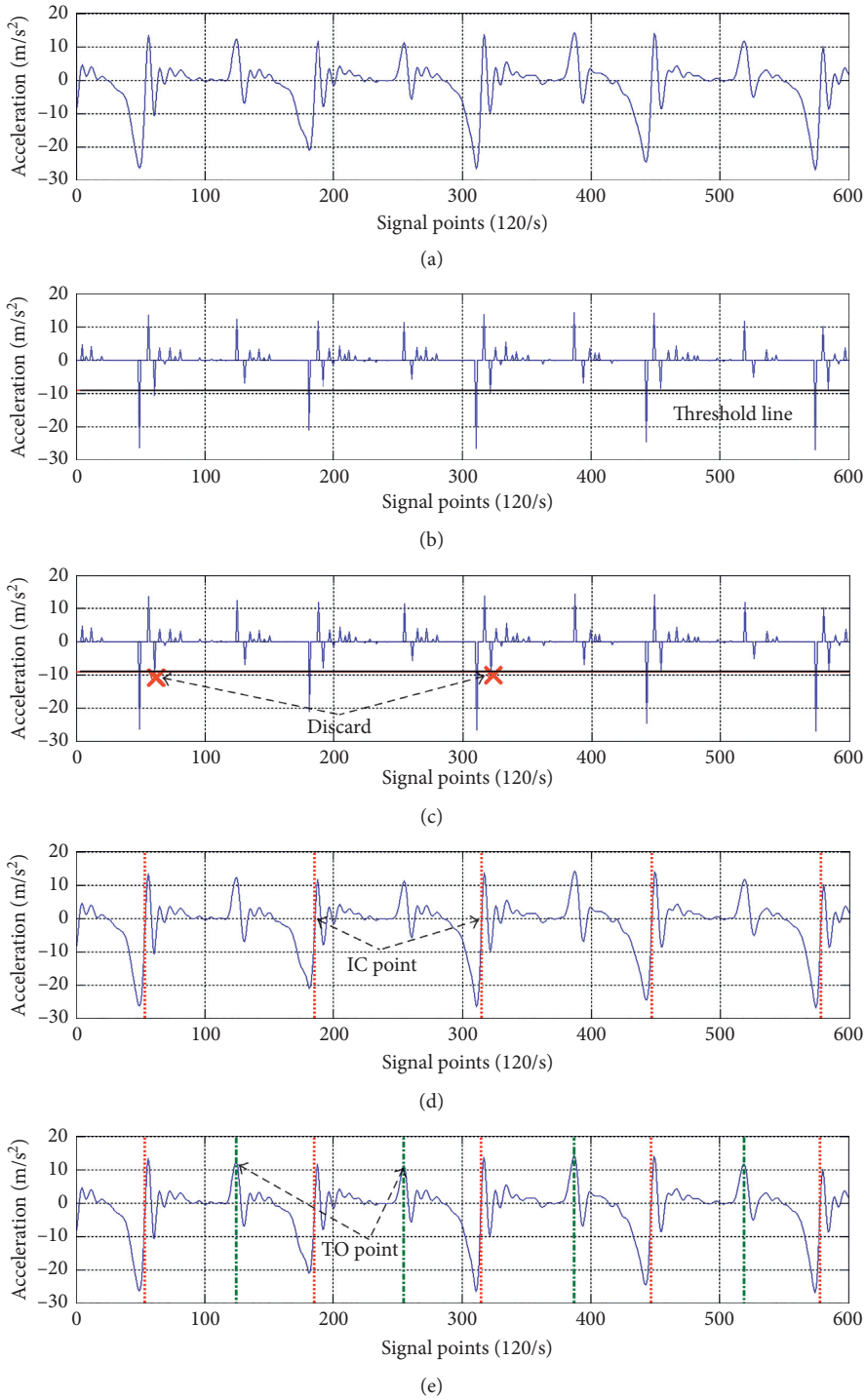


Figure 9: Gait signals processed by the gait cycle segmentation algorithm: (a) filtered signals, (b) setting up a threshold line, (c) unqualified valley points that are discarded, (d) "IC" finding, and (e) "TO" finding.

found. Therefore, the temporal positions of the HR event were extracted manually by observing the recorded video under the supervision of the medical doctor in this study.

3.4. Correlation Analysis. In fast walking, we are interested in exploring the relationship of tripping risks to the gait cycles/phases/events/periods under a chosen cadence. Through a correlation analysis, we could examine the relationship between the cadence and the gait characteristics, which were obtained by the aforementioned algorithms. Subsequently, the correlation coefficients could be calculated to express the trend of positive, negative, or no correlation between two data sets. Let $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_n\}$ be two data sets for the gait characteristics. Then, Pearson's correlation coefficient r_{xy} is computed by

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}. \quad (1)$$

For the interpretation of the correlation coefficient, the positive/negative sign means the trend instead of the degree. The correlation degree between $0\sim\pm0.4$, $0.4\sim\pm0.7$, and $0.7\sim\pm1.0$ indicates that the two data sets have low, moderate, and high correlation, respectively. The correlation degree is ±1.0, which means that the two sets are totally correlated.

4. Results

In the walking experiment, there were totally 2600 gait cycles acquired and stored into 175 records. With the segmentation and extraction algorithms, the characteristics listed in Table 1 were obtained by locating their temporal positions. Then, all gait periods and cycles belonging to each participant were normalized by the ratio in percentage corresponding to the beginning of each gait cycle.

4.1. Cadence versus Gait Events. The correlation analysis for the cadence and relevant characteristics were calculated as shown in Table 1. In this table, we can find that the most important characteristics correlated to the cadence are the TO and V-IC characteristics (Figure 3), which can be used to delineate the phases of a gait cycle. Note that the V-IC point is closely corresponding to the IC event. The correlation coefficients of the TO and V-IC to the cadence are −0.74 and +0.86, respectively. Thus, the higher the cadence is, the larger the motion strength in the terminal swing period (between the TV and IC events) will be applied and, also, the smaller the motion strength in the TO event due to the negative correlation.

In Table 1 and Figure 8, the FA and OHR (opposite HR) characteristics, which are corresponding to the FA and opposite HR events, show the relevant correlation with the cadence. These two characteristics can be corresponded to the gait periods of the midswing and the opposite terminal stance with the correlation coefficients of −0.64 and −0.72, respectively. It indicates that the higher the cadence is, the shorter the two periods will be.

Basically, the normal walking speed in daily life of a person is equal to the product of his/her step length and the cadence. Since the average step length in walking of a person is highly related to his/her height, the walking speed is mostly affected by the cadence. In the order of absolute values of the correlation coefficients, the cadence variation in walking is mainly affected by the characteristics of V-IC (related to the TV and IC events, or the terminal swing period), TO, and OHR (or the opposite terminal stance period). For example, the signal curve around V-IC changes prominently in the terminal swing period because the swinging foot conducts greater acceleration and deceleration in fast walking. When walking slowly, the situation becomes contrapositive.

From the above analysis, it can be easily found that higher cadences trigger larger motion strength of the swinging gait events and shorten the action time in the terminal swing and opposite terminal stance periods. This situation can likely raise the tripping risks provided that an accident hinders the progression of continuous gait events in the swing phase.

4.2. Cadence versus Gait Periods. In Figure 9, the V-IC characteristic acts as an anchor point for segmenting each gait cycle. We analyzed the correlation of the cadence and gait periods for each gait cycle and obtained the average correlation coefficients as listed in Table 2. In this table, the correlation coefficient of "No_cue" was obtained based on playing a music tempo without reminding the participants to notice the tempo, and for "Cue," the participants were informed to follow the music tempo. Obviously, the music tempo influences slightly the swing phase of every gait cycle in the "No_cue" case, since the beats can still influence the participants by hearing. In "Cue," every gait period was influenced greatly by different musical rhythms so that the tempo changed the gait cycles and produced different walking speeds.

Table 1: Correlation analysis of the cadence and relevant gait characteristics.

	Cadence	TO	FA	OHR	TV	V-IC	IC-P
Cadence	1	−0.74	−0.64	−0.72	−0.32	0.86	0.03
TO	−0.74	1	0.75	0.59	0.28	−0.72	0.27
FA	−0.64	0.75	1	0.76	0.18	−0.52	0.04
OHR	−0.72	0.59	0.76	1	0.54	−0.73	0.25
TV	−0.32	0.28	0.18	0.54	1	−0.52	0.19
V-IC	0.86	−0.72	−0.52	−0.73	−0.52	1	−0.29
IC-P	0.03	0.27	0.04	0.25	0.19	−0.29	1

Table 2: Correlation analysis of the cadence and gait periods.

Period	Type	
	No_cue	Cue
Preswing	0.10	0.82
Initial swing	−0.10	−0.83
Midswing	−0.22	−0.79
Terminal swing	0.20	0.82
Loading response	−0.02	0.82
Midstance	0.05	−0.86
Terminal stance	−0.03	0.84

Table 3: Percentage of the left/right gait phase with respect to the normalized gait cycle in seven different walk types.

Phase (%)	Type						
	Normal	NC_80	NC_100	NC_120	C_80	C_100	C_120
Left swing	42.22	42.00	41.67	42.00	38.00	41.22	43.00
Right swing	41.00	41.78	40.67	41.33	37.89	40.89	41.33
Left – right	1.22	0.22	1.00	0.67	0.11	0.33	1.67
Left stance	57.78	58.00	58.33	58.00	62.00	58.78	57.00
Right stance	59.00	58.22	59.33	58.67	62.11	59.11	58.67
Left – right	−1.22	−0.22	−1.00	−0.67	−0.11	−0.33	−1.67

4.3. Gait Phase Variation versus Walk Types. Since the music tempos could greatly affect gait cycles, the gait phase separation on both feet was done with Algorithms 1 and 2 as shown in Table 3, where "Normal" means walking comfortably and naturally; NC_80, NC_100, and NC_120 mean walking in the "No_cue" cases with tempo 80, 100, and 120 BPM; and C_80, C_100, and C_120 mean walking in the "Cue" cases with tempo 80, 100, and 120 BPM.

As walking in "Normal" and "No_cue," the gait phases of both feet differ only around the range of less than 1.22%. For walking in "Cue," the difference in C_80 and C_100 is small ($\leq$0.33%), but it becomes large (=1.67%) in C_120. In addition, the gait phase variation increases as the cadence increases in "Cue."

In Table 3, it can be seen that the left foot spends longer time in the swing phase and shorter time in the stance phase in comparison with the right foot. It reveals that the participants would like to match the beat with the left IC gait event so that the gait cycles of the right foot tend to vary greatly to minimize the speculation of the next beat. In addition, there is no obvious tendency of the gait phases in the cases of "Normal" and "No_cue." However, in "Cue," a greater swing time variation happens between C_100 and C_120. It shows an increasing tendency that the left foot spends longer time in swing.

To disclose more messages from the above observations, we examine the gait signals on the y-axis, which can reveal the situations of left and right body sways, as shown in Figure 10. Note that the gait signals in acceleration on y-axis in Figure 10 are corresponding to the gait signals on x-axis in Table 3. Since the gait cycle (particularly in stance) and body sway have a close relationship to the balance, the extreme values of acceleration on y-axis can be used as a balance indicator for the fast walk. Thus, the walking balance in gait is investigated according to different walking speeds based on with and without musical cue. In Figure 10, a box plot shows an interval between a pair of acceleration's maximum and minimum and an interquartile range of acceleration for one walk type.

According to the walk types, we can see that the box plots for the types in "No_cue" are almost similar to each other, and for "Normal," its box plot is a little smaller to the cases of "No_cue." This means that walking comfortably and naturally and walking without caring about the cue may have stable and balanced movements in gait with the lower body sway. As walking in "Cue," the increasing tendency of the body sway matches the increase of cadence. Especially, the case of C_120 is the most unstable and unbalanced in all cases due to the large range of acceleration from the left and right body sways. In addition, the case of C_80 is the most stable and balanced in all cases since the low cadence regulated by the slow tempo benefits the balance control of body's gait movements.

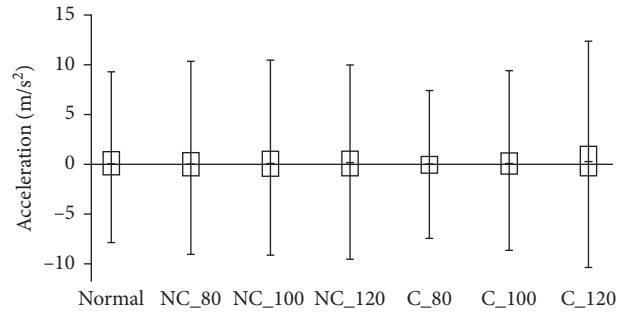


Figure 10: The box plot of the gait cycle variations in y-axis.

4.4. Gait Periods versus Walk Types. In Table 4, the percentage of each left/right gait period is calculated with respect to each gait cycle occurring in the seven different walk types. Noteworthily, the gait periods of all participants are consistent after the normalization of various gait cycles. In this table, we can see that the walk types of "Normal" and "No_cue" have no big difference in the seven gait periods since the walking cadence was determined by the participants. For the walk types of "Cue,", it can be observed that the gait periods of the left/right preswing and loading response are not sensitive to the higher cadence variation (i.e., 100 and 120 BPM) but is sensitive to the slow tempo such as 80 BPM.

For the gait periods of the initial swing, midswing, and midstance of both feet, their period lengths in percentage increase as the cadence increases. For the terminal swing and terminal stance, their period lengths in percentage decrease as the cadence increases. Thus, we can find that the musical rhythm can intervene with the walking speed by regulating the cadence, which will affect the increase or decrease of gait period lengths.

In addition, we find that the increments of the gait period lengths for the midstance, midswing, and initial swing are more than those in all other cases, and the same situation happens for the decrements of the period lengths for the terminal swing and terminal stance. This indicates that there exists a tripping risk: the longer time for doing the midswing

Table 4: Percentage of the left/right gait periods with respect to each gait cycle occurring in the seven walk types.

Period (%)	Type						
	Normal	NC_80	NC_100	NC_120	C_80	C_100	C_120
Left preswing	9.51	9.09	9.97	8.86	11.21	9.57	9.35
Left initial swing	13.00	13.22	12.78	13.22	11.33	13.56	14.11
Left midswing	12.89	12.22	13.78	13.78	11.00	13.78	15.00
Left terminal swing	16.33	16.56	15.11	15.00	15.67	14.89	13.89
Left loading response	9.27	9.13	9.69	8.80	10.90	9.32	9.32
Left midstance	19.89	19.56	20.11	19.78	18.11	19.22	21.33
Left terminal stance	19.11	20.22	18.56	20.56	21.78	19.67	17.00
Right preswing	9.19	9.21	9.86	9.40	11.16	9.37	9.30
Right initial swing	13.78	13.33	13.00	13.00	10.78	12.44	15.00
Right midswing	14.44	13.89	13.89	14.44	11.22	13.33	14.56
Right terminal swing	12.78	14.56	13.78	13.89	15.89	15.11	11.78
Right loading response	9.59	9.01	9.80	9.26	10.95	9.52	9.36
Right midstance	19.89	19.33	19.11	20.00	18.78	19.44	21.78
Right terminal stance	20.33	20.67	20.56	20.00	21.22	20.78	18.22

means the stride becomes bigger, and the shorter terminal swing means the movements in the tibia vertical to knee extension (before the initial contact) become faster.

If the walking paths were not flat and there were bumps that might block the movements in the tibia vertical to knee extension, a tripping risk would raise greatly due to the longer time in midstance blocking the next stance support while the center of body mass is moving ahead of the next stance support point. In summary, larger steps or strides in walking maintain larger gait speed and consume more calories. In this way, cadence can be kept slower to prevent tripping risks, but we still have higher gait speed to consume calories and promote health.

5. Conclusion

Due to technological and spatial limitations, fast walking for long distance in daily life is difficult to be observed by scientific equipment. In the literature, most of the gait experiments were done by walking on treadmills [5, 7, 8, 10]. Actually, gait signals acquired on treadmills are quite different from the signals on the ground. This hinders the observation of a fast gait on ordinary ground. In this study, we conducted all the walking tests in a flat and straight pathway, intended to reveal important clues for tripping risks in fast walking.

In the walking tests, music cueing in three tempos were employed to regulate the walking speed of participants, and seven walk types, which include one "Normal," three "No_cue," and three "Cue" tests, were devised for the acquisition of gait signals. From the extracted gait characteristics, we found that "V-IC" is an important sign that has the largest signal variation in one gait cycle and is highly correlated with cadences. Thus, it can be used to compute the cadence of one specific walk.

In the "Normal" and "No_cue" tests, the difference in cadences, gait phases, and gait periods in both types is minor since the background music tempos in "No_cue" are mostly ignored by the participants. Meanwhile, the left and right body sways in the frontal plane are insignificant. It means that the gaits are stable and balanced in body movements.

In the "Cue" tests, every gait period is influenced greatly by different music tempos, which may cause various walking speeds. Firstly, a larger time variation in foot swing exists between the cadences in 100 and 120 beats. By observing the gait period's variation in the midswing and terminal swing, the longer midswing time means the bigger stride, and the shorter terminal swing time indicates the faster movements in the tibia vertical to knee extension. Secondly, by the observation to a higher cadence (e.g., 120 beats), another factor of tripping risks may exist due to the high negative correlation of midstance and midswing to the higher cadence. This condition causes smaller foot clearance over the ground and increases greatly the probability of tripping. As walking through ordinary or uneven ground, the tripping risk of passing over bumps becomes substantial in faster movements.

A walking speed is determined from a cadence multiplied by a stride length, and it indicates that a bigger stride length needs only a lower cadence to keep the same walking speed. Therefore, in the clinical practice, the principle of fast walking safely in the outdoors should be based on a lower cadence with bigger strides for health promotion. Remarkably, the aforementioned results show that the cadence below 120 beats is preferred. From another point of view, a completely flat ground becomes a necessary and sufficient condition for fast walking. The principle can also apply to indoor activities or locomotion since our experimental ground is close to a home environment.

Ethical Approval

This study was approved by the Institutional Review Board of the National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University (Approval no. A-ER-104-268).

Consent

All participants gave written informed consent to take part in this research.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors are grateful to the MOST, Taiwan, for supporting this study under Contract nos. MOST105-2622-E-224-003-CC3 and MOST106-2314-B-006-017-MY3.

References

[1] G. Abbud, K. Li, and R. DeMont, "Attentional requirements of walking according to the gait phase and onset of auditory stimuli," *Gait and Posture*, vol. 30, no. 2, pp. 227–232, 2009.

[2] M. Tsuruoka, R. Shibasaki, and Y. Tsuruoka, "Analysis of 1/f fluctuations of walking listening to Mozart's music," in *Proceedings of the International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 62–65, Perth, Australia, October 2010.

[3] A. Delval, C. Moreau, S. Bleuse et al., "Auditory cueing of gait initiation in Parkinson's disease patients with freezing of gait," *Clinical Neurophysiology*, vol. 125, no. 8, pp. 1675–1681, 2014.

[4] A. B. Bourgeois, B. Mariani, K. Aminian, P. Zambelli, and C. Newman, "Spatio-temporal gait analysis in children with cerebral palsy using, foot-worn inertial sensors," *Gait and Posture*, vol. 39, no. 1, pp. 436–442, 2014.

[5] L. R. Nascimento, C. Q. de Oliveira, L. Ada, S. M. Michaelsen, and L. F. Teixeira-Salmela, "Walking training with cueing of cadence improves walking speed and stride length after stroke more than walking training alone: a systematic review," *Journal of Physiotherapy*, vol. 61, no. 1, pp. 10–15, 2015.

[6] M. Roerdink, P. J. Bank, C. L. E. Peper, and P. J. Beek, "Walking to the beat of different drums: practical implications for the use of acoustic rhythms in gait rehabilitation," *Gait and Posture*, vol. 33, no. 4, pp. 690–694, 2011.

[7] C. Mendonca, M. Oliveira, L. Fontes, and J. Santos, "The effect of instruction to synchronize over step frequency while walking with auditory cues on a treadmill," *Human Movement Science*, vol. 33, pp. 33–42, 2014.

[8] D. Eikema, L. Forrester, and J. Whitall, "Manipulating the stride length/stride velocity relationship of walking using a treadmill and rhythmic auditory cueing in non-disabled older individuals. A short-term feasibility study," *Gait and Posture*, vol. 40, no. 4, pp. 712–714, 2014.

[9] J. E. Wittwer, K. E. Webster, and K. Hill, "Music and metronome cues produce different effects on gait spatiotemporal measures but not gait variability in healthy older adults," *Gait and Posture*, vol. 37, no. 2, pp. 219–222, 2013.

[10] K. Byung-Woo, H.-Y. Lee, and W.-K. Song, "Rhythmic auditory stimulation using a portable smart device: short-term effects on gait in chronic hemiplegic stroke patients," *Journal of Physical Therapy Science*, vol. 28, no. 5, pp. 1538–1543, 2016.

[11] B. Mariani, H. Rouhani, X. Crevoisier, and K. Aminian, "Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors," *Gait and Posture*, vol. 37, no. 2, pp. 229–234, 2013.

[12] T. Khurelbaatar, K. Kim, S. Lee, and Y. H. Kim, "Consistent accuracy in whole-body joint kinetics during gait using wearable inertial motion sensors and in-shoe pressure sensors," *Gait and Posture*, vol. 42, no. 1, pp. 65–69, 2015.

[13] L. Z. Rubenstein, "Falls in older people: epidemiology, risk factors and strategies for prevention," *Age and Ageing*, vol. 35, no. 2, pp. ii37–ii41, 2006.

[14] S. Lord, C. Sherrington, H. Menz, and J. Close, *Falls in Older People: Risk Factors and Strategies for Prevention*, Cambridge University Press, Cambridge, UK, 2007.

[15] M. W. Whittle, *Gait Analysis: An Introduction*, Butterworth-Heinemann, Edinburgh, Scotland, 4th edition, 2007.

[16] M. Nordin and V. H. Frankel, *Basic Biomechanics of the Musculoskeletal System*, Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2001.

[17] D. J. Magee, *Orthopedic Physical Assessment*, Elsevier Health Sciences, Oxford, UK, 2014.

[18] Texas Instruments, *MSP430F552x, MSP430F551x Mixed-Signal Microcontrollers*, 2015, http://www.ti.com/lit/ds/slas590m/slas590m.pdf.

[19] STMicroelectronics, *LIS3DH: MEMS Digital Output Motion Sensor*, 2011, http://www.st.com/resource/en/application_note/cd00290365.pdf.

[20] Sony, *Digital HD Video Camera Recorder Handbook*, 2014, http://www.sony.ee/support/en/content/cnt-man/HDR-AS100V/list.

[21] I. Monson, *Saying Something: Jazz Improvisation and Interaction*, University of Chicago Press, Chicago, IL, USA, 2009.

[22] K. R. Rao, D. N. Kim, and J. J. Hwang, *Fast Fourier Transform–Algorithms and Applications*, Springer, Netherlands, 2010.

[23] Electronics Hub, *Butterworth Filter*, 2015, http://www.electronicshub.org/butterworth-filter/.

[24] L. M. Bruce and R. R. Adhami, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Transactions on Medical Imaging*, vol. 18, no. 12, pp. 1170–1177, 1999.

[25] K. S. Erer, "Adaptive usage of the Butterworth digital filter," *Journal of Biomechanics*, vol. 40, no. 13, pp. 2934–2943, 2007.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 2548537, 13 pages
https://doi.org/10.1155/2018/2548537

Research Article

Unsupervised Medical Entity Recognition and Linking in Chinese Online Medical Text

Jing Xu, Liang Gan, Mian Cheng, and Quanyuan Wu

School of Computer, National University of Defense Technology, Changsha, China

Correspondence should be addressed to Jing Xu; jing.xu@nudt.edu.cn

Received 28 September 2017; Accepted 2 January 2018; Published 18 April 2018

Academic Editor: Robertas Damaševičius

Online medical text is full of references to medical entities (MEs), which are valuable in many applications, including medical knowledge-based (KB) construction, decision support systems, and the treatment of diseases. However, the diverse and ambiguous nature of the surface forms gives rise to a great difficulty for ME identification. Many existing solutions have focused on supervised approaches, which are often task-dependent. In other words, applying them to different kinds of corpora or identifying new entity categories requires major effort in data annotation and feature definition. In this paper, we propose unMERL, an unsupervised framework for recognizing and linking medical entities mentioned in Chinese online medical text. For ME recognition, unMERL first exploits a knowledge-driven approach to extract candidate entities from free text. Then, the categories of the candidate entities are determined using a distributed semantic-based approach. For ME linking, we propose a collaborative inference approach which takes full advantage of heterogenous entity knowledge and unstructured information in KB. Experimental results on real corpora demonstrate significant benefits compared to recent approaches with respect to both ME recognition and linking.

1. Introduction

In recent years, due to the rapid development of techniques and the increasing concern of people with their health, many medical websites have emerged which not only provide diverse medical information, including health knowledge and medical news, but also provide the online consultation service about diseases. Some well-known Chinese medical websites are Family-doctor (http://www.familydoctor.com.cn/), Muzhi-doctor (http://muzhi.baidu.com/), Qiuyi (http://www.qiuyi.cn/) and so on, which produce a large amount of medical question and answer (Q&A) data based on real patients and doctors every day. This data contains many real individual cases with high medical value, motivating many medical applications, such as disease prevention and self-treatment.

Medical Q&A data, as unstructured text expression, contains many diverse and ambiguous references to medical entities. The diversity is that an entity is referred to in multiple ways, including aliases and abbreviations. The ambiguity means that different entities have the same surface form. For example, "传染病" (epidemic) could refer to either a disease or a film. This gives rise to a great difficulty in ME identification. Only using entity recognition technology is limited in terms of its ability to effectively mine the data. To fully mine and exploit useful medical knowledge, ME recognition and linking is a good solution. Specifically, it first detects and classifies the ME mentions in text and then understands their meanings by linking the mentions to the correct entities in a given KB. For example, given a text such as "MF, 即骨髓纤维化, 症状为脾肿大…" (the symptom of MF, namely, myelofibrosis, is splenomegaly), ME recognition determines that the strings "MF" and "骨髓纤维化" (myelofibrosis) are diseases and that "脾肿大" (splenomegaly) is a symptom. ME linking performs the next step, inferring that "MF" and "骨髓纤维化" (myelofibrosis) actually refer to an entity at URL "http://baike.baidu.com/item/骨髓纤维化" and that "脾肿大" (splenomegaly) refers to an entity at URL "http://baike.baidu.com/item/脾大."

Medical entity recognition (MER) is a well-known problem which has been studied for decades. Medical entity linking (MEL) is a newer research issue which has attracted

much attention because of its importance in many applications, such as understanding medical text, KB construction, and Q&A systems. However, existing works on this topic mainly focus on well-formed English text, such as electronic patient records and medical reports. Few studies have focused on Chinese online medical Q&A text data. The research challenges can be summarized as (1) the online medical Q&A text is characterized by unreliable tokenization, abbreviation, and misspellings. This gives rise to a great difficulty in recognizing the correct entity boundary. (2) It is generally brief, lacking rich context information. This affects the availability of context that can be leveraged to assist the linking. (3) Compared to English, Chinese has more complicated syntax rules, so it is difficult to use solutions for the English language.

In this paper, we design an unsupervised framework that recognizes and links ME mentions in Chinese online medical text, namely, unMERL. To the best of our knowledge, this is the first paper that describes such a comprehensive framework for Chinese medical text. The main contributions of this work are as follows.

(1) unMERL utilizes a knowledge-driven approach to detect the ME boundaries, which incorporates the offline and online process, thereby significantly improving recognition performance. In addition, the strategy exploiting the dependency relationships between words can capture the nested and combined medical entities well.

(2) unMERL uses an improved classifier based on text feature computation and semantic signature similarity, which can efficiently classify medical entities and further filter nonmedical entities.

(3) The linking approach synthetically considers the name similarity, entity popularity, category consistency, context similarity, and semantic correlation between entities, which can better distinguish and determine the candidate entities. In addition, to solve the imperfection problem of the KB, we introduce an incremental evidence mining process, thereby significantly improving the linking performance.

(4) We extensively evaluate unMERL for the ME recognition and linking task over real datasets. The experiment results show that unMERL can achieve a significantly higher performance compared to current mainstream methods.

The remainder of this paper is organized as follows. Section 2 discusses related works; Section 3 presents our framework in detail; Section 4 describes our experiments, results, and discussion; and Section 5 concludes this paper.

2. Related Works

Entity recognition has been widely studied in the context of the medical domain. Early works on this topic relied on heuristic rules and lexical resources [1–4]. Based on the name characteristics of medical entities, the researchers encoded and mapped the terms in clinical text to the lexical resources. In particular, the widely applied systems include MedLEE [1], EDGAR [2], and MetaMap [3]. The most well-known medical lexicons included MetaThesaurus [5], MeSH (Medical Subject Heading) [6], and SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) [7]. The Chinese version of SNOMED-CT was published in 1997. The rule-based and lexicon-based systems depended on name regularity and lexicon size, which were restricted to extracting the limited and normative entities. However, by incorporating the dependency relationships between words and the online detection process with a search engine, our approach solves these problems well.

More recently, Zhang and Elhadad proposed an unsupervised approach to biomedical-named entity recognition, leveraging terminologies, syntactic knowledge, and corpus statistics [8]. In addition, the bootstrapping algorithm attracted much attention in the context of medical entity recognition [9, 10]. Bootstrapping is an unsupervised machine learning approach which starts from small sets of seeds or rules and iteratively labels the corpus with them by pattern matching [11]. However, it relies on the quality of seeds and the normalization of the corpus, which easily produces semantic deviation, due to involvement of the incorrect seed categories and irregular context information.

In recent years, many researchers have focused on using statistical machine learning approaches in the medical field. The ME recognition problem is transformed into a sequence annotation or a classification problem. The lexical, syntactic, and semantic features of words are used for training various learning models such as HMM (hidden Markov model) [12, 13], MEM (maximum entropy model) [14, 15], CRF (conditional random field) [13, 16–19], and (structured) SVM (support vector machine) [19]. In addition, to alleviate the limitation of a single model, some researchers proposed the cascading methods [15, 20] which combine multiple models, including CRF, (structured) SVM, and MEM. However, the supervised nature of the machine learning-based approaches relies on a large amount of training corpus which need to be annotated by humans. Besides, it is difficult for the feature set to cover all entity types. As a result, they are usually task-dependent. To solve this problem, we propose an unsupervised approach which leverages syntactic knowledge, corpus statistics, and lexical resources for ME recognition.

Medical entity linking is a newer problem. Some effective approaches to English corpora have been proposed. Glavas exploited semantic textual similarity for linking entity mentions in clinical text [21]. Zheng et al. proposed a collective inference approach which leverages semantic information and structures in ontology to solve the entity linking problem for biomedical literature [22]. Wang et al. proposed a graph-based linking approach which first constructs graphs for mentions, KB, and candidates and then exploits the information entropy and similarity algorithm to link biomedical entities [23]. These approaches are dependent on the context and KB. Therefore, the noise and lack of information in the context reduce the accuracy of the linking. In addition, the

graph-based approaches have a high computation cost, and the imperfection of KB also impacts the performance of the linking.

Our linking approach synthetically considers multiple entity knowledge, which is more accurate in distinguishing and determining the candidate entities, with lower computational costs. Moreover, our solution adds the step of extracting the relevant context, to solve the noise problem. To optimize the local KB, we still introduce an incremental evidence mining process with the third KB. Entity linking in the Chinese medical domain has been studied less than entity linking in the English medical domain. To our knowledge, our linking approach is the first solution for Chinese online medical text.

3. UnMERL

The framework of unMERL is shown in Figure 1, in which unMERL consists of two modules. The ME recognition module consumes an input corpus and performs entity boundary detection and entity classification. The output is a set of medical entities and categories. For each recognized medical entity, the ME linking module generates the candidates from the KB and then acquires the target object by ranking them.

3.1. Medical Entity Recognition. The ME recognition module aims to detect and classify all ME mentions in the input corpus. Named entity recognition (NER) [24] involves two main steps: detecting entity boundaries and classifying the entities into predefined categories. Based on the thesis, our ME recognition module is implemented in the sequence of two separate processes: boundary detection and entity classification.

3.1.1. Boundary Detection. This step requires the detection of boundaries of medical entities, collecting candidates for entity classification. In our solution, unMERL exploits a knowledge-driven method, mapping the input text to concepts in the lexical resources. Compared to the existing dictionary-based approaches, our approach differs in the following ways: (1) The entity candidates are identified based on the dependency relationships between words. The strategy can well capture the combined and nested entities and reduce the computational cost of the subsequent process by downsizing the candidate set. (2) The search engine is included as a lexical resource, which breaks the conditionality of the limited terms in the dictionary and has good performance in terms of its ability to detect variational and rare entity names. The detection process is roughly divided into two stages: candidate entity generation and medical entity detection.

(1) Candidate Entity Generation. Through corpus analysis, we find that a long medical entity is usually segmented into several fragments by a common nature language processing tool. The POS tag of each fragment is included in Table 1. Moreover, these fragments generally have an attributive dependency relationship. For example, the text "骨髓纤维化简称髓纤, 是一种骨髓增生性疾病, 武汉协和医院有很好的治疗效果" (myelofibrosis, or MF in brief, is a myeloproliferative disease, for which Wuhan Concorde Hospital has a very good therapeutic effect), parsed by the HanLP dependency parser (http://hanlp.linrunsoft.com/), is shown in Figure 2. The dependency labels are shown in Table 1. Based on the hypothesis that entities should be noun phrases (NPs), from the automatically parsed dependency trees, we extract native NPs as candidate entities. A native NP is a single noun (without the attributive modifiers) or a maximum noun phrase with the POS tags in Table 1 and the dependency "ATT." The candidate entities extracted from the above text are shown in the third row of Table 1.

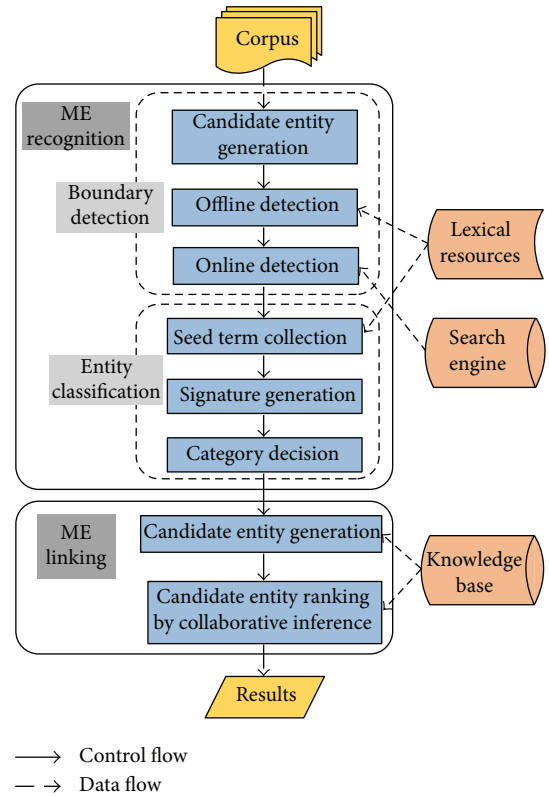


Figure 1: Architecture of unMERL.

However, not all noun phrases are medical entities. In order to remove the nonmedical NPs, we employ a knowledge-driven method whose aim is to discover the concepts in the lexical resources referred to in the text. Here, we use Chinese SNOMED-CT [25], the medical KB of Baidu Baike (https://baike.baidu.com/science/medical), and Sogou medical dictionaries (http://pinyin.sogou.com/dict/cate/index/132?rf=dictindex) as the offline lexical resources (LRs). In order to mitigate the limited coverage of the above resources, we still use Baidu Search (https://www.baidu.com/) as an online lexical resource to help recognize the medical entities.

(2) LR Description. As mentioned before, Chinese SNOMED-CT, translated from SNOMED-CT (English), is a standard of clinical medicine and contains more than 140,000 clinical terms, covering most aspects of clinical information. To correct incorrect terms in the translated version, we add the

Table 1: Constraints on POS tags, description of dependency labels, and candidate entities of the sentence in Figure 2.

Notation	Description
POS tags	f (preposition of locality), m (measure word), b (distinguishing word), rr (personal pronoun), v (verbal word), gb* (word related to biology), or n* (noun)
Dependency labels	HED (head), SBV (subject-verb), VOB (verb-object), ATT (attribution), COO (coordination), RAD (right adjunct)
Candidate list	骨髓纤维化 (myelofibrosis), 髓纤 (MF), 骨髓增生性疾病 (myeloproliferative disease), 武汉协和医院 (Wuhan Concorde Hospital), 治疗效果 (therapeutic effect)

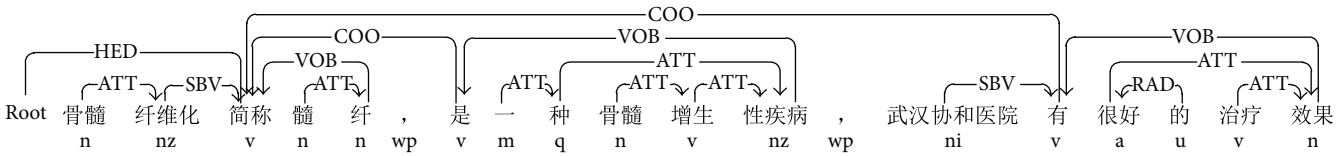


Figure 2: An example sentence with dependency parsing and POS tagging.

medical KB of Baidu Baike and Sogou medical dictionaries. Baidu Baike contains more than 25,000 medical terms edited by authoritative organizations and experts. Sogou medical dictionaries, as the lexicon resource of Sogou's input method, collect data from multiple medical websites. Baidu Search is the largest Chinese search engine, using which we can obtain information that standardized LRs do not cover, such as emerging, rare, and variational medical entities.

(3) LR Preprocessing. Considering the heterogeneity and redundancy of the above offline LRs, we extract and fuse the medical terms from them to build a dictionary. In particular, we select specific categories of interest, which are also the goal of our entity classification. Table 2 presents the statistics in the self-built dictionary. In addition, to improve retrieval efficiency, we build indices by using the first phonetic alphabet of each term.

$$t_m = \text{maxOccur}(r_1, r_2, \ldots, r_j),$$
$$r_j = \begin{cases} \text{LCS}(t_k, s_k), (k \in K), \text{if Len}(\text{LCS}(t_k, s_k)) > 1, \\ \text{LCS}(t_k, t_i), (k, i \in K, k \neq i). \end{cases} \quad (1)$$

Given a candidate entity, the results returned by Baidu Search contain not only the objective medical term but also other noise information that impacts the performance of the entity recognition. Therefore, we need to process the search results to obtain an unmixed medical term. Based on the common knowledge that there are more correct results than incorrect results, the method is implemented based on corpus statistics and "LCS" (a function of achieving the longest common substring), as shown in (1). Given the search result set $S = \{(t_k, s_k)\}_{k=1}^{K}$ (t_k represents a title, and s_k represents a summary.), we first get the kernel term from each result by using the "LCS" function. However, not all summaries contain the kernel terms in the titles. Therefore, we add the process $\text{LCS}(t_k, t_i)$ for the search results without common substring. Finally, in the kernel term set, we select the most frequent term t_m as a correct medical term.

Table 2: Statistics of the user-defined dictionary.

Category	Term number
Body	1802
Disease	48,120
Symptom	3698
Medicine	42,047
Treatment	7403
Check	768

In addition, considering that the search engine has no distinguishing ability to filter the nonmedical entities in the candidate set, we establish a medical keyword set. This includes "医" (medicine), "药" (drug), "病" (disease), and "症" (symptom). If a search result contains one or more keywords in the above set, we identify the candidate as a medical entity. If not, it is removed as a nonmedical entity.

(4) Medical Entity Detection. Once the medical terms are acquired from the offline and online LRs, detecting the medical entities from the candidate set can be performed. Based on the different characteristics of LRs (that the offline LRs have high accuracy but limited coverage and the online LR have high coverage but lower accuracy), we divide the detection process into offline and online processes. Given the candidate set, unMERL first performs the offline detection with the self-built dictionary. For the output nonmedical candidates, unMERL performs the online detection with Baidu search engine. Here, we exploit the string matching and text distance constraint to implement the detection process.

$$\text{Sim}_r(t_{\text{cm}}, t_m) = \frac{\text{Len}(\text{LCS}(t_{\text{cm}}, t_m))}{\min(\text{Len}(t_{\text{cm}}), \text{Len}(t_m))}, \quad (2)$$

$$\text{Dis}(t_{\text{cm}}, t_m) = \left|\text{Loc}\left(d_{t_{\text{cm}}}, t_{\text{cm}}\right) - \text{Loc}\left(d_{t_{\text{cm}}}, t_m\right)\right|. \quad (3)$$

Given a candidate t_{cm} and a medical term t_{m} in LRs, we use the length proportion of their longest common substring and the shorter term as their name similarity, as (2). The similarity computation can capture the nested entities. For example, for a candidate "胸肌内膜炎" (endomysitis), if t_m is "胸肌" (chest muscle), we can regard "胸肌" (chest muscle) as a nested entity. In addition, considering that some terms and their fractional terms exist together in a text, we use the text distance constraint to improve the detector's accuracy. For example, a text contains both "头孢" (cephalosporin) and "头孢拉定" (cefradine), and the compared medical term is "头孢拉定" (cefradine). For the candidate "头孢" (cephalosporin), if the text distance constraint is not used, the output medical entity is "头孢拉定" (cefradine). Obviously, this is the incorrect surface form for the candidate "头孢" (cephalosporin). In (3), the sign $d_{t_{\mathrm{cm}}}$ represents the text containing the candidate entity. Function "Loc" computes the location of the second parameter in the first parameter. Using the above two equations, the specific detection process is as Algorithm 1.

In Algorithm 1, the input includes the candidate entity set, the medical term set from the offline and online LRs, and the input text. The output is a set of medical entities. Given a candidate, we first compute its name similarity with each medical term in MT. If they are the same, the candidate is regarded as a medical entity. If not, we select the medical terms exceeding the predetermined similarity threshold θ for performing the text distance calculation. For each medical term ranked by name similarity, if the text distance between the medical term and the candidate is under the threshold δ, the medical term is output as the correct expression of this candidate. In addition, considering the existing of misspelled ME names, we add the "Diff" function to recognize them. This involves counting the number of different characters between a candidate and the medical term (with the highest name similarity). If the number is less than the threshold ϵ (in our experiments, it is set to the number of half the characters in a medical term in our experiments), we output the medical term as the correct expression of this candidate. For example, for a candidate "头孢拉丁" (cefradine), the compared medical term is "头孢拉定" (cefradine), meeting the above condition. Therefore, we output this medical term instead of the candidate.

3.1.2. Entity Classification. Entity categories are additional information for characterizing the entities mentioned. They are essential ingredients in many medical applications, such as medical dictionaries, medical KBs, and medical service systems. Our classification approach is partly inspired by the use of seed knowledge and context signature similarity in [8]. The difference between our approach and the classification approach in [8] is in the following four ways: (1) In the collection of seed terms, we use the framework information in the terminology instead of the category tags, reducing classification error. Meanwhile, we classify some ME mentions based on text feature computation, thereby avoiding the constraint of dissimilar context and the lack of context. (2) Signature vector computation is refined through word embedding, which can better measure the semantic similarity than the TF-IDF method. (3) The filtering threshold is automatically generated by averaging the signature similarity of seed terms, thereby reducing labor costs and increasing filtering accuracy. (4) The seed set is scaled up continually to improve coverage. The classification is implemented by applying the following three steps: seed term collection, signature generation, and category decision.

Input: candidate set C, medical term set MT, text set T.
Output: medical entity set ME

for $c_i \in C$, $mt_j \in MT$, $t_z \in T$ **do**
  if $c_i == mt_j$ **then**
    $ME \leftarrow c_i$;
  end if
  set $M = \varnothing$;
  if $\mathrm{Sim}(c_i, mt_j) > \theta$ **then**
    $M \leftarrow \mathrm{Rank}(mt_j)$;
  end if
  for $m_k \in M$ **do**
    if t_z contains m_k and $\mathrm{Dis}(c_i, m_k) < \delta$ or $\mathrm{Diff}(c_i, m_k) < \epsilon$ **then**
      $ME \leftarrow m_k$;
      break;
    end if
  end for
end for
return ME;

Algorithm 1: Medical entity detection.

(1) Seed Term Collection. This step involves collecting seed terms for entity categories, based on which the signature vectors of the categories will be generated in the subsequent step. Here, we utilize Baidu Baike to automatically gather the seed terms. In an in-depth analysis, we find that the medical entities of the same class have similar framework information in Baidu Baike, which is more accurate than only using category tag, in identifying the entity category. Therefore, we design a text feature computation-based seed collection approach. Here, we define $T = \{\mathrm{s}, \mathrm{a}, \mathrm{d}, \mathrm{c}\}$ as the set of text features, with a subtitle "s," the attribute names "a" of the infobox, the directory names "d" of the content, and the category tags "c" in the entry page of Baidu Baike. The approach is implemented as follows. (1) From the self-built dictionary, we randomly select 50 terms from each category, to extract and fuse their text features as the category signature. In particular, we exploit the perfect string matching algorithm to produce unambiguous Baidu Baike entries. (2) For each candidate, we also crawl the feature information from Baidu Baike. Then, we calculate its string similarity with all category signatures by using (4) and classifying this candidate to the category with the highest similarity. In particular, the signs W_c and W_{cm} represent the word sets of a category signature and the feature information for a candidate, respectively. Finally, the classified candidate entities are used as the seeds for the category signature computation in the next step.

$$\mathrm{Sim}_c(W_{\mathrm{cm}}, W_c) = \frac{|W_{\mathrm{cm}} \cap W_c|}{|W_{\mathrm{cm}}|}. \quad (4)$$

(2) Signature Generation. This step involves transforming the medical terms (including candidates and seeds) and categories into signature vectors. Here we use the phrase "term signature" to denote the vector of a ME mention or a seed term. Considering that the internal words have descriptive ability for a term, we use the internal and context words for signature generation. To capture the semantic similarity between words, we exploit a word embedding approach to calculate the vector value of a word. Here we use the Word2-Vec model, a distributed representation model, to express the words in text as vectors based on deep learning technology [26]. The training corpus is the input corpus, the description content of all medical terms in Baidu Baike, and the search results of Baidu Search. The final term signature vector is computed by averaging all word vectors, in accordance with (5). In addition, we use the phrase "category signature" to denote the vector of an entity category. This is computed by averaging the signature vectors of all seed terms belonging to the same class, following (5).

(3) Category Decision. Once all term signatures and category signatures are generated, the category of each candidate is identified by using Algorithm 2. The symbol description is shown in Table 3. The similarity calculation between vectors adopts a cosine similarity algorithm, following (6). Though Algorithm 2, each candidate exceeding the filtering threshold is assigned to the category with the highest similarity. In addition, the filtering threshold is automatically computed by averaging the signature similarity of seed terms, following (7). In particular, $|c|$ is the number of seed terms belonging to a class, corresponding to $|t_k|$ in Algorithm 2. $C^2_{|c|}$ is the combination function, counting the number of combinations of any two seeds. Finally, to increase the coverage of the seed set, we add the classified candidate to the relevant seed signature set and then update the filtering threshold and the category signature.

$$v^c = \frac{1}{|S|} \sum_{m \in S} v^m, \quad (5)$$

$$\mathrm{Sim}_{\cos}\left(v^a, v^b\right) = \frac{\sum_{i=1}^{I} \left(v_i^a \times v_i^b\right)}{\sqrt{\sum_{i=1}^{I} \left(v_i^a\right)^2} \times \sqrt{\sum_{i=1}^{I} \left(v_i^b\right)^2}}, \quad (6)$$

$$F\left(v_i, v_j\right) = \frac{1}{C^2_{|c|}} \sum_{i,j=1, i \neq j}^{|c|} \mathrm{Sim}_{\cos}\left(v_i, v_j\right). \quad (7)$$

3.2. Medical Entity Linking. We use the medical KB of Baidu Baike as a basic KB. To increase the accuracy of the similarity calculation, we use the medical KB of Hudong Baike (http://www.baike.com/sitecategory-10.html) to expand the description information of the entities in this basic KB. The method is as follows: for each entity in KB, we acquire its page from Hudong Baike and then extract the description content and category information.

Input: candidate set M, seed signature set T, category signature set C
Output: medical entity-category set E

```
for (m_i, s_{m_i}) ∈ M do
    set F = ∅, D = ∅;
    for t_k ∈ T, s_a, s_b ∈ t_k do
        F ← F(s_a, s_b);
    end for
    for s_{c_j} ∈ C, f_{c_j} ∈ F do
        if Sim_cos(s_{m_i}, s_{c_j}) > f_{c_j} then
            D ← (Sim_cos(s_{m_i}, s_{c_j}), c_j);
        end if
    end for
    if D ≠ ∅ then
        c_t ← arg max(D);
        E ← (m_i, c_t);
        T ← (m_i);
        update s_{c_t} ∈ C by (5);
    end if
end for
return E;
```

Algorithm 2: Medical entity classification.

Table 3: Symbol description in Algorithm 2.

Symbol	Description
M	A set containing each candidate entity m_i and its signature vector s_{m_i}
F	A threshold set filtering the nonmedical entities
t_k	A seed signature set of the same class
s_a, s_b	Seed signature
c_j, c_t	Category name
s_{c_j}, s_{c_t}	Category signature of c_j or c_t
f_{c_j}	Filtering threshold of c_j

In accordance with the procedure of entity linking [27], the ME linking module has two stages: candidate entity generation and ranking. For each ME mention, the module first obtains its candidate entities from the KB, and then selects the top candidate (after ranking) as the linking entity. The mentions without linking entities are regarded as NIL.

3.2.1. Candidate Entity Generation. In this stage, our goal is to increase the probability of the candidate set containing a target entity and to control its size. To accomplish the first goal, we use the fuzzy string matching algorithm to compute the name similarity between a mention and all entities in the KB, in accordance with (8). The function "MCC" acquires the most common characters between two strings in order. It can well process the abbreviations and acronyms besides the standard names. The entities exceeding the similarity threshold α are included in the candidate set. However, this algorithm may result in a large candidate set.

To reduce the computational cost in the subsequent processing, we introduce the condition of category consistency to control the size. The specific method is as follows: for each candidate, we acquire its text features in Baidu Baike and then compute the similarity between the category signatures acquired in the section of seed term collection, following (4). The candidates under a predefined threshold β are removed from the candidate set. This strategy can still well process the terms that have the same name but different meanings. For example, for a ME mention "传染病" (epidemic), its candidate set includes "传染病 (疾病)" (epidemic [disease]), "传染病 (游戏)" (contagion [game]), and "传染病 (电影)" (contagion [film]). Through category constraint, the latter two candidates are removed.

$$\mathrm{Sim}_l(t_{\mathrm{me}}, t_e) = \frac{\mathrm{MCC}(t_{\mathrm{me}}, t_e)}{\min(\mathrm{Len}(t_{\mathrm{me}}), \mathrm{Len}(t_e))}. \tag{8}$$

3.2.2. Candidate Entity Ranking. This stage aims to acquire the linking entity in the candidate set by ranking using a confidence score. We propose a collaborative inference method which synthetically exploits the name similarity, entity popularity, context similarity, and the semantic correlation between entities.

Specifically, the name similarity of the mention and its candidates is computed using (8). In addition, based on the common knowledge that the most important entity is the most frequently mentioned, we introduce the entity popularity for distinguishing and discriminating between the candidate entities. Here, we utilize the number of visits in the Baidu Baike page to indicate the entity popularity, which is a positive integer (e.g., 15,348). Considering that the entity popularity is not the only decisive criterion, we establish a conversion to ensure its effectiveness and to avoid impacting other measuring conditions. Given the visiting number n, the entity popularity is computed as

$$p(n) = \frac{\left(|n| \times 10^{(|n|)} + n\right)}{10^{(|n|+1)}}, \tag{9}$$

in which $|n|$ expresses the digit number. For instance, the above integer is translated into 0.515348.

The existing context similarity-based approaches generally extract the words in a fixed window, which ignores the noise information in the context. To increase the description ability of the context words of a mention, we explore a relevant information extraction approach based on the dependency relationships between words. Specifically, this extracts all words that have a dependency relationship with a mention as the context information. For example, in Figure 2, the relevant information of "骨髓纤维化" (myelofibrosis) is "髓纤" (MF) and "骨髓增生性疾病" (myeloproliferative disease). Then, we compute its string similarity with the description content of each candidate by using (4). In particular, the signs W_{cm} and W_c represent the context word sets of a mention and a candidate. Of note, before similarity computation, we need to remove the stop words in the context and the description content.

However, the context information acquired by the above extraction approach is limited. It may result in the same context similarity between different candidates. Moreover, some mentions may have no context information. For the mentions, we add the semantic correlation knowledge for ranking based on the hypothesis that the linking entities of the cooccurring entities in text are also correlated, and they have overlapping context information. The special method is as follows: (1) In the context of a mention, we select some ME mentions (with the linking entities) as the collaborators. (2) We extract the anchors and other noun phrases (which are more descriptive than other words) from the description content of these linking entities and the candidates of the mention, respectively. (3) The context similarity between each candidate and all linking entities is computed, and the candidate with the highest similarity is regarded as the target entity.

In conclusion, the confidence score of the candidate entities can be computed by using (10). λ is a control factor (the value is 1 or 0), controlling whether the semantic correlation is computed. If the context similarity of each candidate is 0 or the same, $\lambda = 1$. If not, $\lambda = 0$. Given a mention t_{me} and a candidate t_{ce}, the linking entity set of the collaborators L, the confidence score is computed using

$$\begin{aligned}\mathrm{CS}(t_{\mathrm{me}}, t_{\mathrm{ce}}, L) = {} & \mathrm{Sim}_l(t_{\mathrm{me}}, t_{\mathrm{ce}}) + P(t_{\mathrm{ce}}) + \mathrm{Sim}_c(I(t_{\mathrm{me}}), D(t_{\mathrm{ce}})) \\ & + \lambda \mathrm{Sim}_c\left(A(t_{\mathrm{ce}}), \sum_{t_{le_k} \in L} A\left(t_{le_k}\right)\right).\end{aligned} \tag{10}$$

The signs "P," "I," "D," and "A" express the entity popularity, the relevant context information, the description content, and the special content containing only anchors and noun phrases in the KB, respectively.

In order to better understand the ranking process, we provide an example. Given the text "NS,…,功能紊乱体现在失眠, 多梦, 盗汗…" (NS,⋯, the dysfunction is reflected in insomnia, dreaminess, and night sweats⋯), the recognized ME mentions are "NS," "失眠" (insomnia), "多梦" (dreaminess), and "盗汗" (night sweats). Through the previous process, we find that "NS" has multiple candidate entities with the same name, such as "NS (nervous system)," "NS (nephrotic syndrome)," and "NS (normal saline)." Their name similarity is 1, and their other measuring scores are as in Figure 3. It must be noted that we only present the partial value of the entity popularity for the purpose of saving space. According to the confidence scores computed using (10), the candidate "NS (nervous system)" is selected as the linking entity with the highest score (0.7577).

4. Experiments

4.1. Experimental Data. We crawl 5000 medical Q&A text records from three Chinese medical websites to evaluate our proposed framework, including "家庭医生在线" (Family-doctor), "拇指医生" (Muzhi-doctor), and "求医网" (Qiuyi). Next, we randomly select 500 records from each corpus to recognize all medical entities, classify them to the six categories in Table 3, and link them to the KB manually. In total, we recognize 6596 ME mentions and link 3821

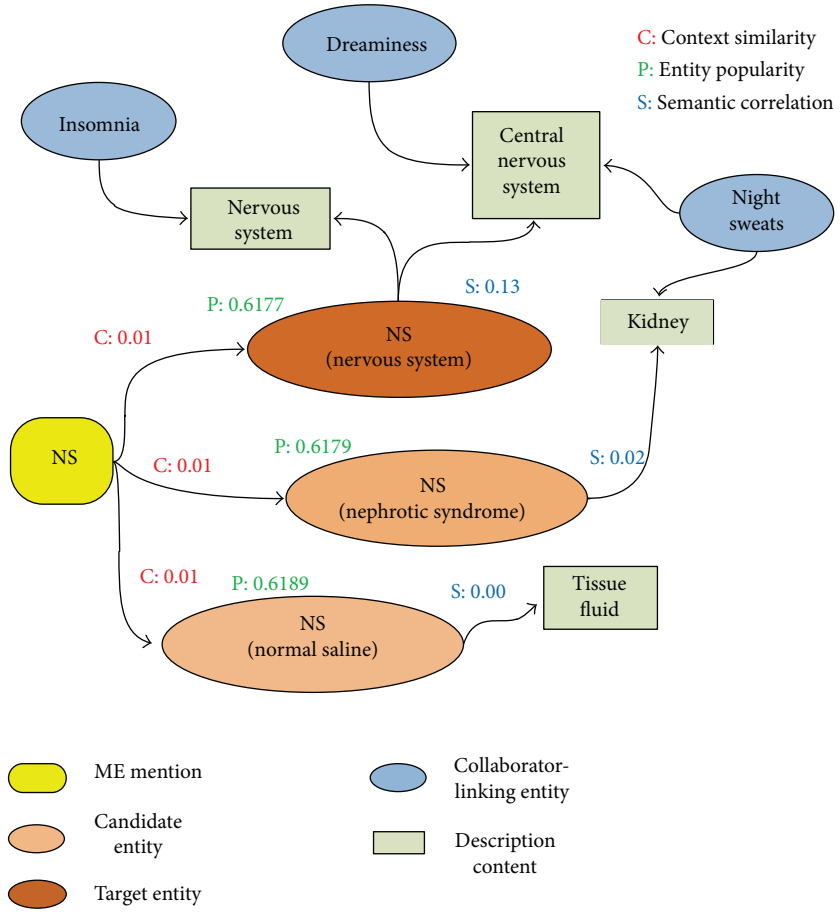


Figure 3: Example of linking the ME mention "NS."

mentions to the correct entries in the KB, whose statistics are shown in Table 4. The sign "NIL" expresses the ME mentions without the linking entities in the KB.

4.2. Experimental Evaluation

4.2.1. Comparative Methods. To thoroughly validate the effectiveness of unMERL, we conduct a comparison between the representative state-of-the-art methods and our proposed methods in the recognition and linking modules, respectively. For ME recognition, we select BM-NER [8] and bubble-bootstrapping [10], which are unsupervised methods, as well as Dic-CRF (to distinguish the method, we have given it this name, as it is a supervised method) [18] as the comparative methods. In particular, for the BM-NER method, we use the Stanford parser (https://nlp.stanford.edu/software/lex-parser.html) for chunking. The seed terms are taken from our built dictionary. For the Dic-CRF method, we split 500 records into two subsets: two-thirds for training and one-third for testing. In ME linking, we select QCV (a language independent and unsupervised method) [13] as a comparative method. In addition, it is necessary to state that we use the same seeds in [10] for the bubble-bootstrapping method, the same features in [18] and our built dictionary for the Dic-CRF method, as well as the anchors in the KB to build a KB graph for the QCV method.

Table 4: Statistics of the corpus.

Corpus	MEs	MEs linking to KB	NIL
Family-doctor	2531	1524	1007
Muzhi-doctor	1876	109	780
Qiuyi	2189	1201	988

4.2.2. Measuring Methods. We use P (precision), R (recall), and F_1 to measure performance. P is the fraction of the correct objects in all objects acquired by the method. R is the fraction of the correct objects acquired by the method, in the valid objects in the corpus. F_1 is defined as $2 \times P \times R/(P + R)$. In addition, we still use "accuracy" to measure the whole linking accuracy, as shown in (11). In particular, $|S_{\text{link}}|$ and $|S_{\text{NIL}}|$ express the number of ME mentions that are linked or not linked to the correct entities in the KB by the method. $|T|$ represents the number of ME mentions in the corpus.

$$\text{Accuracy} = \frac{|S_{\text{link}}| + |S_{\text{NIL}}|}{|T|} \times 100\%. \tag{11}$$

4.3. Experimental Results and Discussion. To simulate ME recognition and linking tasks in an open environment (note:

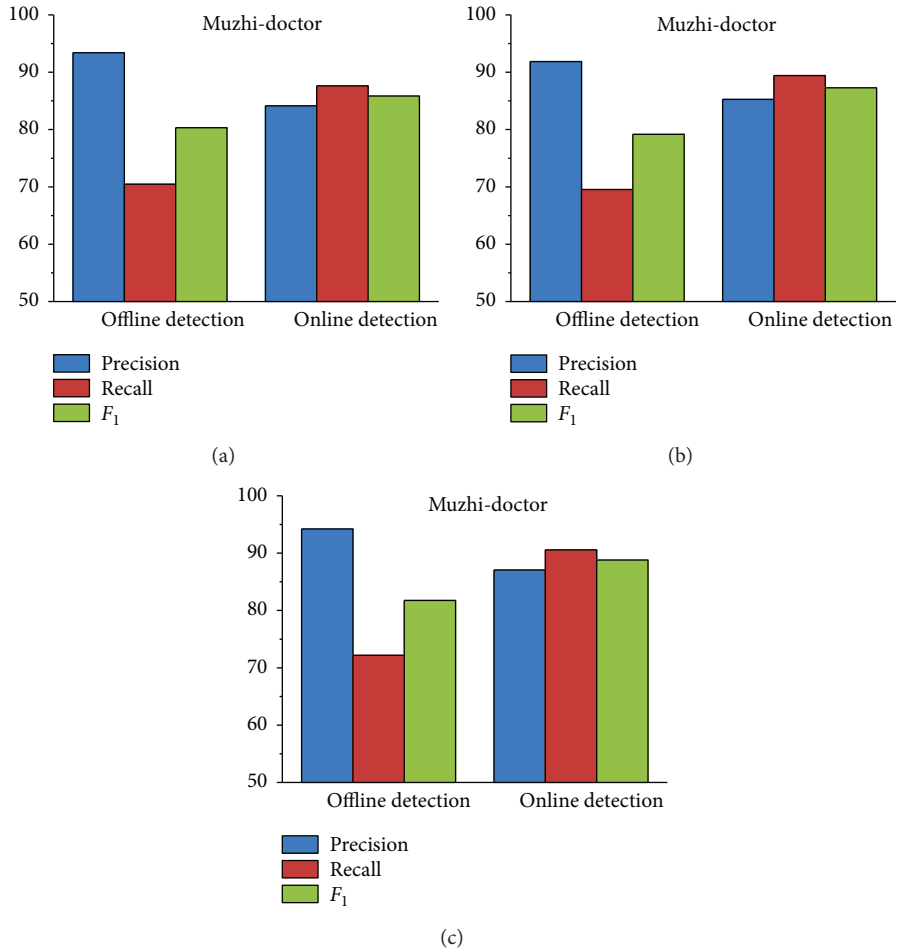


Figure 4: Experimental results after offline detection and online detection on the corpus (%).

the experimental data has low coverage for real data), we randomly select 30 records for learning all the above-mentioned threshold values. In the ME recognition module, the threshold θ for name similarity between a candidate and a medical term in LRs is experimentally set as 0.5. The threshold δ for the text distance constraint is experimentally set to 3. This means that if the text distance between a medical term and a candidate is lower than 3, the medical term is output instead of the candidate. In the ME linking module, the thresholds α and β are experimentally set as 0.5 and 0.47, respectively. In particular, the threshold α is used to compute the name similarity between a mention and an entity in the KB, and the threshold β is used for category consistency.

4.3.1. Medical Entity Recognition. As mentioned above, our ME recognition module is divided into two stages: boundary detection and entity classification. In order to evaluate the effectiveness of our proposed methods fully, we show the experimental results of each stage in detail.

(1) Boundary Detection. To validate the effectiveness of online detection, Figure 4 presents the experimental results after offline detection and online detection for all datasets. Recall has a noticeable improvement after the online detection process. It is therefore proven that online detection is efficient in solving the limitation problem of the dictionary-based method. However, the precision has some limitations. The main limitation is that some irrelevant terms in the candidate set are not filtered by the online detection process. Therefore, in the entity classification stage, we add the filtering threshold to remove these terms.

(2) Entity Classification. To evaluate the entity classification method on its own, we conduct an experiment with the standard entity boundaries for all medical entities in the corpus. Assuming that all medical entities have been extracted correctly from text and that our task is to classify them into the predefined categories, Table 5 presents the classification results of each corpus. The overall performance is significant at an 81.85% precision level and a 75.84% recall level. The lower recall is because when filtering the nonmedical entities, some medical entities are removed by the filtering threshold, thereby reducing the recall. The performance of the target categories "symptom," "treatment," and "check" is somewhat low. One possible reason for this is that these entities are mostly classified based on the context signature similarity. However, the lack of the identifying information in the context reduces the similarity score, thus impacting the classification performance.

Table 5: Entity classification results on the corpus (%).

Entity category	Family-doctor			Muzhi-doctor			Qiuyi		
	P	R	F_1	P	R	F_1	P	R	F_1
All	82.49	79.13	80.78	80.91	73.86	77.22	82.16	74.35	78.06
Body	85.17	81.20	83.14	83.62	80.13	81.84	83.53	80.21	81.84
Disease	80.45	82.16	81.30	81.96	82.67	82.31	79.26	81.68	80.45
Symptom	78.19	60.84	68.43	74.54	61.73	67.53	76.26	61.52	68.10
Medicine	82.31	79.62	80.94	80.63	75.84	78.16	84.57	78.39	81.36
Treatment	76.86	67.25	71.73	75.24	63.59	68.93	76.61	61.82	68.42
Check	75.14	65.53	70.01	74.50	67.26	70.70	73.48	63.72	68.25

(3) Overall Recognition Performance. We compare the overall performance of our recognition approach (named "unMER") with the unsupervised and supervised methods described above in Figures 5–7.

Figure 5 shows the experimental results of unMER compared with the bubble-bootstrapping approach. This is because we only acquired the seeds of the symptom category for bubble-bootstrapping. The results show that unMER significantly outperforms bubble-bootstrapping in terms of recall. However, unMER's precision is slightly low. One possible reason for this is that the symptomatic entity mentions are diverse, resulting in low coverage in the offline LRs. Therefore, they are mainly recognized by the online detection method. However, the combined mentions produce diverse search results, from which it is difficult to get a complete term. For example, for the mention "手脚无力" (powerless hands and feet), the returned results contain "手脚无力" (powerless hands and feet), "四肢无力" (powerless limbs), and "手脚发软" (limp hands and feet). After online detection, the acquired entity is "手脚" (hands and feet) or "无力" (powerless). In addition, the low recall of the bubble-bootstrapping approach is because the online Q&A text lacks normalization in its description, reducing the performance of pattern matching.

Figure 6 shows the experimental results of unMER compared with the BM-NER approach. Obviously, unMER outperforms BM-NER in both precision and recall. The value of F_1 of unMER increases 26.12%, 27.52%, and 25.78% on three corpora. The reasons are as follows: (1) The BM-NER approach uses a noun phrase chunker to extract candidate entities, which does not consider the nested entities, thereby reducing the recall. In addition, the chunker utilizes a common NLP tool, which had poor recognition performance for the medical entity boundary. (2) The IDF filter removes many common medical entities. (3) We exploit a distributed word embedding approach to acquire the word vector, which well considers the semantic similarity between words than the TF-IDF algorithm of BM-NER. (4) Our built dictionary contains many incorrect seed categories, and this resulted in semantic deviation for the BM-NER approach, reducing the classification performance.

Figure 7 shows the experimental results of unMER compared with Dic-CRF on each corpus. Note that for the body category, we do not have the features of Dic-CRF and hence do not present its measuring result. On three corpora, the F_1 value of unMER increases 15.01%, 13.68%, and 12.68% than Dic-CRF approach, respectively. By analyzing the experiments, we find that the high recall of unMER mainly depends on the online detection process, which demonstrates the validity of using a search engine for recognizing medical entities. However, Dic-CRF uses a medical dictionary for word segmentation, this can easily lead to incorrect segmentation, especially for the combined entities. In addition, the defined features have low coverage in all entity types, which is also a reason for the low recall. Moreover, the informal description of the online medical text also reduces the recognition performance of the CRF model. In terms of precision, unMER yields comparative results and even exceeds Dic-CRF in some categories. This is due to our combination of multiple offline LRs, thereby increasing the coverage of medical entities. Moreover, unMER has good recognition performance in the nested entities.

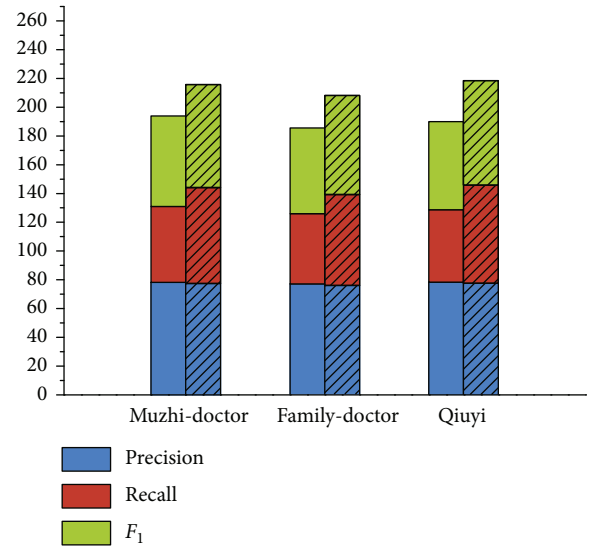


Figure 5: Experimental results of unMER versus bubble-bootstrapping on the symptom category only (note: the cylinder with bias represents unMER, and the other cylinder represents bubble-bootstrapping).

4.3.2. Medical Entity Linking. Figure 8 shows the linking results of our approach (named "unMEL") compared with the QCV approach on each corpus. To evaluate the linking

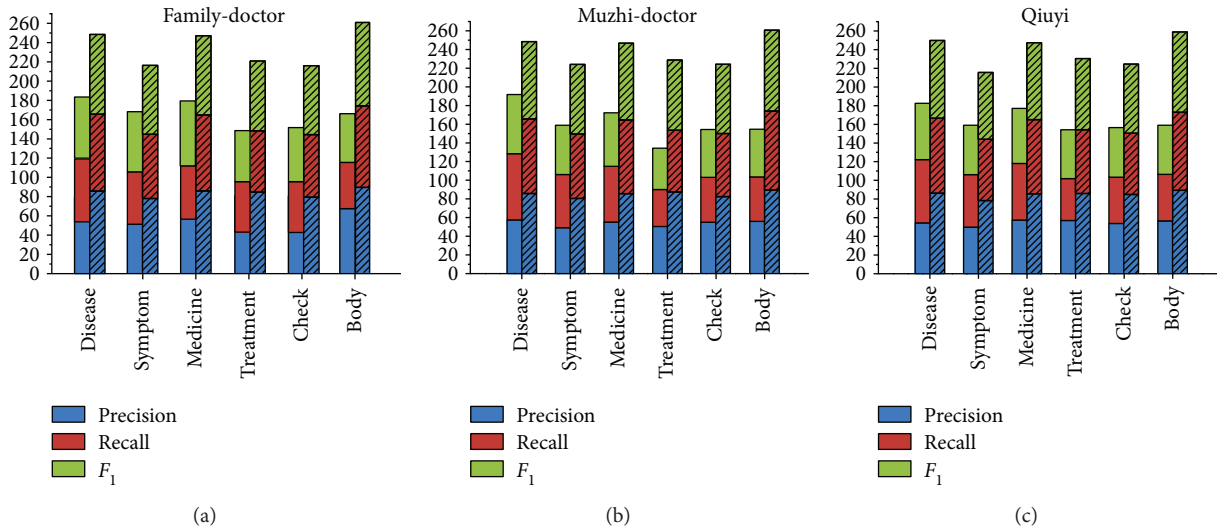


Figure 6: Experimental results of unMER versus BM-NER on the corpus (note: the cylinder with bias represents unMER, and the other cylinder represents BM-NER).

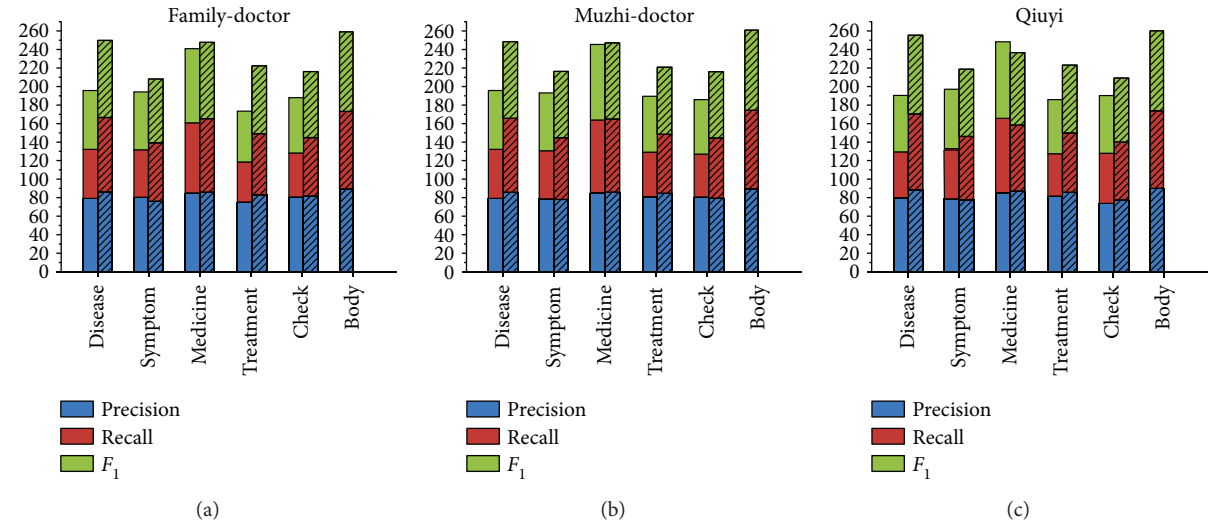


Figure 7: Experimental results of unMER versus Dic-CRF on the corpus (note: the cylinder with bias represents unMER, and the other cylinder represents Dic-CRF).

approach on its own, we conduct an experiment with the standard entity boundaries for all medical entities in the corpus. Assume that all entities have been extracted correctly from text, and our task is only to link them to the correct entities in the KB. Compared to the QCV approach, the F_1 value of unMEL increases 6.39%, 6.67%, and 5.81%, and the accuracy value increases 6.03%, 4.6%, and 5.54% on each corpus, respectively. This is possibly due to the similar relationship in the KB between the mentions within the specific window, QCV virtually uses the context similarity for linking. Therefore, the noise and lack of information in the context reduce the linking performance. However, unMEL alleviates the restriction by extracting the relevant context information and using semantic correlation. Moreover, in the recognition module, we modify the misspelled ME mentions, which help link to the correct entities. Nevertheless, unMEL utilizes the fuzzy string matching to generate candidate entities, which omits some target entities that are fully different in the surface form, reducing the linking recall.

4.3.3. Overall System Performance. To evaluate the overall performance of our framework (unMERL), Table 6 shows the linking results by conducting an experiment with our recognized entities. Compared to the above linking results, both the precision and recall show some decline. The reason is that unMERL obtains some inexact entities in the boundary detection step. In addition, unMERL removes some medical entities when filtering the nonmedical terms in the classification step.

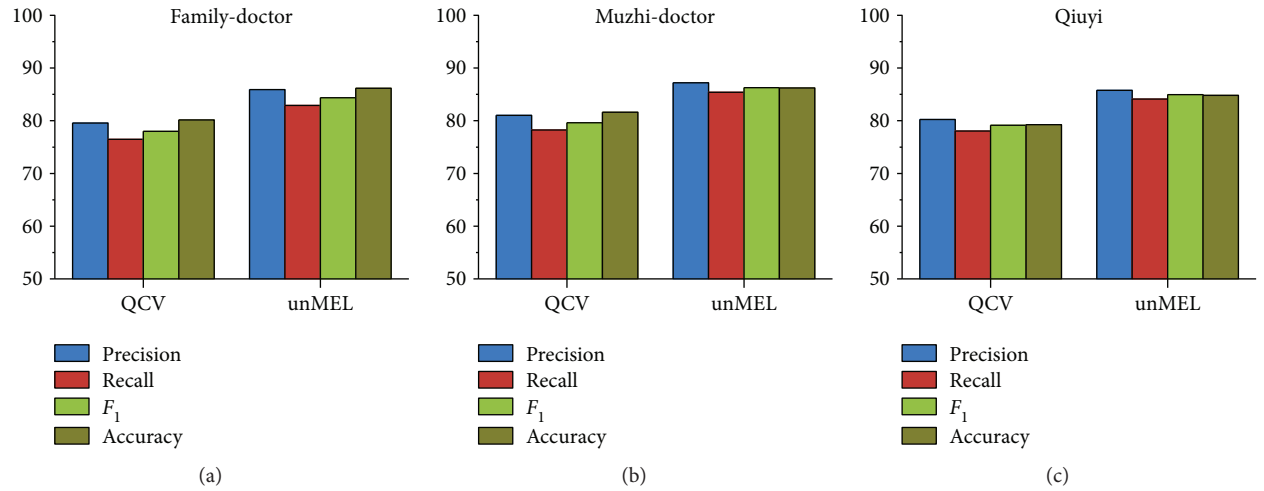


Figure 8: Experimental results of unMEL versus QCV on the corpus (note: the cylinder with bias represents unMEL, and the other cylinder represents QCV).

Table 6: Experimental results of unMERL on the corpus (%).

Corpus	unMERL			
	P	R	F_1	A
Family-doctor	82.64	73.26	77.67	83.23
Muzhi-doctor	83.37	74.41	78.64	83.48
Qiuyi	82.15	72.79	77.19	82.05

5. Conclusions

Medical entity recognition and linking are challenging tasks in Chinese natural language processing. In this paper, we have described an unsupervised framework for recognizing and linking medical entities from Chinese online medical text, namely, unMERL. To the best of our knowledge, this is the first complete unsupervised solution for Chinese medical text with both medical entity recognition and linking. It has considerable value in many applications, such as medical KB construction and expansion, semantic comprehension of medical text, and medical Q&A systems. Experimental evidences show that unMERL consistently outperforms current approaches. In addition, due to its unsupervised nature and language independence, unMERL has good generalizability.

In the future, we will improve unMERL in the following ways. Firstly, we will improve the online detection approach, by adding in-depth textual analysis in extracting medical terms from the search results. Secondly, we will improve the linking approach by introducing semantic analysis.

Disclosure

The authors alone are responsible for the content and writing of the paper.

Conflicts of Interest

The authors report no conflicts of interest.

Acknowledgments

This work is supported in part by the National Basic Research and Development Program (2016YFB0800303), the National Key Fundamental Research and Development Program of China (2016QY03D0601, 2016QY03D0603), the National Natural Science Foundation of China (61502517).

References

[1] C. Friedman, P. O. Alderson, J. H. M. Austin, J. J. Cimino, and S. B. Johnson, "A general natural-language text processor for clinical radiology," *Journal of the American Medical Informatics Association*, vol. 1, no. 2, pp. 161–174, 1994.

[2] T. C. Rindflesch, L. Tanabe, J. N. Weinstein, and L. Hunter, "Edgar: extraction of drugs, genes and relations from the biomedical literature," in *Proceedings of the Pacific Symposium*, pp. 517–528, Honolulu, Hawaii, USA, 2000.

[3] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program," *Proceedings Amia Symposium*, vol. 2001, no. 1, p. 17, 2001.

[4] S. Kraus, C. Blake, and S. L. West, "Information extraction from medical notes," *Opening Schools for All*, vol. 13, pp. 95–103, 2007.

[5] B. L. Humphreys, D. A. B. Lindberg, H. M. Schoolman, and G. Octo Barnett, "The unified medical language system: an informatics research collaboration," *Journal of the American Medical Informatics Association*, vol. 32, no. 4, p. 281, 1993.

[6] C. J. Mcdonald, J. M. Overhage, W. M. Tierney et al., "The Regenstrief medical record system: a quarter century experience," *International Journal of Medical Informatics*, vol. 54, no. 3, pp. 225–253, 1999.

[7] K. Donnelly, "SNOMED-CT: the advanced terminology and coding system for eHealth," *Studies in Health Technology & Informatics*, vol. 121, no. 121, p. 279, 2006.

[8] S. Zhang and N. Elhadad, "Unsupervised biomedical named entity recognition: experiments with clinical and biological texts," *Journal of Biomedical Informatics*, vol. 46, no. 6, pp. 1088–1098, 2013.

[9] D. Movshovitz-Attias and W. W. Cohen, "Bootstrapping biomedical ontologies for scientific text using NELL," in *BioNLP '12 Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, pp. 11–19, Montreal, Canada, 2012.

[10] L. Z. Feng, *Automatic Approaches to Develop Large-scale TCM Electronic Medical Record Corpus for Named Entity Recognition Tasks*, Beijing Jiaotong University, 2015.

[11] E. Riloff and R. Jones, "Learning dictionaries for information extraction by multi-level bootstrapping," in *AAAI '99/IAAI '99 Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, pp. 474–479, Menlo Park, CA, USA, 1999.

[12] G. D. Zhou, J. Zhang, J. Su, D. Shen, and C. L. Tan, "Recognizing names in biomedical texts: a machine learning approach," *Bioinformatics*, vol. 20, no. 7, pp. 1178–1190, 2004.

[13] Y. Wang, Z. Yu, L. Chen et al., "Supervised methods for symptom name recognition in free-text clinical records of traditional Chinese medicine: an empirical study," *Journal of Biomedical Informatics*, vol. 47, no. 2, pp. 91–104, 2014.

[14] Y. F. Lin, T. H. Tsai, W. C. Chou, K.-P. Wu, T.-Y. Sung, and W.-L. Hsu, "A maximum entropy approach to biomedical named entity recognition," in *BIOKDD'04 Proceedings of the 4th International Conference on Data Mining in Bioinformatics*, pp. 56–61, London, UK, 2004.

[15] Y. Wang and J. Patrick, "Cascading classifiers for named entity recognition in clinical notes," in *WBIE '09 Proceedings of the Workshop on Biomedical Information Extraction*, Borovets, Bulgaria, 2009.

[16] B. Settles, "Biomedical named intity recognition using conditional random fields and rich feature sets," in *JNLPBA '04 Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications*, pp. 104–107, Geneva, Switzerland, 2004.

[17] J. Liang, X. Xian, X. He et al., "A novel approach towards medical entity recognition in Chinese clinical text," *Journal of Healthcare Engineering*, vol. 2017, Article ID 4898963, 16 pages, 2017.

[18] Y. Su, J. Liu, and Y. Huang, "Entity recognition research in online medical texts," *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 52, no. 1, pp. 1–9, 2016.

[19] J. Lei, B. Tang, X. Lu, K. Gao, M. Jiang, and H. Xu, "A comprehensive study of named entity recognition in Chinese clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 808–814, 2014.

[20] C. Y. Qu, *Research of Named Entity Recognition for Chinese Electronic Medical Records*, Harbin Institute of Techonology, 2015.

[21] G. Glava, "TAKELAB: medical information extraction and linking with MINERAL," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 389–393, Denver, CO, USA, 2015.

[22] J. G. Zheng, D. Howsmon, B. Zhang et al., "Entity linking for biomedical literature," *BMC Medical Informatics and Decision Making*, vol. 15, no. S1, 2015.

[23] H. Wang, J. G. Zheng, X. Ma, P. Fox, and H. Ji, "Language and domain independent entity linking with quantified collective validation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 695–704, Lisbon, Portugal, 2015.

[24] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.

[25] B. L. Li, E. S. Li, and Y. H. Wei, "Function design, implementation and applications of Chinese SNOMED 3.4," in *National Conference on Medical Informatics*, 1999.

[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, https://arxiv.org/abs/1301.3781.

[27] W. Shen, J. Wang, and J. Han, "Entity linking with a knowledge base: issues, techniques, and solutions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 443–460, 2015.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 4581272, 10 pages
https://doi.org/10.1155/2018/4581272

Research Article

ANN and Fuzzy Logic Based Model to Evaluate Huntington Disease Symptoms

Andrius Lauraitis[1], **Rytis Maskeliūnas**[2] **and Robertas Damaševičius**[3]

[1]*Department of Multimedia Engineering, Kaunas University of Technology, Studentų 50, Kaunas, Lithuania*
[2]*Center of Real Time Computing Systems, Kaunas University of Technology, K. Baršausko 59, Kaunas, Lithuania*
[3]*Department of Software Engineering, Kaunas University of Technology, Studentų 50, Kaunas, Lithuania*

Correspondence should be addressed to Andrius Lauraitis; andrius.lauraitis@ktu.lt

Received 28 September 2017; Accepted 24 January 2018; Published 11 March 2018

Academic Editor: Terry K.K. Koo

We introduce an approach to predict deterioration of reaction state for people having neurological movement disorders such as hand tremors and nonvoluntary movements. These involuntary motor features are closely related to the symptoms occurring in patients suffering from Huntington's disease (HD). We propose a hybrid (neurofuzzy) model that combines an artificial neural network (ANN) to predict the functional capacity level (FCL) of a person and a fuzzy logic system (FLS) to determine a stage of reaction. We analyzed our own dataset of 3032 records collected from 20 test subjects (both healthy and HD patients) using smart phones or tablets by asking a patient to locate circular objects on the device's screen. We describe the preparation and labelling of data for the neural network, selection of training algorithms, modelling of the fuzzy logic controller, and construction and implementation of the hybrid model. The feed-forward backpropagation (FFBP) neural network achieved the regression R value of 0.98 and mean squared error (MSE) values of 0.08, while the FLS provides a final evaluation of subject's reaction condition in terms of FCL.

1. Introduction

Huntington disease (HD) is a progressive genetic neurodegenerative disorder causing involuntary movement and cognitive problems that significantly affect daily life of HD patients. HD affects about 1 in 10,000 to 20,000 people of European (Caucasian) descent [1], though in some isolated populations it is much higher. HD reduces life expectancy due to heart disease, pneumonia, physical injury from falls, and suicide. The most visible symptom of HD is chorea, which consists of jerky, involuntary movements of the upper and lower extremities, face or body, and occurs in about 90% of patients at some stage of their illness [2]. Other symptoms include behavioural problems, cognitive impairment, psychiatric disorders, and dementia, which have a serious impact on daily living of a patient and often result in hospitalization. The societal and financial cost of HD on health and social care systems is significant and is estimated to be £195 million per year in the UK alone [3].

HD is currently incurable so most of the current research in this area focuses on identifying the deficits at the early stage of the disease, to benefit from future medical interventions that may help delaying the progress of the disease [4, 5]. This is also the case of the work presented in this paper. Traditional HD research often include magnetic resonance neuroimaging (MRI) measures of striatum and white matter volume, CAG repeat length in chromosome analysis, age, and striatal atrophy [6, 7]. Moreover, medical personnel and doctors who have experience in caring after HD patients and knowing that disease is cureless are not usually motivated to conduct scientific research themselves or to support multidiscipline (e.g., bioinformatics) investigations.

Any scientific result (device, technology, and theoretical model) that could contribute towards improvement of daily life of HD patient's and help to monitor or predict the progress of the disease can be useful for both doctors and HD patients.

Table 1: Comparison of various ML methods adapted for neurodegenerative disorders such as Huntington or Parkinson disease to solve prediction and classification problems.

Work ref.	ML method	Learning approach	ML problem	Size of data set	Number of test subjects	Target group	Involve HD patients
[8]	ANN, MLP	Supervised	Classification	—	21	PD, healthy	No
[9]	RBFNN	Supervised	Regression	—	—	PD	No
[10]	DNN	Supervised	Classification	—	12	PD (8), healthy (4)	No
[11]	Decision tree, ID3	Supervised	Classification	195	31	PD (23), healthy (8)	No
[15]	Adaptive neurofuzzy	Hybrid		100	—	PD	No
[16]	Neurofuzzy system	Hybrid		—	—	ALS	No
[17]	Fusion of classifiers (Bayesian, SVM, k-nearest neighbor)	Hybrid		640	ALS (13), PD (15), HD (16), healthy (16)	ALS, PD, HD	Yes
[18]	Neurofuzzy system	Hybrid		—	—	Only survey was done	No
[19]	ANN + MLP, RBFNN	Hybrid		—	—	PD	No
[20]	Neurofuzzy system	Hybrid		—	—	—	No
[21]	PBL-McRBFN	Supervised	Classification	22,283	72 (50 PD, 22 healthy)	PD, healthy	No
[22]	Multistate Markov model	Hybrid		2500	72 (82 PD, 62 healthy)	PD, healthy	No
[23]	Random tree, (C-*RT*), ID3, binary logistic regression, k-NN, (PLS), (SVM)	Supervised	Classification	195	31 (23 PD, 8 healthy)	PD, healthy	No
[24]	FCM	Unsupervised	Clustering	195	—	PD	No

The problematics of data prediction evolved with the rise of artificial intelligence (AI) and machine learning (ML) methods and algorithms. Artificial neural networks (ANN) such as multilayer perceptron (MLP) can be used for classification of accelerometer-based tremor signals invoked by Parkinson patient's involuntary movements [8]. Prediction of Parkinson disease onset by adapting radial basis function neural network (RBFNN) for tremor activity data recorded via stimulation electrodes using electromyography (EMG) signals is described in [9]. Dynamic neural network (DNN) is used to detect time-varying occurrences of tremor and dyskinesia from time series data acquired from EMG sensors and triaxial accelerometers worn by Parkinson patients [10]. Another approach of designing a prediction model for Parkinson's disease uses a decision tree and Iterative Dichotomiser (ID3) methods to analyze data collected from HD symptoms such as trembling in the legs, arms, hands, impaired speech articulation, and production difficulties [11]. Hybrid models combine different AI and ML approaches for reproducing intelligent human reasoning process [12]. By using information fusion, hybrid models combine heterogeneous ML approaches and improve quality of reasoning for complex regression and classification problems [13]. Neurofuzzy systems combine neural network and fuzzy logic paradigms to avoid the limitations of neural network explanations to reach decision and limitations of fuzzy logic to automatically acquire the rules used for making those decisions [14]. Fuzzy expert systems such as neurofuzzy system (ANFIS) can be applied in assessment of Parkinson's disease with a noninvasive screening system for quantitative evaluation and analysis by using amplitude, frequency, spectral characteristics, and trembling localization parameters of input data [15]. Hybrid model is adapted in designing a decision support system (DSS) for the intelligent identification of Alzheimer where neurofuzzy system explores approximation techniques from neural networks to find the parameter of a fuzzy system [16]. Hybrid systems are also used as a classifier fusion strategy (Bayesian, SVM, k-nearest neighbours) in the prevalence of age-related diseases like Alzheimer's and dementia [17], diagnostics and measurement [18] with wavelet transform (WT) and norm entropy feature extraction methods. The DSS that uses MLP and RBFNN is applied for monitoring patients with neurological disorders [19]. The data is collected using noninvasive smart devices (modified mouse and 3-axis accelerometer sensor). Integration of neurofuzzy networks and information fusion for multimodal human cognitive state recognition is described in [20]. Projection-based learning for metacognitive radial basis function network (PBL-McRBFN) is applied to predict Parkinson's disease [21]. Other hybrid systems and applications include nonlinear adaptive system, which fuses brain and gait information algorithmically using multistate Markov model [22]. Accurate Parkinson disease diagnosis model based on cluster analysis uses random tree, classification and regression tree (C-*RT*), ID3, binary logistic regression, k-NN, partial least square regression (PLS), support vector machines (SVM) [23], and fuzzy c-means clustering (FCM) [24]. Table 1 provides a summary of methods used by other authors.

Our previous work included the development of text input-based system for evaluating the condition of Huntington's patients [25]. The use of ANN for predicting the functional capacity of a Huntington's patient was proposed in [26].

The aim of this paper is to create a computerized behavioural model, which predicts an impaired reaction condition for HD patients. We develop a mobile application to collect a dataset using finger touch coordinates and reaction time features extracted from test subjects (healthy and HD patients); create an ANN to predict the functional capacity level and fuzzy logic system (FLS) to determine the reaction condition (stage) for individual person; combine ANN with FLS into a hybrid model to predict the impaired reaction condition for HD patients; and simulate an experimental setup for test subjects to perform a provided exercise (test) at the different moments in time in order to predict a possibly impairing reaction condition with the help of the proposed hybrid model.

2. Materials and Methods

2.1. Subjects. The study included ten (10) Huntington disease (HD) patients living in Lithuania. Each HD patient agreed to participate and allowed the data collected during the test to be used for scientific purposes. Every HD patient fall in the early clinical descriptor category of Huntington disease, that is, I and II stages according to Shoulson–Fahn evaluation system [27]. Such HD patients have hand tremors, body movement distractions, but are capable to perform the test on a mobile application without extra help, for example, from medical personnel, nurses, or family members. Other ten (10) participants were healthy people with no signs of any neurological or neurodegenerative disorder.

2.2. Procedure. The test can be performed using various mobile devices that support Android OS. The mobile application randomly generates circular shape objects (2, 3, and 5 circles at time) of particular color that are generated on the mobile device's screen. Each circle is located in different positions of the screen, thus no possible collisions (overlapping) between two particular circles are possible. An active circle that needs to be touched is marked by a black contour so as to differ from other objects.

The subjects are instructed to touch every object, starting from first in sequence, by finger as close to center and as quickly as possible. When subject finishes the test, collected data is stored in external mobile device storage and sent to the database using the internet connection.

2.3. Dataset. The collected dataset consists of 3032 data examples collected from 20 test subjects (10—healthy and 10—HD patients). The dataset (see a sample in Table 2) contains the ground truth coordinates of the generated object, the coordinates of subject's touch, subject's reaction time, subject's label, and the marker of Huntington's disease.

2.4. Feature Extraction and Class Labelling. The subject's reaction time (*rt*) and the Euclidian distance between the two points of true and touched positions (*delta*) serve as features which are incorporated as input variables to ANN. We assume that smaller *rt* and *delta* values indicate better functional capacity level. The bigger *delta* value can show stronger hand tremoring, whereas higher *rt* value is an indicator of body stagnancy.

Table 2: Collected data from mobile application (random sample data of 5 records).

x	*y*	*xt*	*yt*	nC	*rt*	*delta*	User	IsSick
126	871	125	872	5	1.665	1.414	1	0
411	403	390	408	3	3.886	21.587	1	0
243	609	299	592	3	0.573	58.523	2	1
580	377	618	449	5	0.545	81.413	2	1
501	634	437	585	2	0.741	56.436	3	1

x,y: screen coordinates of the center of circular object to touch; *xt*, *yt*: screen coordinates of user touch; nC: number of circular objects rendered on the device screen; *rt*: user's reaction time in seconds; *delta*: the Euclidean distance between object's center and touch position. User: user ID; IsSick: indicates if test subject has Huntington disease (1 yes, 0 otherwise).

The statistical analysis of the *rt* and *delta* values has revealed that the values are not normally distributed, but after the applying the log transformation, which is commonly used in regression analysis of biological data with highly skewed distribution [28], the values become normal as confirmed by visual inspection in Figure 1 and skewness γ and kurtosis κ tests ($\gamma_{rt} = 1.046, \kappa_{rt} = 4.239$ and $\gamma_{delta} = 0.028, \kappa_{delta} = 4.779$). For data samples greater than 300, values $|\gamma| < 2$ and $|\kappa| < 7$ are considered as acceptable for normality [29].

To analyze the power of *rt* and *delta* values to correctly predict the healthy or sick state of the subject, we have performed feature evaluation using the relative entropy (also known as the Kullback–Leibler distance or divergence) criterion, considering different number of objects presented at the screen. The results are presented in Figure 2. In all cases, *delta* feature has larger discriminative power than *rt*, and the features from 3 and 5 objects test are more statistically discriminative.

2.5. ANN for Functional Capacity Level Prediction. We have analyzed the following neural network models: (1) feed-forward backpropagation (FFBP); (2) feed-forward time delay neural network (FFTD); (3) cascade-forward backpropagation (CFBP); (4) nonlinear autoregressive exogenous model (NARX); (5) Elman neural network; (6) layer recurrent neural network (RNN); and (7) generalized regression neural network (GRNN).

FFBP is a simple neural network without any cycle connections between neurons [30]. FFTD has no internal state and adds delayed copies as other inputs as an input signal to obtain time-shift invariance [31]. In CFBP, the input values calculated after every hidden layer are backpropagated and the weights adjusted [32]. NARX have a limited feedback, which comes only from the output neuron rather than from hidden layer [33]. Elman network additionally has context units, which are connected to the hidden units, thus providing the network with memory [34]. RNN represent an architecture where connections between

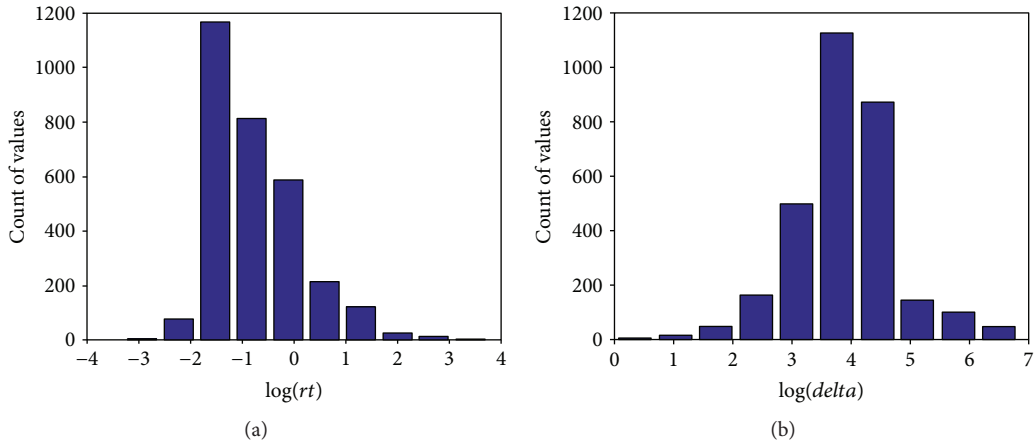


Figure 1: Statistical distribution of log-transformed *rt* (a) and *delta* (b) values.

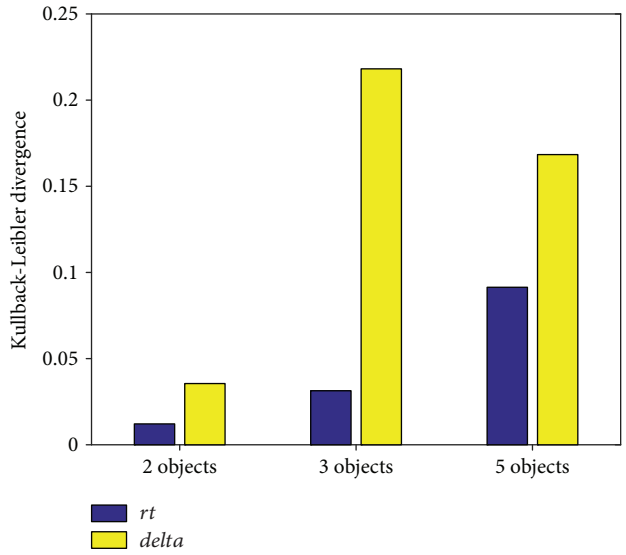


Figure 2: Feature ranking according to the Kullback–Leibler distance.

units form a directed cycle [35]. GRNN has only one (smoothness) parameter, and its convergence is guaranteed; fast and stable [36].

Each neural network has 2 inputs (*rt*, *delta*) and 1 output (Y). Neural network is composed of single neurons that are treated as a simple unit carrying signals (data) to each other or different layers via transfer functions, which correspond to sum of input signal. Training function is the optimization algorithm used for finding global minimum of a function. The outputs of ANN are class labels for determining the functional capacity of a person (the larger value indicates that a person is more capable to do motoric activities). Such scenario imitates the TFC scale measurement system for Huntington disease patients presented in Table 3 [27].

Table 4 illustrates the setup for analyzed ANN models with their parameters.

2.6. Training and Testing. The dataset was randomly divided into 3 sets: training, validation, and testing. Training set uses all samples from 70% of users. Validation set (15%) is used to measure network generalization and to stop training when necessary. Testing set (15%) provides independent performance of the network afterwards. We also analyzed a different partition of the dataset (40% for training, 30% for validation, and 30% for testing); however, there were no significant differences in the performance of ANN.

Table 3: Total functional capacity score (TFC) and its relationship to Shoulson–Fahn stages and clinical descriptors [27].

Descriptor	TFC	Stage
Early	11–13	I
	7–10	II
Moderate or mid	4–6	III
Advanced or late	1–3	IV
	0	V

Overfitting was prevented by using the early stopping technique, which controls error on the validation set which is monitored during training process: when error increases for a specified number of iterations then the training is stopped and the weights and biases at the minimum of the validation error are returned.

For each neural network model, we have repeated the training and testing process for 20 times in order to allow calculation of statistical characteristics (mean, standard deviation) of ANN performance measures and to perform statistical comparison.

2.7. Reaction Stage Determination Using Fuzzy Logic System (FLS). The aim of the FLS system is to determine the reaction stage of a patient (test subject) according to some predefined parameters. The FLS consists of three main parts: fuzzification block, inference mechanism, and defuzzification block. Membership functions, linguistic variables are created in fuzzification module. Inference engine is responsible for applying logical rules (fuzzy rule base) to the knowledge base and deduce new knowledge. Defuzzification module converts all the fuzzy terms created by the rule base of the controller to crisp terms (numerical values). The FLS uses triangular membership Mamdani-type functions with fuzzy set inference mechanism (minimum implication,

Table 4: Summary of different neural network models and their configuration parameters.

Network	Hidden layer (neurons)	Transfer function	Training function	Number of weight elements	Time delay
FFBP	1 (10)	Log-sigmoid, linear	Gradient descent with adaptive learning rate backpropagation	41	–
FFTD	1 (10)	Tan-sigmoid	Levenberg–Marquardt	101	+
CFBP	1 (10)	Tan-sigmoid	Levenberg–Marquardt	43	–
NARX	1 (10)	Tan-sigmoid	Levenberg–Marquardt	81	+
Elman, RNN	1 (10)	Tan-sigmoid	Levenberg–Marquardt	141	+
GRNN	1 (size of dataset)	Radial basis, linear	Levenberg–Marquardt	800	–

Table 5: FLS rule base (5 random examples chosen for each reaction stage).

AVG1	AVG2	AVG3	Reaction stage
HIGH	HIGH	HIGH	HEALTHY/PRECLINICAL
HIGH	LOW	HIGH	EARLY
AVERAGE	AVERAGE	AVERAGE	AVERAGE
LOW	LOW	HIGH	LATE
LOW	LOW	LOW	ADVANCED

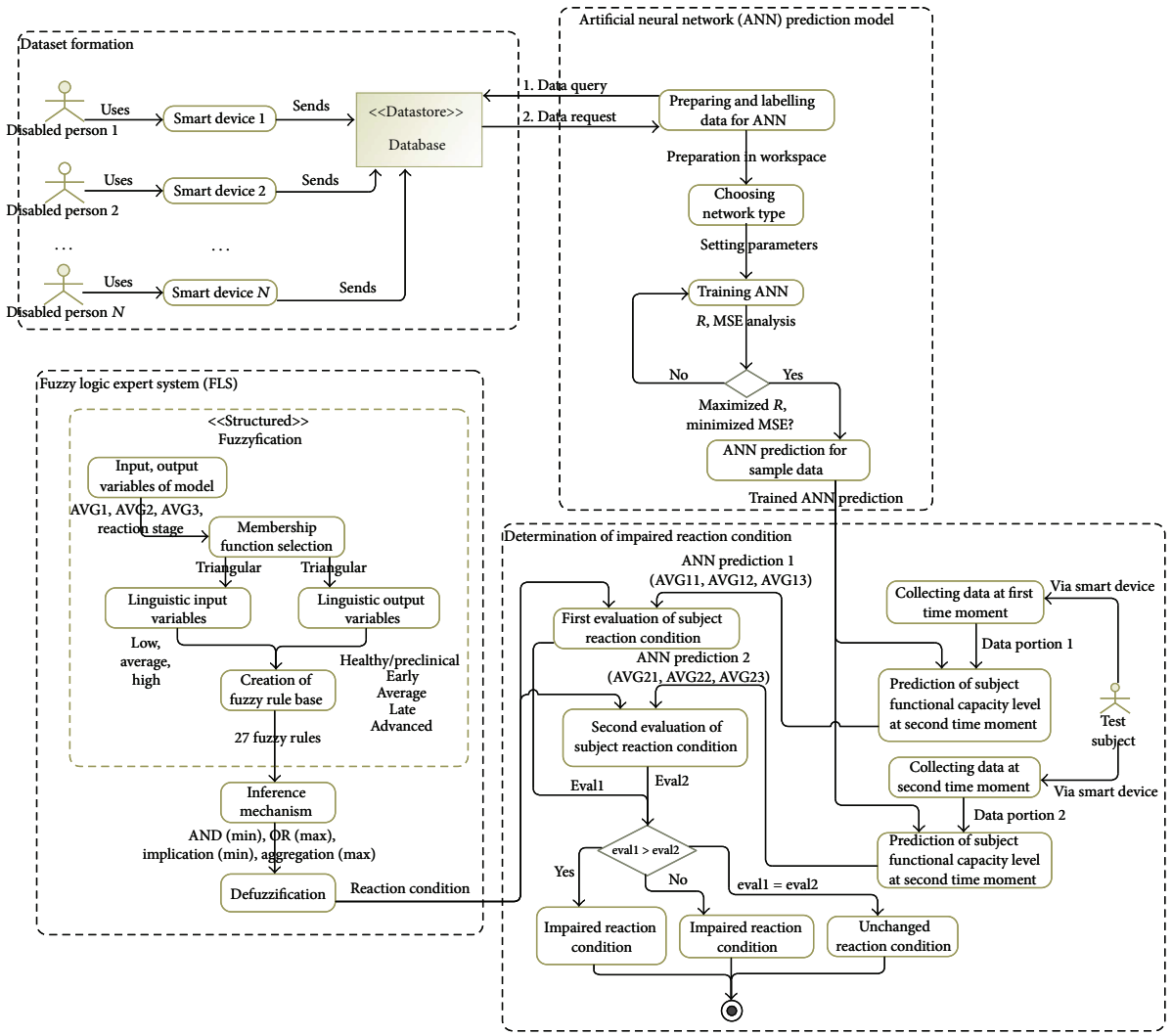


Figure 3: Schema of prototype hybrid model to forecast impaired reaction condition.

Table 6: Regression and prediction result comparison of different ANN models.

Functional capacity level predictions (data sample of 10 records from 1 test subject)										
rt	2.65	0.25	0.67	0.26	0.24	0.78	0.4	0.29	0.34	0.25
delta	22.36	108.3	87.8	267	60.1	20.8	37	113	41.4	68.8
TFC	8	3	4	1	5	8	7	3	7	5
FFBP	8.02	3.31	3.90	1.32	5.39	8.55	7.25	3.19	6.85	4.83
FFTD	7.99	3.26	3.94	0.97	5.41	8.51	7.24	3.15	6.83	4.87
CFBP	7.94	3.32	3.59	1.28	5.45	8.84	7.15	3.16	6.83	4.87
NARX	8.06	3.28	3.83	1.20	5.43	8.81	7.24	3.16	6.85	4.87
Elman	8.01	3.27	3.87	1.19	5.40	8.79	7.30	3.16	6.88	4.84
RNN	7.97	3.28	3.86	1.15	5.42	8.54	7.27	3.15	6.87	4.87
GRNN	8.20	2.92	3.86	0.99	5.18	8.93	6.83	2.91	6.88	4.87

Table 7: Performance comparison of analyzed ANN models.

Neural network model	Mean *R*	95% confidence intervals of *R*	Mean MSE	95% confidence intervals of MSE
FFBP	**0.9876**	**[0.9871, 0.9880]**	**0.0809**	**[0.0782, 0.0835]**
FFTD	0.9861	[0.9855, 0.9867]	0.0906	[0.0865, 0.0948]
CFBP	0.9827	[0.9787, 0.9868]	0.1125	[0.0860, 0.1389]
NARX	0.9868	[0.9868, 0.9869]	0.0858	[0.0855, 0.0860]
Elman	0.9868	[0.9868, 0.9868]	0.0857	[0.0856, 0.0857]
RNN	0.9870	[0.9870, 0.9870]	0.0845	[0.0845, 0.0845]
GRNN	0.9849	[0.9841, 0.9858]	0.0977	[0.0926, 0.1029]

maximum aggregation, minimum AND operator, maximum OR operator) and centroid defuzzification method.

The parameters of the FLS are derived from the ANN output corresponding to the functional capacity level, so in the FLS design process, the model input and output values need to be considered accordingly. There are three input and one output variable in the FLS. The input parameters are *AVG1*, *AVG2*, and *AVG3*, which correspond to the average of ANN output values when test subject is working with two, three, and five objects, respectively. All three inputs can have values in range [0; 10].

The linguistic variables (terms) for *AVG1*, *AVG2*, and *AVG3* are

(1) *LOW* [0 2 4];

(2) *AVERAGE* [3.6 5.5 7];

(3) *HIGH* [6.6; 8.5 10].

The model has one output parameter *ReactionStage* can have one of five values: close to peaks 1, 3, 5, 7, or 9, that is, each peak corresponds to particular linguistic variable of *ReactionStage*. The terms for output parameter *ReactionStage* are

(1) *ADVANCED* [0 1 2];

(2) *LATE* [1.5 3 4];

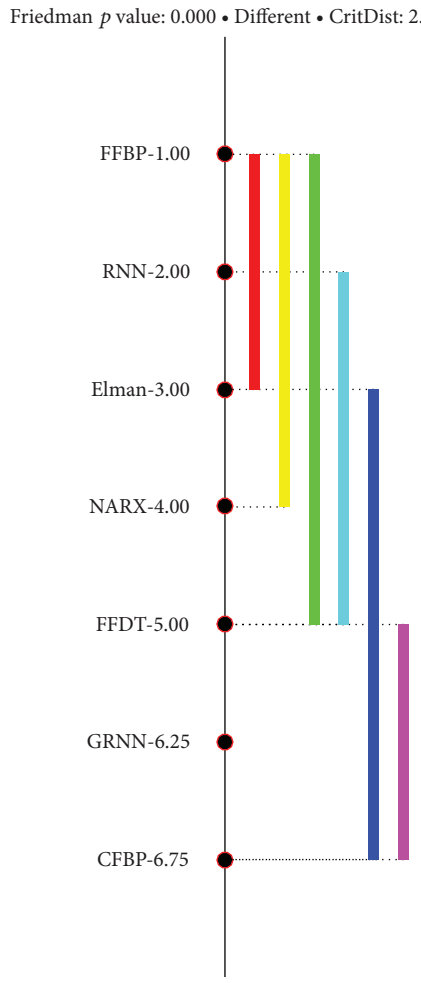


Figure 4: Results of Nemenyi test on performance (MSE) of ANN models.

(3) *AVERAGE* [3.5 5 6];

(4) *EARLY* [5.5 7 8];

(5) *HEALTHY/PRECLINICAL* [7.5 9 10].

The FLS rule base is formed from 27 fuzzy rules. Table 5 illustrates the principles of constructing fuzzy rule base.

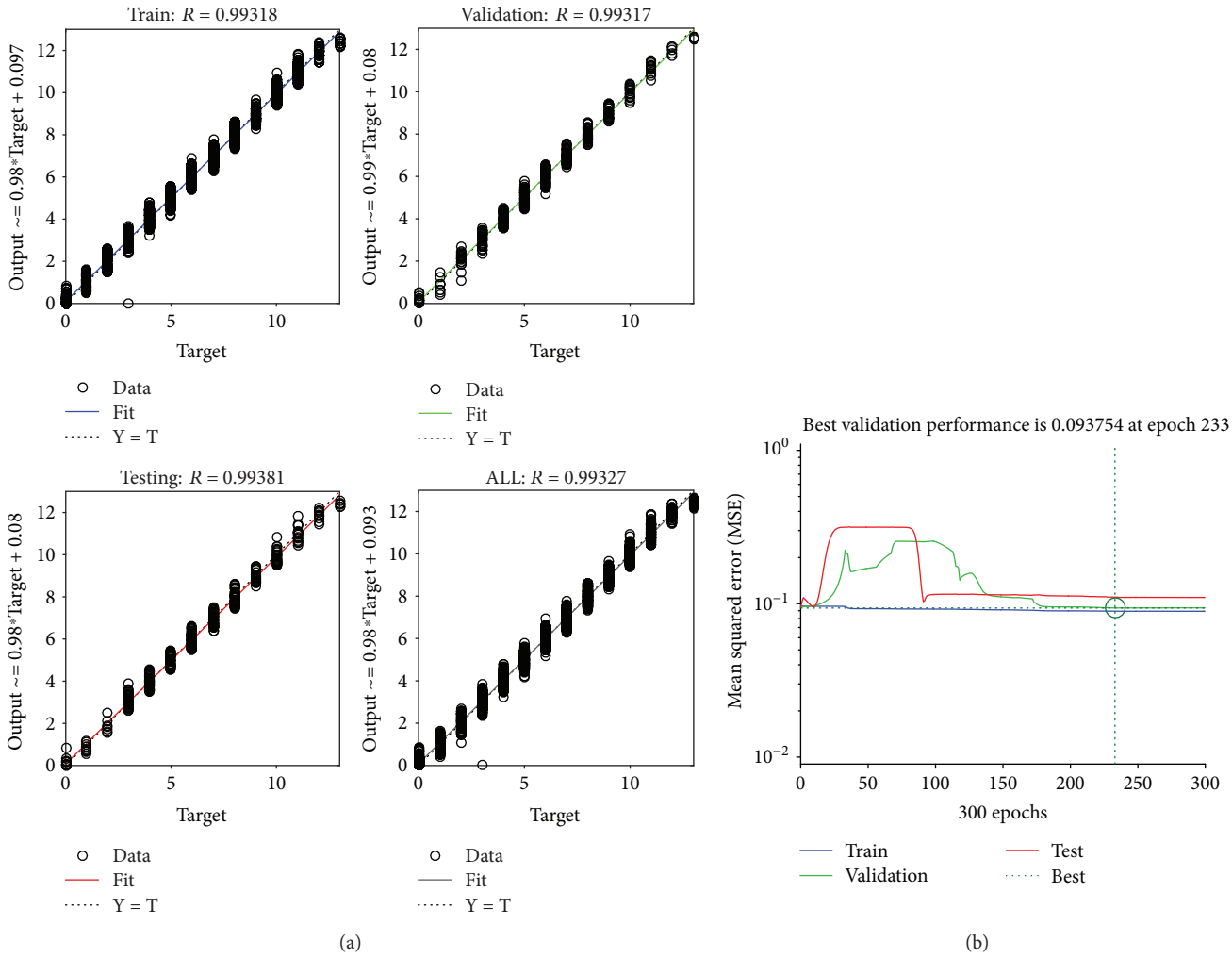


Figure 5: FFBP performance evaluation using R (a) and MSE (b) metrics.

These can be interpreted as general fuzzy IF-THEN rules containing only fuzzy logical AND operators, for example,

IF *AVG1* is LOW AND *AVG2* is LOW AND *AVG3* is LOW

THEN *ReactionStage* is ADVANCED.

3. Proposed Hybrid Model

The hybrid model (see Figure 3) is composed of four sub models: (1) dataset formation; (2) ANN prediction model; (3) fuzzy logic expert system (FLS); and (4) decision module for determination of person's condition.

During dataset formation, test subjects (under the supervision of a healthy person—a medical doctor or a nurse) use smart devices to perform reaction and accuracy test experiments with their fingers. The collected data is stored in the database. The ANN submodel predicts the functional capacity level of a person using the data from the database. The network is trained by observing regression (R), that is, correlation measurement between outputs and targets and mean squared error (MSE) values. Once the network is trained, it can make predictions on new sample data. Finally, to evaluate the reaction condition of a test subject, the test session is repeated at a different time and the ANN predictions are aggregated, and the reaction stage of a person is evaluated using a fuzzy rules system.

4. Experimental Results

The hybrid model was implemented with MATLAB Neural Network and Fuzzy Logic Toolbox software (MathWorks Inc.). The results of regression and comparison of the prediction results of the analyzed ANN models is presented in Table 6, whereas the performance of neural networks in terms of means and 95% confidence intervals of R and MSE is given in Table 7. The "TFC" field indicates the ground truth evaluation of the patient state provided by a medical neurologist expert according to the TFC scale. The R metric measures the correlation between output and targets, whereas the MSE metric is the average squared difference between outputs and targets.

Nonparametric Friedman test was conducted to compare the performance results (MSE) among ANN models. Results show that there is a significant difference in performance among all ANN models (chi-square = 133.15; $p = 2 \cdot 10^{-26}$).

Table 8: Example of hybrid model results for evaluation of a single test subject.

(a)

Feature	Mode 1 2 objects		Mode 2 3 objects				Mode 3 5 objects			
Dataset formation: data portion 1										
rt1	4.75	8.05	5.30	2.27	7.09	1.48	6.58	6.33	2.29	1.82
delta1	3.32	2.99	4.05	19.09	0.31	19.15	0.51	19.42	5.95	10.50
Artificial neural network (ANN) prediction model: ANN prediction 1										
	8.94	8.00	8.43	6.68	8.99	6.79	9.00	5.46	9.03	8.10
Dataset formation: data portion 2										
rt2	8.62	8.96	1.89	6.60	9.41	1.48	6.58	6.33	2.29	1.82
delta2	1.42	9.78	16.99	19.94	0.89	10.85	17.27	18.18	16.90	17.57
Artificial neural network (ANN) prediction model: ANN prediction 2										
	8.01	7.94	7.00	5.10	8.01	7.26	7.01	6.93	5.99	7.04

(b)

Fuzzy logic expert system (FLS)									
AVG11	AVG12	AVG13	AVG21	AVG22	AVG23	eval1	eval2	Condition 1	Condition 2
8.47	8.03	7.67	7.98	6.70	6.84	9.00	7.00	Healthy/Preclinical	Early
Conclusion: Impaired reaction condition.									

Posthoc Nemenyi tests further reveal that the performance of FFBP is the best among all ANN models (Figure 4).

Figure 5 shows an example of FFBP best performance equal to $R = 0.993$ and MSE = 0.094 on the validation set.

Table 8 illustrates impaired reaction condition simulation example on a single test subject using the FLS system. In order to make comparison, data samples were collected at different time moments. Feature (*rt1*, *delta1*, *rt2*, and *delta2*) values are presented in all three modes (10 attempts), thus giving two separate ANN (in the example provided, FFBP model was used) prediction outputs, which are used to calculate average values and evaluate the reaction condition in the FLS.

5. Conclusions

We have presented an actual experimental framework to assess finger-tapping tests performed by patients suffering from the Huntington's disease (HD). The proposed model was validated using a dataset of 3032 data records collected from 20 test subjects (both healthy and HD patients). The reaction condition was determined using the developed Mamdani Type-1 fuzzy logic expert system (FLS) with 3 input (3 linguistic variables), 1 output (5 linguistic variables), triangular membership functions, and 27 fuzzy rules base.

We describe an architecture that combines several artificial neural networks (ANN) of different type (FFBP, FFDT, CFBP, NARX, Elman, RNN, and GRNN) to create a hybrid (neurofuzzy) model, which integrates feature extraction, prediction, and classification routines to forecast the impaired reaction condition for HD patients. The best results were achieved using the feed-forward backpropagation (FFBP) neural network model, which predicts the total functionality capability (TFC score) with high performance results, that is, it has obtained regression R value not less than 0.98 and mean squared error (MSE) values of 0.08, while FLS evaluates several measurements taken time apart to provide a final evaluation of the subject's reaction condition.

Future work will focus on the validation of the proposed system using a larger dataset, which includes the data collected from the Parkinson's and Alzheimer's patients as well, the analysis and use of more sophisticated finger-tapping features, and the comparison of the ANN results with those of SVM regression.

Additional Points

Human Studies. Research on human subjects was approved by the Institutional Review Board of the Faculty of Informatics of Kaunas University of Technology.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

Authors would like to thank the president of Lithuania Huntington disease association, Dr. Zivile Navikiene, for contacting family members of HD patients to help carry out experiments for the investigation described in this paper, as well as for practical support and advice.

References

[1] R. A. C. Roos, "Huntington's disease: a clinical review," *Orphanet Journal of Rare Diseases*, vol. 5, no. 1, p. 40, 2010.

[2] M. S. Haddad and J. L. Cummings, "Huntington's disease," *The Psychiatric Clinics of North America*, vol. 20, no. 4, pp. 791–807, 1997.

[3] C. Jones, M. Busse, L. Quinn et al., "The societal cost of Huntington's disease: are we underestimating the burden?," *European Journal of Neurology*, vol. 23, no. 10, pp. 1588–1590, 2016.

[4] M. Bennasar, Y. Hicks, S. Clinch et al., "Huntington's disease assessment using tri axis accelerometers," *Procedia Computer Science*, vol. 96, pp. 1193–1201, 2016.

[5] T. V. Wiecki, C. A. Antoniades, A. Stevenson et al., "A computational cognitive biomarker for early-stage Huntington's disease," *PLoS One*, vol. 11, no. 2, article e0148409, 2016.

[6] J. S. Paulsen, D. R. Langbehn, J. C. Stout et al., "Detection of Huntington's disease decades before diagnosis: the Predict-HD study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 8, pp. 874–880, 2008.

[7] E. H. Aylward, B. F. Sparks, K. M. Field et al., "Onset and rate of striatal atrophy in preclinical Huntington disease," *Neurology*, vol. 63, no. 1, pp. 66–72, 2004.

[8] M. Engin, S. Demirag, E. Engin et al., "The classification of human tremor signals using artificial neural network," *Expert Systems with Applications*, vol. 33, no. 3, pp. 754–761, 2007.

[9] D. Wu, K. Warwick, Z. Ma et al., "Prediction of Parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization," *International Journal of Neural Systems*, vol. 20, no. 02, pp. 109–116, 2010.

[10] B. T. Cole, S. H. Roy, C. J. De Luca, and S. H. Nawab, "Dynamic neural network detection of tremor and dyskinesia from wearable sensor data," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, vol. 2010, pp. 6062–6065, Buenos Aires, Argentina, 2010.

[11] A. Chandrashekhar, S. Jain, and V. Kumar Jha, "Design and analysis of data mining based prediction model for Parkinson's disease," *International Journal of Computer Science Engineering (IJCSE)*, vol. 3, no. 03, 2014.

[12] T. Furuhashi, S. Tano, and H.-A. Jacobsen, *Deep Fusion of Computational and Symbolic Processing*, vol. 59, Publish in Physica, Fuzziness and Soft Computing, New York, NY, USA, 2012.

[13] P. Kazienko, E. Lughofer, and B. Trawinski, "Hybrid and ensemble methods in machine learning," *Published in Journal of Universal Computer Science*, vol. 19, no. 4, pp. 457–461, 2013.

[14] R. Fuller, "Introduction to neuro-fuzzy systems," in *Advances in Intelligent and Soft Computing*, vol. 2, Springer Science & Business Media (2013), 2013.

[15] O. Geman, "Parkinson's disease assessment using fuzzy expert system and nonlinear dynamics," *Advances in Electrical and Computer Engineering*, vol. 13, no. 1, pp. 41–46, 2013.

[16] J. C. Obi and A. A. Imainvan, "Decision support system for the intelligent identification of Alzheimer using neuro fuzzy logic," *International Journal on Soft Computing (IJSC)*, vol. 2, no. 2, pp. 25–38, 2011.

[17] S. Iram, P. Fergus, D. Al–Jumeily, A. Hussain, and M. Randles, "A classifier fusion strategy to improve the early detection of neurodegenerative diseases," *International Journal of Artificial Intelligence and Soft Computing (IJAISC)*, vol. 5, no. 1, pp. 23–44, 2015.

[18] Z. J. Viharos and K. B. Kis, "Survey on neuro-fuzzy systems and their applications in technical diagnostics and measurement," *Measurement*, vol. 67, pp. 126–136, 2015.

[19] I. Chiuchisan and O. Geman, "An approach of a decision support and home monitoring system for patients with neurological disorders using internet of things concepts," *WSEAS Transactions on Systems*, vol. 13, pp. 460–469, 2014.

[20] G. Yang, Y. Lin, and P. Bhattacharya, "Multimodality inferring of human cognitive states based on integration of neuro-fuzzy network and information fusion techniques," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, article 371621, 2007.

[21] G. Sateesh Babu and S. Suresh, "Parkinson's disease prediction using gene expression – a projection based learning meta-cognitive neural classifier approach," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1519–1529, 2013.

[22] O. Geman, "Towards an inclusive Parkinson's screening system," in *2014 18th International Conference on System Theory, Control and Computing (ICSTCC)*, pp. 470–475, Sinaia, Romania, 2014.

[23] G. R. Ramani and G. Sivagami, "Parkinson disease classification using data mining algorithms," *International Journal of Computer Applications*, vol. 32, no. 9, pp. 17–22, 2011.

[24] I. Rustempasic and M. Can, "Diagnosis of Parkinson's disease using fuzzy c-means clustering and pattern recognition," *Southeast Europe Journal of Soft Computing*, vol. 2, pp. 42–49, 2013.

[25] J. Gelsvartas, R. Simutis, and R. Maskeliunas, "User adaptive text predictor for mentally disabled Huntington's patients," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3054258, 6 pages, 2016.

[26] A. Lauraitis and R. Maskeliūnas, "Investigation of predicting functional capacity level for Huntington disease patients," in *Information and Software Technologies. ICIST 2017. Communications in Computer and Information Science*, R. Damaševičius and V. Mikašytė, Eds., vol. 756, pp. 142–149, Springer, Cham, 2017.

[27] I. Shoulson and S. Fahn, "Huntington disease: clinical care and evaluation," *Neurology*, vol. 29, no. 1, pp. 1–3, 1979.

[28] S. Gustavsson, B. Fagerberg, G. Sallsten, and E. M. Andersson, "Regression models for log-normal data: comparing different methods for quantifying the association between abdominal adiposity and biomarkers of inflammation and insulin resistance," *International Journal of Environmental Research and Public Health*, vol. 11, no. 12, pp. 3521–3539, 2014.

[29] H.-Y. Kim, "Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis," *Restorative Dentistry & Endodontics*, vol. 38, no. 1, pp. 52–54, 2013.

[30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[31] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, no. 3, pp. 328–339, 1989.

[32] S. Goyal and G. K. Goyal, "Cascade and feedforward backpropagation artificial neural network models for prediction of sensory quality of instant coffee flavored sterilized drink," *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition*, vol. 2, no. 6, pp. 78–82, 2011.

[33] S. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*, Wiley, Chichester, UK, 2013.

[34] J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," *Machine Learning*, vol. 7, no. 2-3, pp. 195–225, 1991.

[35] C. Goller and A. Küchler, "Learning task-dependent distributed representations by backpropagation through structure," in *Neural Networks, 1996, IEEE International Conference on*, vol. 1, pp. 347–352, Washington, DC, USA, 1996.

[36] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 6183468, 8 pages
https://doi.org/10.1155/2018/6183468

Research Article

The Evolving Role of Information Technology in Haemovigilance Systems

Augusto Ramoa,[1,2] **Jorge Condeço**,[1,2] **Maria Antónia Escoval**,[1] **Jean-Claude Faber**,[3] **Florentino Fdez-Riverola**,[2,4] **and Anália Lourenço**[2,4,5]

[1]*Instituto Português do Sangue e da Transplantação, IP, Rua de Bolama 133, 4200-139 Porto, Portugal*
[2]*Department of Computer Science, ESEI, University of Vigo, Campus Universitario As Lagoas, s/n, 32004 Ourense, Spain*
[3]*LuxConsulTrans, Banglamung, Thailand*
[4]*Centro de Investigaciones Biomédicas (CINBIO), University of Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain*
[5]*Centre of Biological Engineering (CEB), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*

Correspondence should be addressed to Anália Lourenço; analia@uvigo.es

Received 29 September 2017; Accepted 9 January 2018; Published 8 March 2018

Academic Editor: Reima Suomi

This work provides an overview and appraisal of the general evolution of IS/IT in haemovigilance, from which lessons can be learned for its future strategic management. An electronic survey was conducted among the members of the International Haemovigilance Network to compile information on the mechanisms implemented to gather, process, validate, and store these data, to monitor haemovigilance activity, and to produce analytical reports. Survey responses were analysed by means of descriptive statistics, and comments/observations were considered in the final discussion. The answers received from 23 haemovigilance organizations show a direct relationship between the number of collected notifications (i.e., communication of adverse effects and events) and the technical specifications of the haemovigilance system in use. Notably, IT is used in the notification reception of 17 of these systems, out of which 8 systems are exclusively based on Web solutions. Most assessments of the evolution of IS/IT tend to focus on the scalability and flexibility of data gathering and reporting, considering the ever-changing requirements of haemovigilance. Data validation is poorly implemented, and data reporting has not reached its full potential. Web-based solutions are seen as the most intuitive and flexible for a system-user interaction.

1. Introduction

The transfusion of blood and blood components is a critical procedure to consider when delivering healthcare services to patients. Haemovigilance aims to improve the safety of the blood supply by monitoring the entire value chain and contravening possible threats to the safety of transfusion recipients [1–3], with the recommendation of adequate corrective actions [4]. As such, haemovigilance information systems are required to give support to the monitoring of the safety of the blood supply to the organizations that provide or use blood products for patient treatment.

Among healthcare information systems, blood establishments and clinical pathology laboratories are considered pioneers in the use of information technologies (IT) [5]. Innovative efforts are mainly focused on improving the quality of the overall process responsible for the safe administration of blood components to patients [6]. Also, efforts have been done in increasing the efficiency of the blood chain process, namely, through a better management of the growing volume of information, a substantial reduction of paper records, the operation of mechanisms to decrease transcription errors, and a more efficient management of blood stocks [7]. For example, the implementation of labelling standards, such as the ISBT128, has enabled the unambiguous and language-independent identification of blood component units worldwide.

IT plays a key role in contributing to the safety and traceability of the transfusion chain, implementing/supporting the growing number of international guidelines

and standards [8–13]. Current European Union legislation requires all member states to design and implement haemovigilance information systems to monitor the quality and safety of blood components for transfusion [13]. These information systems receive, process, and analyse notifications of transfusion reactions and adverse incidents occurring throughout the national blood transfusion chain, that is, from donor to recipient, where the healthcare professionals are responsible for such notification—hence the designation of notifiers.

These notifications consist in an extended set of questions that aim to characterize the adverse event or reaction that took place. For example, the notification of a patient adverse reaction should contain data about patient age, gender, diagnostics and reason for transfusing, symptoms and signs, blood component transfused, time and place of transfusion, reaction severity and imputability, type of transfusion reaction, and so forth.

Since the 90s, several national competent authorities (e.g., the Ministry of Health and the Health Authority) have created nationwide haemovigilance systems. In 1998, European countries founded the European Haemovigilance Network [14] with the purpose of exchanging best practices and benchmarks between the national haemovigilance systems [15]. Later, the International Haemovigilance Network (IHN) was created to accommodate non-EU members, becoming the single worldwide haemovigilance institution and a collaborating partner of the International Society of Blood Transfusion and the World Health Organization [16]. Nowadays, IHN has 29 nation members, namely, Australia, Austria, Belgium, Canada [17], Brazil [18], Croatia [19], Denmark, Finland, France [20], Germany, Greece, Iceland, Ireland [21], Italy [22], Japan, Luxembourg, Malta, Netherlands [23], Norway [24], New Zealand, Portugal, Singapore, Slovenia, South Africa, Spain, Sweden, Switzerland, United Kingdom [25], and United States of America [26].

Although different expert groups have publicly discussed both the events to be registered and the notification criteria these systems should comply with, little has been disclosed about their implementation and, most notably, the general evolution of IS/IT in haemovigilance. With the goal of gaining a better understanding about the IT involved in active and evolving haemovigilance systems, we conducted two anonymous Web surveys among the IHN members. These surveys collected information on the IS/IT portrait of the systems at two different points in time and focused on key aspects of the implementation of data registering, validation, and reporting. Also, they covered some primary aspects of data structuring and data security.

The rest of the paper is structured as follows: Section 2 describes the material and methods supporting these surveys, that is, the experimental design; Section 3 presents the results of the study, exposing several procedural and technological aspects; and Section 4 discusses the general evolution of haemovigilance systems and the main lessons learned. Finally, Section 5 provides some concluding remarks and outlines future work.

2. Materials and Methods

With the support of the IHN, two Web surveys were conducted among its members to collect information about the means used by their haemovigilance systems to collect data about patient and donor adverse reactions and events.

2.1. First Survey. The first survey aimed to obtain the first portrait of the haemovigilance system notification process. Therefore, this survey covered the main procedural aspects, notably the mechanisms of notification, the structure of the notifications, data management and analysis, and a general description of the IS/IT solution in use.

The mechanisms of reception of notifications were categorized into electronic (e.g., via e-mail or Web site), paper records (e.g., via fax or mail), or both simultaneously. In the case of Web-supported notifications, questions about the implementation included in-house versus outsourced development, the programming language, the database engine, and the inclusion of data safety measures. For the structure of the notification, attention was set on the use of plain, free-text descriptions, prestructured questionnaires with or without some free-text areas, and "guided" questionnaires, that is, where questions would be prompted according to previous answers.

Questions were made about the storage of notifications (e.g., the paper record was stored or the notification was transcribed), data validation (e.g., done automatically or by an expert), and reporting functionalities (e.g., transcribing data into spreadsheets or data access through the Web site).

The survey also covered the natural evolution of the systems, in particular, the dates of the system debut and its last update/revision, the number of staff members involved in haemovigilance management, and the number of registered institutions (healthcare facilities where blood is collected from donors and/or transfusion of blood components is performed), registered notifiers, and annual notifications received. Finally, the observation field allowed respondents to add any comments they considered relevant and not anticipated by the questionnaire.

A copy of this survey is presented in Supplementary 1.

2.2. Second Survey. Four years after the first survey, the same participants were asked to fill in a follow-up survey that accounts for any changes made to their systems, namely, in terms of the mechanisms of reception of notifications, the structure of the notifications, and notification management, validation, and reporting.

In both surveys, answers were received by an e-mail and through an Online Google Form, and results were anonymised.

A copy of this second survey is presented in Supplementary 2.

3. Results

The first survey was conducted between July 15 and November 7, 2013. At the end, we collected 23 answers from haemovigilance organizations of 21 different countries. While 21 countries might seem a small number at a global

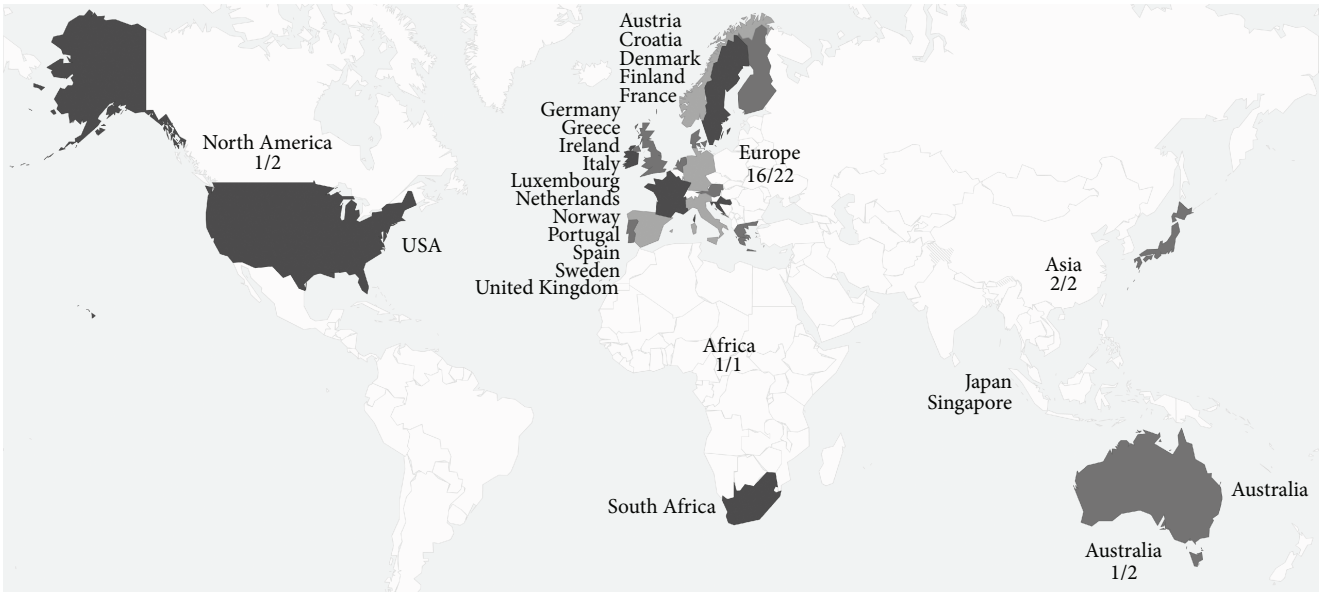
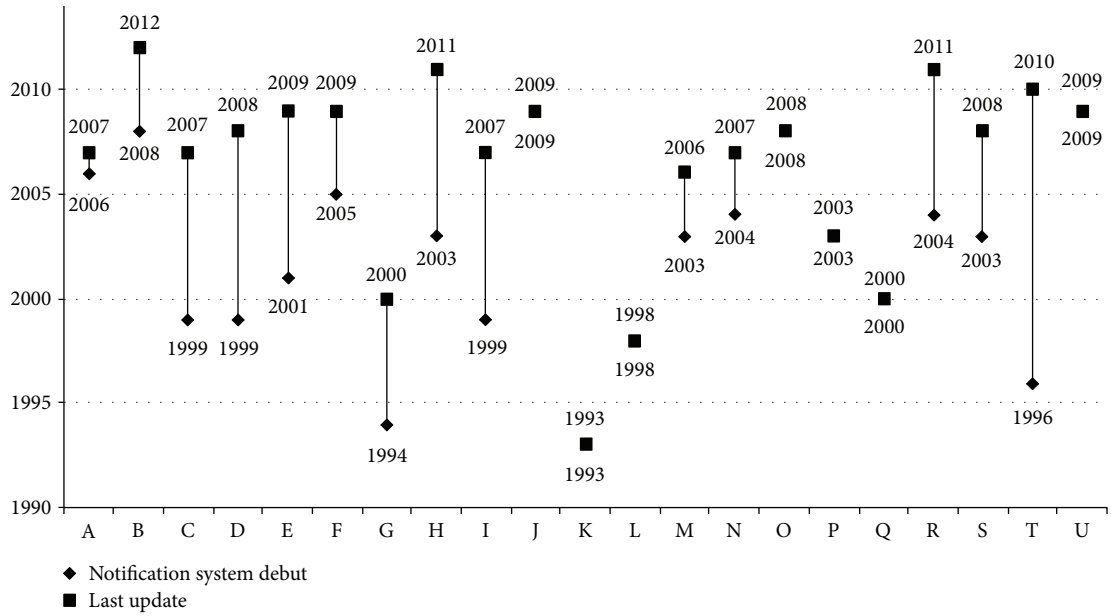


Figure 1: Distribution of IHN member responses.



Notification system debut
Last update

Figure 2: Chronology of the debut and last update of the haemovigilance systems that responded to the first survey. Letters A to V represent each system.

scale, they represent a response rate from IHN members of 72.4% (Figure 1). The fact that haemovigilance is a relatively new area and that IHN is the worldwide reference in this field grants that this response number is quite significant.

Two of the IHN members that participated in the survey indicated that they maintain separate data management systems for patient and donor notifications, and another participant indicated that the corresponding data management system does only record the notifications of a region of the country.

Concerning the institutions and notifiers registered in the system, as well as the number of annual notifications, the participants described very different operational realities. For comparison purposes, the extent of the system is considered directly related to the number of annual notifications. Systems were grouped in the following intervals: less than 50, between 51 and 100, between 101 and 250, between 251 and 500, and more than 500 notifications per year.

Regarding longevity, 31.8% of the haemovigilance systems were implemented between 1993 and 1999 (Figure 2). The year when a larger number of debuts was registered was 2003 (18.2%), and the last system launch occurred in 2009. Moreover, participants declared that 63.6% (14 of the 22 systems) of these haemovigilance systems have been

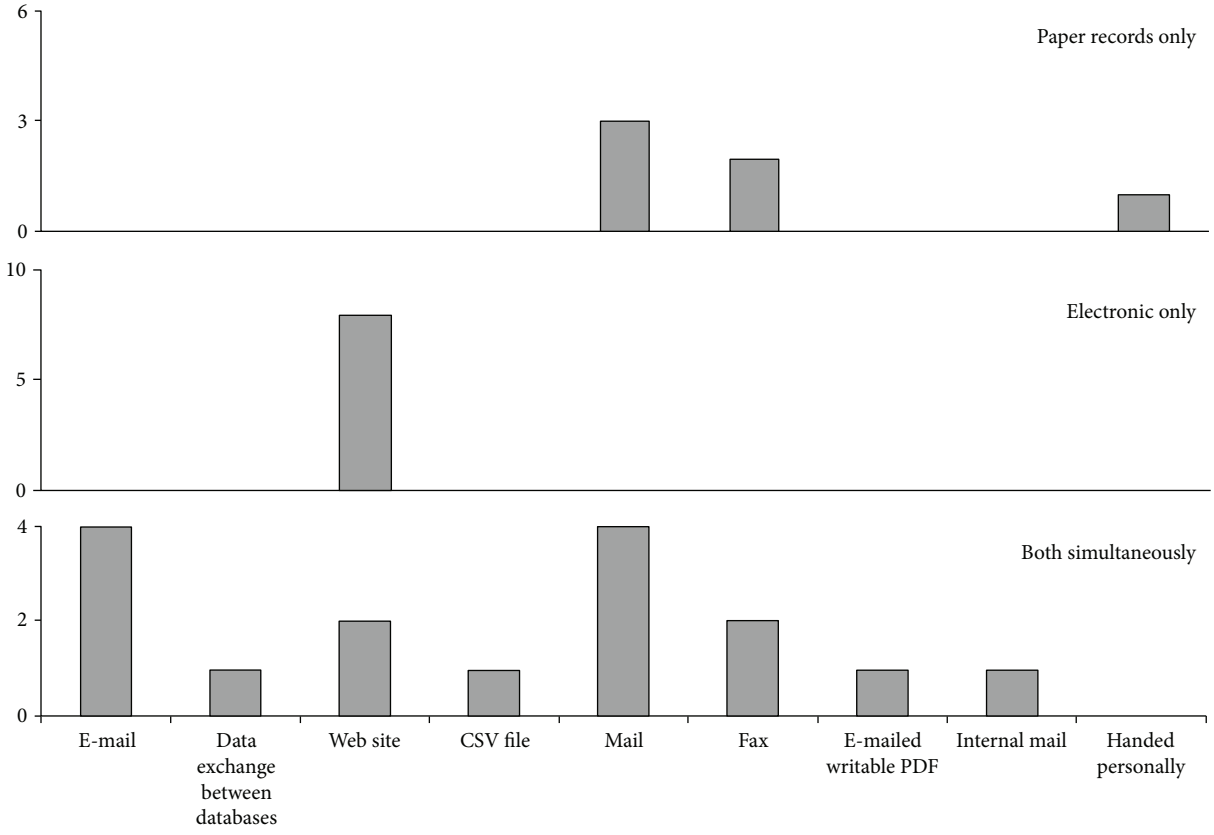


Figure 3: Categorization of haemovigilance systems according to the ways they collect notifications: systems that receive paper record notification, those that receive electronic notification, and those that receive electronic and paper record notifications simultaneously. There is one more haemovigilance system that uses a Web site, but we excluded it from this chart, since it has not answered the question regarding what type of notification reception methods was used.

redesigned from an infrastructure reengineering perspective. The oldest system, which started its activity in 1993, still maintains the original notification method, but the respondent stated in the observation field that a new electronic version is going to be released shortly.

The number of people enrolled in the haemovigilance system, namely, IT staff, office workers, and scientific staff, grows as the number of notifications increases. In systems receiving 50 or less notifications per year, the average number of staff is 1.0. For systems dealing with hundreds or more notifications, the average number of staff visibly increases: 1.8 in systems receiving between 101 and 250 notifications, 4.7 in systems receiving between 251 and 500 notifications, and 7.0 in the larger systems.

Looking into the IT staff of the larger systems, the 5 larger systems that do not use a Web site have a low average of IT staff (0.8); in the 6 larger systems that have a contracted Web site, this number almost doubles (1.58); and in the 2 larger systems, which developed their own Web site, the average of IT staff is 0.75.

Focusing on the notification method, 23.8% of the 21 responses receive only paper record notifications (one of the systems receives notifications both by a mail and by being handed personally), 38.1% receive only electronic notifications, and 38.1% receive both (paper record and electronic notifications) simultaneously. Figure 3 details the reception methods used. Noteworthy, the systems that rely only on electronic notifications use the Web site as means of submission.

Table 1: Notification reception methods. While paper records (either exclusively or simultaneously with electronic notifications) are present in systems from all sizes, all exclusively electronic interfaces are found in systems with more than 250 annual notifications.

Reception method	Number of annual notifications				
	≤50	51–100	101–250	251–500	>500
Paper records only	1			1	3
Both simultaneously	1		2	1	4
Electronic only				2	6

Tables 1 and 2 summarize the collected data regarding notification reception methods and the structure of the notifications.

In 45.5% of the systems, the notifications are automatically stored; that is, the notification record is directly stored in the database. Regarding the rest, 4.5% of the systems transcribe (e.g., to an Excel spreadsheet) the notifications, 18.2% of the systems store the notification in paper records, 22.7% of the systems store and transcribe the notifications, and the remaining 9.1% of the systems have a mixed system

Table 2: Structure of the notification. 90.9% of the participants receive notifications using prestructured questionnaires.

Structure of the notification	Number of annual notifications				
	≤50	51–100	101–250	251–500	>500
Prestructured questionnaire	1				2
Prestructured questionnaire and free-text areas	1		2	4	9
Questions placed according to previous answers					1
No answer					1

(some notifications are automatically stored in a database and others are transcribed). Automatic data validation is fully implemented in only 9.1% of the systems and is partially implemented in 13.6% of the systems. In turn, 68.2% of the systems validate the notification manually, and apparently, the remaining 9.1% do not validate the notifications.

Reporting and analysis tools are fully integrated in 27.3% of the haemovigilance systems, data importation is supported in 31.8% of the systems (e.g., database connection or file upload), and manual transcription of data to third-party analytical software like SPSS or Excel is used in most systems (40.9%).

Finally, the predominant database solutions are Oracle (36.4%) and MySQL (36.4%) database management systems. 18.2% of the systems use Microsoft SQL Server or InterSystems Caché, and 9.1% of the participants did not specify the database engine in use. The most common Web programing languages are JavaScript (31.8%), PHP/HTML (27.3%), Java (13.6%), and C, C++, or C# (9.1%). The remaining systems indicated the use of .NET, InterSystems Caché Object Script, XML, and webMethods or did not know such specifics. Table 3 summarizes the protection mechanisms implemented by the systems. The maximum number of measures is 7 (one system), and the minimum number of measures is zero (one system). Login-based access, password protection, data privacy, and data anonymity are the measures more frequently implemented.

The second survey took place between September 18 and 27, 2017, and aimed to identify recent or soon-to-happen developments and/or procedural changes. Half of the respondents stated no changes in the haemovigilance systems in the 4-year interval between surveys. Three haemovigilance systems have changed the notification procedures: the first has replaced a word document, the second transcribes the received data to a database through a specific software, and the third now collects data directly from their notifiers. One haemovigilance system has changed the notification structure, replacing free-text questions by check boxes (i.e., avoiding free-text descriptions), and the other is implementing "adaptable questions," that is, questions placed according to previous answers.

Among the systems maintaining paper records, one has started digitizing these records and storing the obtained PDF in a database. Also noticeably, one system has incorporated benchmark software to give feedback to the notifiers, namely, to obtain summary reports.

In terms of the validation of notifications, the most significant developments are the manual validation on all notifications now implemented in one of the systems and the testing of an "external validation method" in another system.

4. Discussion

This survey offers a unique view of the IS/IT landscape, and general practices, of the haemovigilance systems worldwide. The high rate of response (72.4%) reflects the strong involvement and interest of IHN members in these matters.

The surveyed systems present a great heterogeneity, most notably in terms of the number of notifications received and the number of notifiers and registered institutions. Arguably, this variability can be explained by the intricacies of the transfusion chain activity and by the level of maturation of the systems at the national level. For example, we notice the existence of countries with less than 10 million habitants where their haemovigilance system receives more than 500 notifications per year, while others with more than twice the population receive half of these notifications. These statistics could be the subject of further study, trying to correlate the number of notifications with the transfusion chain activity, the gross domestic product spent on health (a possible outcome of such a study could confirm or refute that less investment in healthcare may lead to less attention/effort put on these matters), the rate of population, and the number and type of professionals working in the transfusion area, among others.

4.1. Notification Data Reception and Management. Generally, haemovigilance systems that receive a larger number of notifications have computer-assisted solutions. When the system receives only electronic notifications (40%), the Web is the preferred means of user-system interaction. Actually, 25% of the 12 systems that still use paper records stated that the implementation of electronic notification procedures will be conducted in the near future.

Regarding the structure of the notification, only 1 participant uses a "tailored" method where questions are placed according to previous answers, while all the others pose a predetermined set of questions, which may not always apply to the notified event.

Probably, the reduced number of systems performing the automatic validation of the notifications (only 9%) reflects the lack of control over the submitted data. A paper notification that is faxed to the haemovigilance system must be examined to check if all required questions were answered, whereas in a Web-based system, the form submission can be blocked if the required questions are not answered. Moreover, the Web form analyses the answers as they are inserted, guiding the notifier through the notification. As an example, a Web-based system can hamper a notifier from classifying an adverse reaction as "febrile nonhaemolytic reaction" if the symptom "fever" was not previously selected. In a system

Table 3: Safety measures implemented by the surveyed haemovigilance Web sites.

Safety measure	Web-based information system											Total
	A	*B*	*C*	*D*	*E*	F	G	H	I	*J*	*K*	
Access to site is password protected	*Y*	*Y*	*Y*	*Y*	*Y*	N	Y	Y	Y	*Y*	*Y*	10
Data privacy, that is, each participant is granted access only to his/her data	*Y*	*N*	*Y*	*Y*	*Y*	N	Y	Y	N	*Y*	*Y*	8
Data anonymity, namely, by masking the identification of clinical facilities, patients, and so forth	*Y*	*N*	*Y*	*Y*	*Y*	N	Y	N	Y	*Y*	*Y*	8
Data operation conditioned by user credential	*Y*	*N*	*Y*	*Y*	*N*	N	Y	Y	N	*Y*	*Y*	7
User sessions	*Y*	*N*	*Y*	*Y*	*N*	N	Y	N	N	*Y*	*Y*	6
Secure HTTP	*Y*	*N*	*N*	*Y*	*Y*	N	N	N	N	*N*	*Y*	4
Encryption of sensitive data	*N*	*N*	*N*	*Y*	*N*	N	N	N	N	*N*	*N*	1
Authentication with a national professional health card	*Y*	*N*	*N*	*N*	*N*	N	N	N	N	*N*	*N*	1
Total of safety measures implemented	*7*	*1*	*5*	*7*	*4*	0	5	3	2	*5*	*6*	

Letters A to K represent the haemovigilance systems. The columns referring to outsourced systems are in italic.

with automatic validation, the number of notifications that require expert supervision will certainly be lower.

Automatic analytical reporting is also poorly implemented. Only 26% of the systems automatically generate reports, whereas the rest rely on human resources to import or transcribe the data to third-party software for further processing.

4.2. IT Details. IT investment is economically affordable. Most of the existing systems use technologies that are free of charge, namely, MySQL, PHP, HTML, JAVA, and JavaScript. Furthermore, the size of the IT team seems to be related to the volume of data that the haemovigilance system is required to manage, notably the number of notifications, and unrelated to the method of collecting notifications; that is, even in systems with no Web interface, IT staff is still necessary for subsequent management/analytical tasks like data importing, handling, and reporting.

Overall, electronic data storage, computer-aided data validation, and automatic generation of analytical reports represent the added value of nowadays IT-based haemovigilance systems, both at operational and management levels. These modules support daily routines, reducing the time and human effort involved, whereas enforcing data quality and enabling decision-making. The existence of open-answer questions is an example, as they require more attention by haemovigilance experts.

On another point, we would like to raise awareness to in-house development and the enforcement of general safety measures. Survey results show that in-house-developed systems (F, G, H, and I in Table 3) have an average of 2.5 general safety measures, while outsourced systems (A, B, C, D, E, J, and K in Table 3) have an average of 5 general safety measures. This is a matter that, in our opinion, deserves improvement.

4.3. Functionalities Implemented Since 2013. The data collected on the second survey shows that 50% of the systems did not experience significant developments in 4 years. The changes implemented in the other systems aim for the reduction of paper records, modifications in the notification forms (i.e., replacement of free-text questions by multiple-choice questions), and the transcription of information to other software for further reporting.

5. Conclusions

After collecting insights into the experience and opinion of those closely connected to the management of national haemovigilance systems worldwide, it is clear that a significant number of haemovigilance systems have invested in the elimination of paper records and the reduction of free-text questions in notification forms, that is, the automatic reception of notifications and the enforcement of the quality of these notifications. There are an increasing number of systems choosing Web-based solutions, and there is an increasing interest in developing analytical functionalities, namely, to comply with national and European legislation.

Although some of the haemovigilance systems still use nonelectronic notification systems, current developments and feedback from national administrators pinpoint system interconnection and data interoperation/sharing as the future milestones for IS/IT solutions in haemovigilance. The rationale is that sharing data across systems would result in higher data completeness and consistency, enabling the generation of standardized, real-time reports about relevant quality and safety indicators for the European Commission and similar organizations. Therefore, national haemovigilance systems could benefit from stronger international guidelines for the implementation and maintenance of their notification process.

5.1. Future Directions. It is our belief that the discussion of large-scale, integrative data management approaches, namely, data federation and data warehousing systems, would push towards the development of European or international data repositories. An analysis of the data collected until now could lead to the replacement of open-answer fields by multiple-choice questions. For example, instead of asking to describe the signs and symptoms, put a list of them, basing this list in the ones described until now.

Disclosure

This document reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained herein.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work would not be possible without the precious collaboration of the IHN and the generous participation of the survey respondents. The SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from University of Vigo for hosting its IT infrastructure. This work was partially funded by the European Union's Seventh Framework Programme FP7/REGPOT-2012-2013.1 (Grant Agreement no. 316265, BIOCAPS), the Fundação para a Ciência e a Tecnologia (FCT) under the scope of the strategic funding of the UID/BIO/04469/2013 unit, and COMPETE 2020 (POCI-01-0145-FEDER-006684).

Supplementary Materials

Supplementary 1. A copy of the first survey.

Supplementary 2. A copy of the second survey.

References

[1] M.-P. Vo Mai, L. Fauveau, N. Ounnoughene, I. Sandid, and P. Renaudier, "Dix-sept ans d'hémovigilance en France : bilan, perspectives," *Transfusion Clinique et Biologique*, vol. 18, no. 2, pp. 140–150, 2011.

[2] D. Poles on behalf of the Serious Hazards of Transfusion (SHOT) Steering Group, *The 2016 Annual SHOT Report (2017)*, 2016, https://www.shotuk.org/wp-content/uploads/SHOT-Report-2016_web_11th-July.pdf.

[3] R. R. P. de Vries, J.-C. Faber, and P. F. W. Strengers, "Haemovigilance: an effective tool for improving transfusion practice," *Vox Sanguinis*, vol. 100, no. 1, pp. 60–67, 2011.

[4] T. J. Nel, "Clinical guidelines, audits and hemovigilance in managing blood transfusion needs," *Transfusion Alternatives in Transfusion Medicine*, vol. 10, no. 2, pp. 61–69, 2008.

[5] E. Brodheim, W. Ying, and R. L. Hirsch, "An evaluation of the CODABAR symbol in blood-banking automation," *Vox Sanguinis*, vol. 40, no. 3, pp. 175–180, 1980.

[6] I. R. Clark, J. Parekh, M. Peters, I. Frew, and R. N. Ibbotson, "Hospital blood bank laboratory data processing system," *Journal of Clinical Pathology*, vol. 37, no. 10, pp. 1157–1166, 1984.

[7] G. Sharma, A. V. Parwani, J. S. Raval, D. J. Triulzi, R. J. Benjamin, and L. Pantanowitz, "Contemporary issues in transfusion medicine informatics," *Journal of Pathology Informatics*, vol. 2, no. 1, p. 3, 2011.

[8] J. Sampson, R. Nozick, P. Ashford et al., "ISBT guidelines for validation of automated systems in blood establishments," *Vox Sanguinis*, vol. 98, pp. 1–19, 2010.

[9] P. Ashford, D. Gozzard, J. Jones et al., "Guidelines for blood bank computing," *Transfusion Medicine*, vol. 10, no. 4, pp. 307–314, 2000, Available from:.

[10] European Directorate for the Quality of Medicines & HealthCare (EDQM) C of, *Guide to the Preparation, Use and Quality Assurance of Blood Components*, Council of Europe, Strasbourg, 19th edition, 2017.

[11] European Parliament C of the EU, *Directive 2002/98/EC of the European Parliament and of the Council of 27 January 2003 Setting Standards of Quality and Safety for the Collection, Testing, Processing, Storage and Distribution of Human Blood and Blood Components and Amending Directive 2001*, pp. 30–40, 2003, https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-1/dir_2002_98/dir_2002_98_en.pdf.

[12] Commission E, *Commission Directive 2005/62/EC of 30 September 2005 Implementing Directive 2002/98/EC of the European Parliament and of the Council as Regards Community Standards and Specifications Relating to a Quality System for Blood Establishments*, pp. 41–48, 2005.

[13] European Commission, *Commission Directive 2005/61/EC of 30 September 2005 Implementing Directive 2002/98/EC of the European Parliament and of the Council as Regards Traceability Requirements and Notification of Serious Adverse Reactions and Events*, pp. 32–40, 2005, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005L0061&from=EN.

[14] J. C. Faber, "Haemovigilance in Europe: the European haemovigilance network," *Transfusion Clinique et Biologique*, vol. 8, no. 3, pp. 285–290, 2001.

[15] J.-C. Faber, "Work of the European Haemovigilance Network (EHN)," *Transfusion Clinique et Biologique*, vol. 11, no. 1, pp. 2–10, 2004.

[16] C. Politis, J. C. Wiersum, C. Richardson et al., "The International Haemovigilance Network database for the surveillance of adverse reactions and events in donors and recipients of blood components: technical issues and results," *Vox Sanguinis*, vol. 111, no. 4, pp. 409–417, 2016.

[17] P. Robillard, K. I. Nawej, and K. Jochem, "The Quebec hemovigilance system: description and results from the first two years," *Transfusion and Apheresis Science*, vol. 31, no. 2, pp. 111–122, 2004.

[18] D. M. Mota, D. R. Freitas, and W. N. Araújo, "Evaluation of the system of sanitary vigilance of blood at the federal level, Brazil, 2007," *Ciencia & Saude Coletiva*, vol. 17, no. 1, pp. 191–202, 2012.

[19] T. Vuk, M. Barišić, T. Očić, I. Mihaljević, D. Šarlija, and I. Jukić, "Error management in blood establishments: results of eight years of experience (2003-2010) at the Croatian Institute of Transfusion Medicine," *Blood Transfusion*, vol. 10, no. 3, pp. 311–320, 2012.

[20] D. Rebibo, L. Hauser, A. Slimani, P. Hervé, and G. Andreu, "The French haemovigilance system: organization and results for 2003," *Transfusion and Apheresis Science*, vol. 31, no. 2, pp. 145–153, 2004.

[21] D. Lundy, S. Laspina, H. Kaplan, B. Rabin Fastman, and E. Lawlor, "Seven hundred and fifty-nine (759) chances to learn: a 3-year pilot project to analyse transfusion-related near-miss events in the Republic of Ireland," *Vox Sanguinis*, vol. 92, no. 3, pp. 233–241, 2007.

[22] A. Giampaolo, V. Piccinini, L. Catalano, F. Abbonizio, and H. J. Hassan, "The first data from the haemovigilance system in Italy," *Blood Transfusion*, vol. 5, no. 2, pp. 66–74, 2007.

[23] J. C. Wiersum-Osselton and M. R. Schipperus, "Transfusion reactions in patients: haemovigilance reports to the Dutch National Haemovigilance Office in 2003," *Ned Tijdschr Geneeskd*, vol. 149, no. 47, pp. 2622–2627, 2005, http://www.ncbi.nlm.nih.gov/pubmed/.

[24] C. T. Steinsvåg, A. Espinosa, and Ø. Flesland, "Eight years with haemovigilance in Norway. What have we learnt?," *Transfusion and Apheresis Science*, vol. 49, no. 3, pp. 548–552, 2013.

[25] M. F. Murphy, "Surveillance of transfusion errors: putting data to use in the U.K," *Developments in Biological*, vol. 120, pp. 179–187, 2005, http://www.ncbi.nlm.nih.gov/pubmed/16050172.

[26] S. L. Sime, "Strengthening the service continuum between transfusion providers and suppliers: enhancing the blood services network," *Transfusion*, vol. 45, no. s4, pp. 206S–223S, 2005.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 6916204, 6 pages
https://doi.org/10.1155/2018/6916204

Research Article

Projection Mapping User Interface for Disabled People

Julius Gelšvartas,[1] **Rimvydas Simutis**,[1] **and Rytis Maskeliūnas**[2]

[1]*Automation Department, Faculty of Electrical and Electronics Engineering, Kaunas University of Technology, Studentu St. 50-154, Kaunas, Lithuania*
[2]*Centre of Real Time Computer Systems, Kaunas University of Technology, Baršausko St. 59-338a, Kaunas, Lithuania*

Correspondence should be addressed to Julius Gelšvartas; julius.gelsvartas@gmail.com

Received 28 September 2017; Accepted 15 January 2018; Published 1 March 2018

Academic Editor: Reima Suomi

Difficulty in communicating is one of the key challenges for people suffering from severe motor and speech disabilities. Often such person can communicate and interact with the environment only using assistive technologies. This paper presents a multifunctional user interface designed to improve communication efficiency and person independence. The main component of this interface is a projection mapping technique used to highlight objects in the environment. Projection mapping makes it possible to create a natural augmented reality information presentation method. The user interface combines a depth sensor and a projector to create camera-projector system. We provide a detailed description of camera-projector system calibration procedure. The described system performs tabletop object detection and automatic projection mapping. Multiple user input modalities have been integrated into the multifunctional user interface. Such system can be adapted to the needs of people with various disabilities.

1. Introduction

People suffering from severe disabilities such as tetraplegia can also experience some speech pathologies. Such conditions drastically affect the lives of disabled people as well as their relatives. Constant care is usually inevitable, and this is both time-consuming and expensive. There is, therefore, a need for assistive technologies that can be used to perform multiple tasks. Efficient assistive technologies can improve patient's quality of life, reduce the need for care, and increase independence.

Assistive technology is a very broad term covering adaptive, rehabilitative, and assistive devices that help people perform tasks that they were formerly unable to accomplish. Assistive systems usually consist of two parts, namely, assistive devices and human computer interaction (HCI) interface. Unfortunately, most consumer grade-assistive devices are either inefficient or very expensive. Specialized user interface (UI) applications can improve assistive device efficiency and adaptability.

Each disability is unique and can change over time; therefore, the best outcomes can only be achieved by personalizing assistive technologies and UI. Majority of assistive devices have a dedicated HCI software. This makes it difficult to change devices, because the user has to learn new software. This paper presents an architecture of a multifunctional UI that can be used with multiple assistive devices. Moreover, such architecture makes it possible to use multiple devices simultaneously.

Brain computer interface (BCI) has made significant advances in recent years. Advanced BCI devices offer very natural HCI method. It has been shown in [9] that BCI systems can be used to directly control robotic arms. More recently, similar system has been used to control patient's arm [1]. These advances are, however, made with expensive and usually invasive BCI devices. Consumer-grade BCI devices are still inefficient and produce unreliable results when user is distracted [12].

Eye tracking is another technology that is advancing significantly. Moreover, affordable eye tracking devices have

been shown to produce reasonable results [4]. Gaze control interfaces are already used by disabled people [14]. Eye trackers also have some limitations. The tracking can be significantly affected by head movements and sudden lighting changes.

Sip/puff switch is the most popular assistive device. This device can be easily integrated into any assistive system and is inexpensive. Sip/puff device generates two input signals that are mapped to mouse button clicks on a computer. Despite their limitations, switch devices are successfully being used for assistive and rehabilitation purposes [16].

Multifunctional UI presented in this paper can be used to perform several different tasks, namely, text input and object selection. Text input functionality uses a specialized text predictor described in [6]. Advantages of using multiple assistive devices simultaneously are presented in [5]. This paper focuses on object selection functionality of our system. Objects should be placed on a tabletop in front of the user. There are two use cases for object selection UI. First, this UI can be used by a person that is unable to speak but can use an assistive device. The user can then inform a caregiver that he wants a particular object by highlighting it with a projector. This creates a natural interaction method, because both user and caregiver see the highlighted object in the scene. Second, the selected object could be manipulated by an actuator. This would require an intelligent robotic arm that could manipulate the object when it is selected. For example, when the user selects a glass of juice, it is grabbed by a robotic arm and brought to user's mouth so that he can drink from it.

The main advantage of our object selection UI is natural HCI. Conventional UI use displays to present information for the user. In this case, the user would have to switch his attention from the monitor to the scene. Another alternative UI presentation method is using virtual reality. Severely disabled people have communication barriers, and using virtual reality headset would further separate the user from the environment. Whereas, projection mapping UI is visible not only for the system user but also by people nearby. Augmented reality has successfully been used in assistive technologies and rehabilitation [10].

The presented system consists of projector, depth camera (such as Microsoft Kinect), and at least one user action input device. Depth camera output is used to detect tabletop plane and objects positioned on that plane. Finally, projector is used together with depth camera to create a camera-projector system that performs automatic projection mapping.

Projection mapping can either be manual or automatic. Manual projection mapping is mostly used in entertainment and art industries where the scene is static. Automatic projection mapping is a more sophisticated method that works in dynamic environments. Automatic projection mapping needs 3D information of the scene obtained with depth camera. Depth camera information has to be transformed into projector optical frame. This transformation is obtained by calibrating camera-projector system. Method presented in [11] can be used when the camera is already calibrated. Alternatively structured light can be used for camera-projector calibration [13]. Both of these methods are difficult; therefore, this paper utilized a practical calibration method proposed in [22].

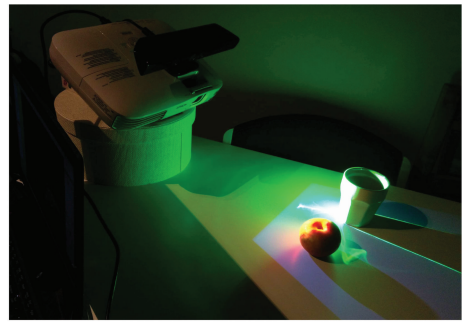


Figure 1: Example scene showing object selection UI that is shown to the user with camera-projector system. The system performs automatic projection mapping using depth camera information.

The system presented in this paper is similar to [2], but without the accounting for deformations caused by physical objects. Accounting for deformations is not necessary in our system, because only nonoccluded objects are detected and highlighted. More advanced dynamic projection mapping methods have been created in recent years [19]. Such systems require more expensive hardware setup. Figure 1 shows the experimental setup of the presented system.

The remaining paper is structured as follows. Section 2 describes the proposed projection mapping-based system. The results and discussions are presented in Section 3. Section 4 is the conclusion.

2. Materials and Methods

The object selection UI highlights objects in the scene using automatic projection mapping. Automatic projection mapping can be split into three distinct stages. First, camera-projector system has to be calibrated. Calibration is a one-time procedure performed during system setup. The second stage is object detection in depth camera images. And the third stage is object highlight calculation and display using projector. Both the second and third steps are performed continuously to create a dynamically changing user interface.

2.1. Camera-Projector Calibration. Calibration process begins by setting up the camera-projector system. Camera and projector have to be fixed sturdily to each other. Ideally, the system should be fixed to a common metal frame or integrated into one case. Camera and projector position or orientation changes with respect to each other invalidate calibration, and the process has to be repeated. The camera-projector system position can change after calibration is finished.

Calibration is performed by placing a small calibration board in a form of camera-projector system. The board has to be stationary for one second before capturing calibration image. The system is calibrated once a predefined number of images are captured. Calibration board with random circular black dot pattern is used. During calibration, projector

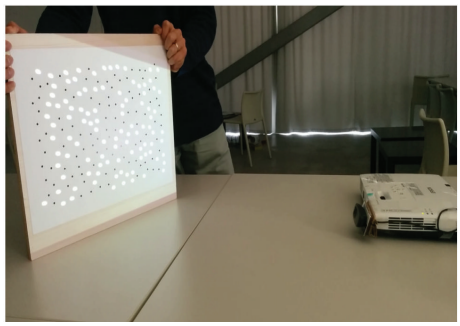


Figure 2: Camera-projector system calibration using a circular black dot pattern. Image was taken from [22].

shows a similar white dot pattern that appears on the calibration board. Figure 2 shows calibration procedure. Pattern detection and tracking are performed using local geometric consensus algorithm [21]. Detected point correspondences are used to estimate camera-projector system intrinsic and extrinsic parameters.

Both camera and projector are described using pinhole camera model. In this case, projector is treated as an "inverse camera." Camera intrinsic parameters consist of two matrices, namely, camera matrix **K** and distortion coefficient matrix **D**. More precisely,

$$\mathbf{K} = \begin{pmatrix} fx & \gamma & px \\ 0 & fy & py \\ 0 & 0 & 1 \end{pmatrix}, \quad (1)$$

where f_x and f_y are the focal lengths, (p_x, p_y) is the position of the principal point, and γ is the camera skew. And distortion coefficients are defined as follows:

$$\mathbf{D} = (k_1 \quad k_2 \quad p_1 \quad p_2 \quad k_3), \quad (2)$$

where k_1, k_2, k_3 are the radial distortion coefficients. p_1 and p_2 are the tangential distortion coefficients. A more detailed description can be found in [3]. Calibration process also estimates extrinsic camera-projector system parameters, namely, a translation vector **T** and rotation matrix **R**. **T** and **R** are defined as transformation from projector optical origin to camera optical origin.

Calibration algorithm continuously detects calibration board circles. Detected circle positions are matched to circle pattern model. The camera is calibrated when sufficient number of corresponding circles is detected. Camera calibration is performed using [23] method. Camera image lens distortions are accounted for after camera is calibrated. For the projector, the board-camera homography H_cb is calculated from detected circle correspondences. The projected circle locations can then be expressed as $H_{cb}^{-1} y^{(c)}$ where $y^{(c)}$ is the projected circle board viewed from the camera. This creates a new set of correspondences that are used to calibrate the projector and estimate camera-projector extrinsic parameters. See [22] for more detailed description of camera-projector system calibration process.

2.2. Object Detection. Object detection is performed on a 3D point cloud that is obtained by reprojecting each depth image pixel (u, v, Z) to 3D point $(X, Y, Z, 1)$. Here, u and v are pixel coordinates along image rows and columns. X and Y are the 3D point coordinates along X- and Y-axis in meters, and Z is the point distance from the camera in meters. This transformation is performed using camera projection matrix $\mathbf{P}_c$. $\mathbf{P}_c$ is obtained by combining camera matrix **K** and distortion coefficients matrix **D** [7].

The created 3D point cloud is in the camera optical frame. Detected object coordinates should be in projector optical frame coordinates. The 3D point cloud coordinates, therefore, have to be transformed using extrinsic camera-projector system parameters. The point cloud is translated using **T** and rotated using **R**.

This paper uses tabletop object detector from *Object Recognition Kitchen* package [20]. The detector has two parts, namely, a table finder and an object recognizer. The tabletop surface is detected using RANSAC algorithm described in [15]. The tabletop plane is used to extract object points from the 3D point cloud. The extracted points are clustered into individual objects. Object type is determined by comparing object clusters to object meshes stored in the database. Exact 3D object pose is estimated using iterative closest point algorithm. Object detector output is pose and type of each detected object. This information is used to create object highlights.

2.3. Highlight Image Creation. The detected object models have to be rendered into a 2D image that is shown by the projector. Visualization module of the Point Cloud Library [18] was used to render object highlights. A virtual 3D scene containing a camera is created in 3D viewer. Camera parameters are chosen such that they match intrinsic parameters of the projector.

Detected object list is tracked and updated each time a new list of detected objects is received. Each object is assigned to a unique identifier. The index of the current object is also stored so that this object can be highlighted in a different color.

Each detected object is inserted into a 3D viewer using a pose obtained during object detection. Inserted object meshes are assigned a uniform color material. When all objects are inserted into the scene an image generated using a virtual camera is shown on a projector. This image projects object highlights on real scene objects.

3. Results and Discussion

The object selection functionality described in the previous section was integrated into a specialized multifunctional UI. The system architecture of this UI application can be seen in Figure 3. Figure 3 highlights the components of the multifunctional UI system.

Depth sensor driver is used to acquire depth images from Kinect senor. Note that other depth sensors can be integrated into our system. The acquired depth images are transformed to projector frame and used in object detection module. The system can use any detector that works with 3D point clouds

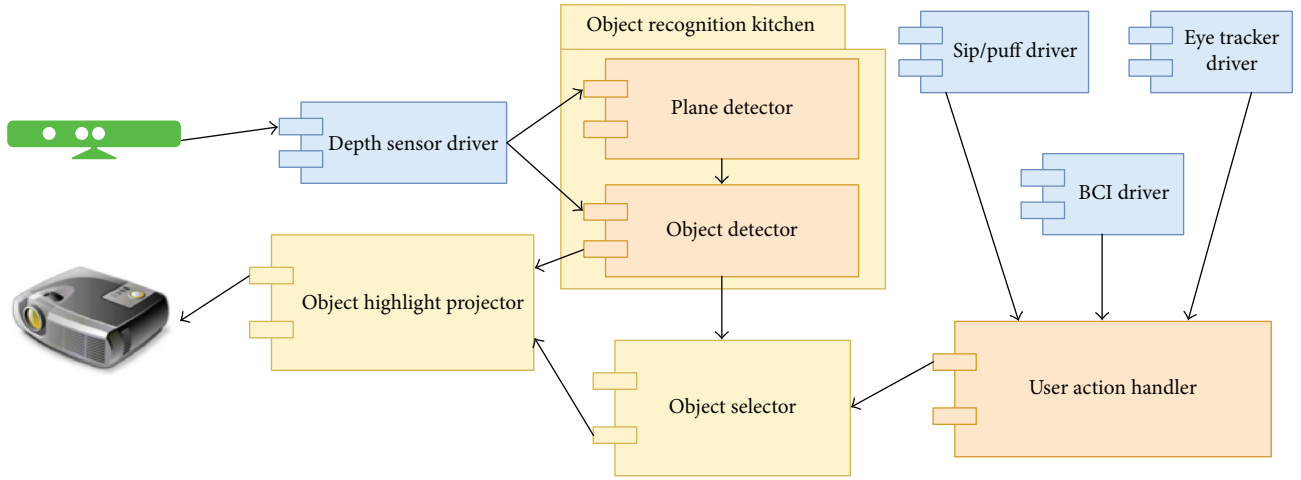


Figure 3: Multifunctional user interface system architecture. UI integrates multiple input modalities and a camera-projector system to perform automatic projection mapping.

or depth images. Current implementation used a simple tabletop object detector. A more advanced detector such as that proposed in [8] would improve detection accuracy.

User action handler is a module that handles input modality actions. This module has simple interface suited for new device integration. It is possible to configure the module to use one or multiple input modalities simultaneously. The interface supports two actions, namely, selecting current object and moving to the next object in the list. In some situations, device can only generate a single action. This problem can be solved by creating a virtual device that performs automatic moving to the next object at the given intervals. This solution is less efficient but is sometimes the only way to use the system. Current implementation supports sip/puff, eye tracking, and consumer grade BCI devices.

Object selector keeps track of all detected objects. The object list is updated each time a new depth image is processed. The update process removes objects that have been removed from the scene and adds newly appeared objects. This module also maintains an index of the current object. This index is updated by input modality actions. Finally, this package can also generate external signals, when a particular object is selected. The external signal can either be sent to the caregiver or used to initiate robotic arm actions.

Object highlight projector module implements automatic projection mapping described in the previous section. This module has access to an object mesh database. Meshes are used to generate realistic and accurate object highlights. This module also receives the current object index so that this object could be rendered in a different color.

The main advantage of described UI is that the user interacts with the environment by looking directly into the scene. Conventional UI presentation method using a display requires that the user constantly changes his attention from the display to the scene. Projection mapping also has some disadvantages. Projected light might not be visible in very bright environments, especially when the projector is not powerful. Moreover, shiny, reflective, and transparent objects can create reflection artifacts that make object highlights difficult to see. The easiest way to overcome these problems is to make sure that the scene contains only diffuse objects. The proposed system is going to be operated in a controlled environment, and the correct object choice should be performed by the system setup personnel. Alternatively, highlight visibility could be improved by projecting animated or textured highlights.

Object selection UI was evaluated using a user experience and system usability questionnaire described in [17]. Ten healthy individuals were asked to use the UI to select one of three cans placed in front of them on a tabletop surface. Each individual used the system three times with a sip/puff device for action input. After the experiment, each individual was asked to fill up the system usability questionnaire containing 16 questions. Each question was given a score from 1 to 7, where 1 means the user strongly agrees with the statement and 7 means strong disagreement. The user experience questionnaire results are summarized in Table 1. One question has been excluded from the questionnaire since it is related to system documentation and our system is only a prototype.

4. Conclusions

This paper presents a multifunctional UI designed specifically for people suffering from severe disabilities. The UI contains a natural augmented reality object selection HCI method. Augmented reality is achieved using automatic projection mapping. Calibrated camera-projector system is used to project object highlights. This paper describes the camera-projector system calibration setup and procedure.

The described object selection UI is intended for two use cases. First, the system can be used by people with speech pathologies. In this case, the user can highlight the objects that he wants. Without such interface, a caregiver has to constantly ask questions to find out what a disabled person

Table 1: Table containing user experience questionnaire and results.

(1)	Overall, I am satisfied with how easy it is to use this system.	1.8 ± 0.2
(2)	It was simple to use this system.	2.3 ± 0.8
(3)	I was able to complete the tasks and scenarios quickly using this system.	1.5 ± 1.3
(4)	I felt comfortable using this system.	4.0 ± 2.0
(5)	It was easy to learn to use this system.	2.0 ± 0.5
(6)	I believe I could become productive quickly using this system.	3.9 ± 0.7
(7)	The system gave error messages that clearly told me how to fix problems.	6.1 ± 0.9
(8)	Whenever I made a mistake using the system, I could recover easily and quickly.	5.9 ± 0.8
(9)	The information (such as online help, on-screen messages and other documentation) provided with the system was clear.	N/A
(10)	It was easy to find the information I needed.	2.1 ± 0.4
(11)	The information was effective in helping me complete the tasks and scenarios.	1.5 ± 3.1
(12)	The organization of information on the system screens was clear.	1.3 ± 0.6
(13)	The interface of this system was pleasant.	3.3 ± 2.3
(14)	I liked using the interface of this system.	2.7 ± 1.8
(15)	This system has all the functions and capabilities I expected it to have.	2.3 ± 1.3
(16)	Overall, I am satisfied with this system.	1.9 ± 0.5

wants. Second, such interface could be used in a more sophisticated setup with a robotic arm. When the user selects an object, the robotic arm can bring this object to him. In this case, the user can, for example, drink independently without the help of a caregiver.

The functionality described in this paper is part of a multifunctional UI that can be used for text entry, item, or image selection from a list and other tasks. The main advantage of a single multifunctional UI is its adaptability. The system can be configured according to the needs of individual person and adapted when his condition changes. This is achieved by integrating multiple user action input modalities.

The ideas presented in this paper could also be used with head-up display (HUD). HUD would show a more realistic object highlight and would not have problems in bright environments. Augmented reality created using HUD, on the other hand, would only be visible to the system user but not to nearby people. Available consumer grade HUD systems, however, are very expensive, whereas projectors are widespread and affordable.

The object selection interface usability has been evaluated using a specific questionnaire. Overall, the interface has been evaluated positively by the system users. Majority of users have learned how to use the system very quickly. As expected, survey relieved that the interface lacks error handling functionality. This was expected as the interface currently is a prototype and little attention has been put into error handling scenarios. These limitations will be addressed in further system developments.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research has been supported by the European Cooperation in Science and Technology (COST) Action CA15122, Reducing Old-Age Social Exclusion: Collaborations in Research and Policy (ROSEnet).

References

[1] A. B. Ajiboye, F. R. Willett, D. R. Young et al., "Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration," *The Lancet*, vol. 389, no. 10081, pp. 1821–1830, 2017.

[2] H. Benko, R. Jota, and A. Wilson, "Miragetable: freehand interaction on a projected augmented reality tabletop," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 199–208, ACM, New York, NY, USA, 2012.

[3] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, O'Reilly Media, Inc, Sebastopol, CA, USA, 2008.

[4] E. Dalmaijer, "Is the low-cost eyetribe eye tracker any good for research?," Technical report, PeerJ PrePrints, San Diego, CA, USA, 2014.

[5] J. Gelšvartas, R. Simutis, and R. Maskeliūnas, "Multifunctional user interface implementation details and evaluation," in *2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 501–504, IEEE, Miedzyzdroje, Poland, 2015.

[6] J. Gelšvartas, R. Simutis, and R. Maskeliūnas, "User adaptive text predictor for mentally disabled huntington's patients," *Computational Intelligence and Neuroscience*, vol. 2016, no. 2, Article ID 3054258, 6 pages, 2016.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2003.

[8] S. Hinterstoisser, C. Cagniart, S. Ilic et al., "Gradient response maps for real-time detection of texture-less objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 34, no. 5, pp. 876–888, 2012.

[9] L. R. Hochberg, D. Bacher, B. Jarosiewicz et al., "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, 2012.

[10] H. M. Hondori, M. Khademi, L. Dodakian, S. C. Cramer, and C. V. Lopes, "A spatial augmented reality rehab system for post-stroke hand rehabilitation," in *Medicine Meets Virtual Reality (MMVR)*, pp. 279–285, IOS Press, Amsterdam, The Netherlands, 2013.

[11] M. Kimura, M. Mochimaru, and T. Kanade, "Projector calibration using arbitrary planes and calibrated camera," in *2007. CVPR '07. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-2, IEEE, Minneapolis, MN, USA, 2007.

[12] R. Maskeliunas, R. Damasevicius, I. Martisius, and M. Vasiljevas, "Consumer-grade EEG devices: are they usable for control tasks?," *PeerJ*, vol. 4, article e1746, 2016.

[13] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pp. 464–471, IEEE, Zurich, Switzerland, 2012.

[14] A. Olsen, A. Schmidt, P. Marshall, and V. Sundstedt, "Using eye tracking for interaction," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems, pages 741–744*, Vancouver, BC, Canada, 2011ACM.

[15] J. Poppinga, N. Vaskevicius, A. Birk, and K. Pathak, "Fast plane detection and polygonalization in noisy 3D range images," *IROS 2008. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 3378–3383, IEEE, Nice, France, 2008.

[16] S. Rojhani, S. A. Stiens, and A. C. Recio, "Independent sailing with high tetraplegia using sip and puff controls: integration into a community sailing center," *The Journal of Spinal Cord Medicine*, vol. 40, no. 4, pp. 471–480, 2017.

[17] J. Rubin and D. Chisnell, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, John Wiley & Sons, 2008.

[18] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011.

[19] T. Sueishi, H. Oku, and M. Ishikawa, "Robust high-speed tracking against illumination changes for dynamic projection mapping," *2015 IEEE Virtual Reality (VR)*, 2015, pp. 97–104, IEEE, Nice, France, 2015.

[20] R. C. Willow Garage, *ORK: Object Recognition Kitchen*, 2017, https://github.com/wg-perception/object_recognition_core.

[21] L. Yang, J.-M. Normand, and G. Moreau, "Local geometric consensus: a general purpose point pattern-based tracking algorithm," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1299–1308, 2015.

[22] L. Yang, J.-M. Normand, and G. Moreau, "Practical and precise projector-camera calibration," in *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 63–70, IEEE, Merida, Mexico, 2016.

[23] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

Hindawi
Journal of Healthcare Engineering
Volume 2018, Article ID 4038034, 15 pages
https://doi.org/10.1155/2018/4038034

Research Article

Analysis of a Pulse Rate Variability Measurement Using a Smartphone Camera

András Bánhalmi,[1] János Borbás,[2] Márta Fidrich,[1] Vilmos Bilicki,[1] Zoltán Gingl,[1] and László Rudas[3]

[1]*Institute of Informatics, University of Szeged, Szeged, 2 Árpád Square 6720, Hungary*
[2]*2nd Department of Internal Medicine and Cardiology Clinic, University of Szeged, Szeged, 6 Semmelweis Street 6725, Hungary*
[3]*Department of Anesthesiology and Intensive Therapy, University of Szeged, Szeged, 6 Semmelweis Street 6725, Hungary*

Correspondence should be addressed to András Bánhalmi; banhalmi@inf.u-szeged.hu

Received 29 September 2017; Revised 20 November 2017; Accepted 7 December 2017; Published 5 February 2018

Academic Editor: Agusti Solanas

Background. Heart rate variability (HRV) provides information about the activity of the autonomic nervous system. Because of the small amount of data collected, the importance of HRV has not yet been proven in clinical practice. To collect population-level data, smartphone applications leveraging photoplethysmography (PPG) and some medical knowledge could provide the means for it. *Objective*. To assess the capabilities of our smartphone application, we compared PPG (pulse rate variability (PRV)) with ECG (HRV). To have a baseline, we also compared the differences among ECG channels. *Method*. We took fifty parallel measurements using iPhone 6 at a 240 Hz sampling frequency and Cardiax PC-ECG devices. The correspondence between the PRV and HRV indices was investigated using correlation, linear regression, and Bland-Altman analysis. *Results*. High PPG accuracy: the deviation of PPG-ECG is comparable to that of ECG channels. Mean deviation between PPG-ECG and two ECG channels: RR: 0.01 ms–0.06 ms, SDNN: 0.78 ms–0.46 ms, RMSSD: 1.79 ms–1.21 ms, and pNN50: 2.43%–1.63%. *Conclusions*. Our iPhone application yielded good results on PPG-based PRV indices compared to ECG-based HRV indices and to differences among ECG channels. We plan to extend our results on the PPG-ECG correspondence with a deeper analysis of the different ECG channels.

1. Introduction

Heart rate variability (HRV) is a rarely used clinical term, but it provides useful information about the variation between consecutive heart beats. HRV parameters could help to describe the activity of the autonomic nervous system (ANS), and through this, we can get a better picture about the status of our health [1].

Most of previous studies in the area of HRV measurements just focused on the technical or the medical aspects. Studies describing relations between these two are quite rare. Many of the medical studies do not investigate the new analytical methods, and most of the new methods have not been validated in medical experiments. The value of our study is based on this economically and medically relevant problem, and we try to solve it using tools taken from information technologies [2]. A medically relevant problem is, for example, the diagnosis and treatment of the cardiovascular diseases, which are the cause of the 37% of global mortality (2012), corresponding to as many as 17.5 million people. Out of these, 6 million people were under 70 years old and 7.4 million of them died from coronary artery disease [3]. The relation between HRV and coronary artery diseases was found relatively early [1]: "The observation that in patients with an acute MI, the absence of respiratory sinus arrhythmias is associated with an increase in "in-hospital" mortality represents the first of a large number of reports which have demonstrated the prognostic value of assessing HRV to identify high-risk patients."

Hence, this is why it would be beneficial to build up a widely accessible service, which can be used easily to measure HRV not just in a hospital but anywhere. This service could

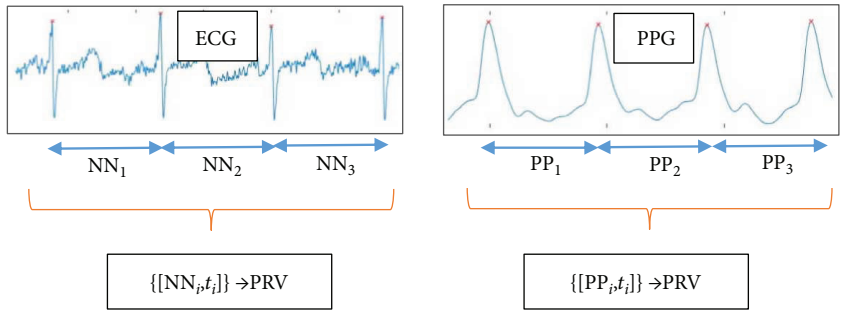


Figure 1: Connection between HRV and PRV analysis. From the ECG signal, the NN intervals (time durations) are determined, with the corresponding timestamps. The timestamps are needed when spectral analysis is applied to the NN time serial. The data for PRV is similarly obtained from the PP durations between the consecutive maximum values in the PPG signal.

improve the quality of life and survival chances of those who are diagnosed with such problems. However, there is a lack of systematic statistical evaluation of the additive prognostic value of the new methods [2]. Over the past 30 years, the literature has not provided much support for the real clinical utilization of HRV. It is anticipated that new methods will aid studies involving large populations, which hopefully will allow us to expand our physiological knowledge and improve our understanding of its clinical relevance [2]. Previously published articles presented new methods and analytical techniques on this topic [4–8]. However, a breakthrough has not been achieved since we have so far failed to collect a critical amount of data from either healthy or sick populations [2].

A new HRV registration technique is needed that can record professionally validated data even by a layman, anytime and anywhere. One possible method might be photoplethysmography (PPG), which can be used to measure pulse rate variability (PRV) on the limbs. PRV and HRV have good correlations, and it was confirmed by several previous publications [9–11]. The relation among the HRV, ECG (electrocardiogram), PRV (pulse rate variability), and PPG (photoplethysmogram) is shown in Figure 1. ECG is a voltage signal, while PPG is the time serial got from digitalizing the measurements of the reflected or absorbed light, which changes with the periodic blood flow. PRV can be easily measured with the help of a smartphone flash and camera [12, 13] (or using other low-cost tools [14]). The operation of PPG is well formulated by the authors of [15]: "PPG is measured via reflection through the illumination of the skin using an LED (e.g., the smartphone's flash) and through the detection of the amount of light that is reflected by a photodetector or a camera located next to the light source. The resulting PPG signal is composed of a direct current (DC) component, which varies slowly depending on tissue properties and blood volume. The alternating current (AC) component varies more rapidly to detect the pulsatile factor. After cardiac systole, local blood volume increases acutely, reducing the received light intensity. During diastole, blood volume decreases and light reflection increases." The intuitive explanation behind the theory of substituting heart rate with pulse rate lies in the common physiological origin of the two signals. However, the ECG signal is an electrical voltage signal, and the PPG signal is measured by light reflection or absorption; the maximum values of both signals are related to cardiac systole.

Using a smartphone as a PPG makes the registration more user friendly than previous ECG measurement techniques. It should be mentioned here that there are some solutions for measuring ECG using smartphones [16–18], and these solutions involve additional devices connected to a smartphone via a cable or radio connection. Although these techniques are easy to use, these still require third-party devices and extra cost. When PPG is measured with a smartphone, no external devices and expensive accessories are required. Nowadays, smartphones are widespread and they can be used for various telemonitoring purposes [19]. In this way, we give patients the chance to control their own monitoring and health. The evaluation can be carried out instantly by using new algorithms run by healthcare professionals, which can be accessed from anywhere via the Wi-Fi or 3/4G Internet. During our study, we examined the acceptability of using "stand-alone" smartphone-based PRV registration with a PPG technique in clinical settings instead of the complicated ECG-HRV registration. We created smartphone-based software to measure PRV with high quality. Then, an environment was devised to measure PPG and ECG at the same time for the sake of an accurate validation. Our aim was to develop a PRV measurement technique, which is widely available and can replace the ECG-based HRV measurements.

This paper is organized as follows. In Section 2, we give a detailed description of how the parallel ECG and PPG measurements were taken. In the same section, the indices derived from the ECG and PPG signals are defined, and a commonly used comparison methodology is introduced. In Section 3, we collect our measurement results, and then, we compare the computed PRV and HRV indices using the usual methodology. Afterwards, we introduce an additional validation aspect, which should be taken into consideration in other comparison studies. Essentially, in the previous studies, the PRV was compared with the HRV using just one channel of ECG as the gold standard. However, the HVR indices derived from different ECG channels also show a nonnegligible deviation, and these correspondences among ECG channels are also investigated. The PRV-HRV correspondence is related to the HRV-HRV correspondence. Finally, in Section 4, we draw

some pertinent conclusions and make some suggestions for future work.

2. Methods and Materials

We will describe the validation methodology, the way that ECG and PPG were recorded in parallel with the intention of having an adequate analysis. Next, we introduce the commonly used HRV parameters, for which the correspondences were investigated in other studies. Then, we describe the comparison methods that are commonly applied to investigate the correspondence between PRV and HRV indices.

The main goal of the study outlined here was to develop a measurement tool that can measure the PRV (pulse rate variability) accurately, and this application can be readily used by a layman. All these requirements can be satisfied using a widely accessible tool called a smartphone like an iPhone 6. There were similar developments in the past, using other types of smartphones [12, 13, 20], but the device and implementation only permitted a low-frequency PPG measurement. The PPG measurement was compared to ECG statistics in that literature, and there was found to be a good correspondence between the HRV (heart rate variability) and PRV parameters. The iPhone 6 smartphone supports a 240-frame-per-second (FPS) video recording, the so-called "Slow-mo" video, and based on this feature, our plan was to develop a PPG measurement application with a sampling rate of 240 Hz. Another goal of this study was to compare our PRV measurements with those HRV parameters computed from the gold standard ECG signals and also to investigate our comparison results among other experiment results like those of [12, 13, 21]. Because different comparison methodologies were used in different research studies, we collected many of the HRV feature computation and comparison methods for the purpose of a thorough investigation.

Later, we investigate a question raised during our comparison process. If there is more than one ECG channel, which channel should be treated as a gold standard? If the ECG device measures just one channel, then, can that measurement be accepted as a gold standard? If there is difference between the statistics calculated from the HRV belonging to different ECG channels, then, the HRV-PRV comparison methodology can also be applied on the different channels of the ECG device. With the results, we should be able to characterize the variability between the channels of the ECG device involved in our investigations. One could compare the PRV-HRV correspondence with the HRV(i)-HRV(j) correspondence; however, up till now, we could not find a similar approach in other studies. So we think that the ECG should be treated as a gold standard including the variance analysis among the derived values got from the different channels.

The measurements were collected from 50 people. Two signals were recorded in parallel from each, namely, an ECG signal (multiple channels) and a PPG signal. The subjects of the experiments were presumably healthy young or middle-aged people (39 males, 11 females; mean age: 27 years). The length of the recordings made was 5 minutes, whose duration is standard in several medical examinations [1, 22]. The participants were asked to sit in a relaxed position and not to speak to others while the measurements were being taken to avoid collecting a lot of artifacts.

2.1. Measuring ECG. ECG signals were recorded using a "Cardiax PC-ECG" device. This type of ECG recording device was chosen for several reasons. This device has reusable clamp electrodes, which allow one to record many subjects easily. More importantly, the recorded signal can be easily saved and converted for a further analysis. Many other ECG devices cannot export the recorded data in an appropriate format, the data are stored in a special format, or the data cannot be accessed. The device was connected to the four limbs of the subjects, which allowed us to collect three channels of ECG signals. The sampling frequency of the signals was 500 Hz, and the device filtered the signal with a notch filter (50 Hz), with a high-pass filter (0.01 Hz), and with a low-pass filter (150 Hz).

After collecting the ECG and PPG signals, the same preprocessing steps were performed digitally on all the raw data. Here, we applied a second-order low-pass Butterworth filter with a cut-off frequency of 80 Hz and a second-order high-pass Butterworth filter with a cut-off frequency of 1 Hz. These transformations effectively reduced the noise from high frequencies and slow changes in the signal.

The next step was to find the peaks in the signals. For this purpose, first, a window length was estimated which corresponded to an average RR duration. The estimation was based on finding the first local maximum in the autocorrelation function computed on the signal. Then, with a moving window whose size is slightly larger than that of the estimated average RR interval (e.g., multiplied by 1.3), the maxima were collected in each window, and after filtering out the maxima on the borders, the set of peaks was determined. This method worked well, which is demonstrated by the fact that after a human review of the automatically detected peaks, there were no false or missing peaks found. Figure 2 shows the results of the peak finding method that we applied here.

Although the participants of the experiments were asked to sit in a relaxed position and not to move, some artifacts appeared in each signal, mainly because of movements. This fact is not unique to our study; other researchers have also reported this issue [23, 24]. The usual method for detecting these parts in the signal is to compare all subsequent RR durations with a median duration, and if the absolute difference is higher than a threshold, then, that RR interval is dropped and it is skipped in the later computations. The condition for accepting an RR interval during our experiments was that $\text{MedRR}/1.2 < \text{RR} < \text{MedRR} * 1.2$, where MedRR is the median of all RRs (durations between subsequent peaks). This method is very similar to the artifact filtering techniques mentioned in other studies [24, 25].

2.2. Measuring PPG. To measure the PPG signal, we decided to use an iPhone 6 smartphone. The procedure was, as in other projects [12, 13, 26], that after switching on the flash, the light would go through a finger of the subject in question and with the camera nearby, the adsorption of the light could be measured.

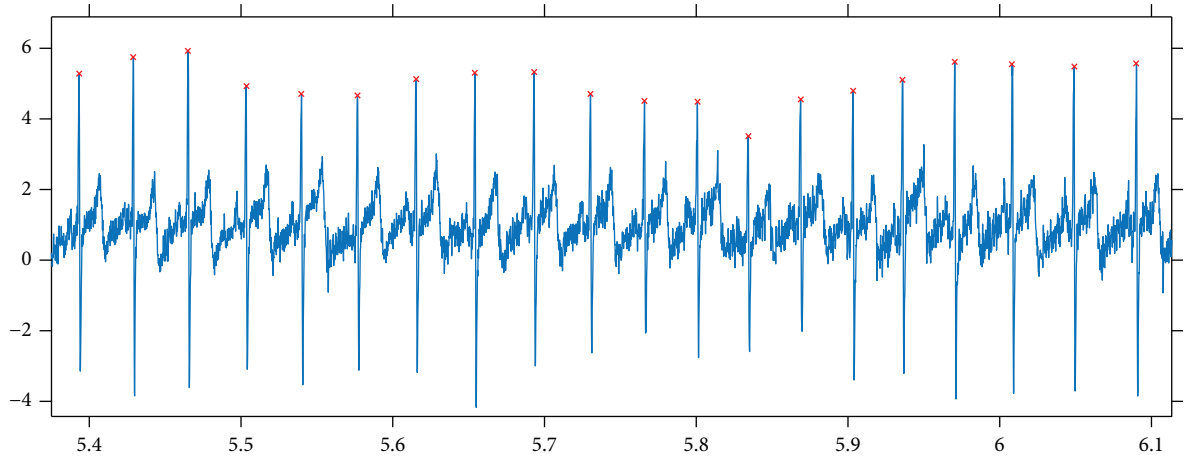


Figure 2: Results got from applying our peak detection method to an ECG signal.

The application was developed in the Swift programming language, which initialized the back camera input for the so-called "Slow-mo" capturing mode (240 Hz, 720 p). A callback method was called when a new image buffer was available with its timestamp, and with this callback method, the PPG signal was computed in real time. In our implementation, the CPU utilization was about 40–50%, while real-time PPG signal production, analysis, and some GUI feedback (signal plots) for the user were carried out.

From the images of the video signal, the luma component was examined (Y component of the supported 420YpCbCr8-BiPlanarFullRange format). In other studies, similar luminosity or brightness data (or just the data of the red channel) were used for computations (in the RGB video recording mode) [27, 28]. It can be seen that these techniques are equivalent, because in the RGB mode, all the blue and green pixel values are zeros and, consequently, all the linear combinations of RGB channels will result in similar curves like those for luminosity. Another technically important fact is that all the automatic functions of the camera can be switched off programmatically (like auto-white balance and autoexposure). Here, the level of the flash ("torch") was set to the maximum.

Unlike that for the ECG signal, here, not just data values but their corresponding timestamps were also available. So it may be interesting and important to investigate the spacing between the timestamps; namely, how much one differs from that of an equally spaced one. Because not just the durations between consecutive peaks (RR or NN intervals) but also their differences will be considered here, a large jump in the duration between timestamps could be a source of error. Fortunately, the durations between the consecutive timestamps have a very small variation. The maximum and minimum differences between consecutive time intervals are of the order of $1e-7$, which means that there is a fairly regular time spacing of the video stream signal.

The next preprocessing steps for the PPG signal were the same as those for ECG, namely, those of low-pass filtering, high-pass filtering, peak detection, and filtering out artifacts from the set of RR intervals.

Figure 3: Experimental arrangement. The subject is sitting in a resting position, the electrodes of the ECG device are connected to the limbs, and the smartphone is held in the subject's palm.

2.3. Parallel Measurements. Many studies already confirm that HRV and PRV parameters, derived from the series of RR and PP durations, are consistent with each other [10, 12, 13]. Our aim here was to investigate this correspondence, when the PPG signal is obtained from the video stream with high frames per second using an iPhone 6. For the sake of a suitable comparison, parallel measurements were made using a standard ECG device and an iPhone 6 smartphone. Figure 3 shows a typical scenario for this. The application developed for the iPhone was designed so that a measurement begins with a 20-second "practice" part, during which the subject can locate his/her finger on the back camera and the flash

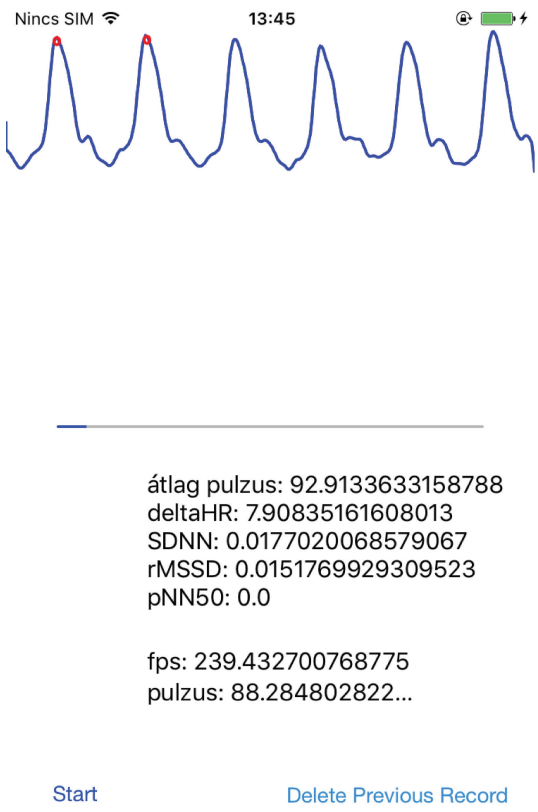


Figure 4: A screenshot of the PPG measurement application during a recording.

appropriately based on feedback (i.e., the filtered PPG signal is shown in real time on the GUI; see Figure 4 for a screenshot). Later, a tone is played, which indicates that the ECG measurement has also to be started. After 5 minutes, a second tone indicates that all the measurements have to be stopped (and the signals must be saved).

When evaluating the parallel measurements, both signals were processed using the methods described above; then, a parallel processing step was performed which attempted to remove from both RR and PP series those values that might correspond to artifacts in at least one of the ECG or PPG signals. For this, after the peaks were detected in both signals, a synchronization step was carried out. Namely, the peak series were paired to each other with a minimal error. Figure 5 shows this synchronization step. The pairing process examined multiple parts taken from both signals to determine the optimal shift value between them, because the artifacts could be anywhere in a signal. Moreover, a time scaling multiplier was calculated, the value of which was very close to 1, since the sampling frequency of the ECG signal was not exactly 500 Hz and the FPS of the video stream was not exactly 240 Hz (actually, it was 239.84 Hz in our experiments).

After this pairing process, RR (and PP) durations corresponding to an artifact in one of the time series were removed. Another filter was applied that deleted RR and PP durations from both series, if they differed by more than 0.3 second.

2.4. Analysis of the Signals. There are many medically relevant parameters which can be derived from the RR series. Some of these parameters are statistical properties of the RR time series, while others characterize the frequency-domain features of the RR data. Some values measure statistical properties of the differences between consecutive RR durations. Figure 6 shows this delta RR series computed on ECG and PPG signals.

When comparing the RR (PP) series got from ECG and PPG signals, the usual way is to compare the derived HRV (PRV) measures [1, 12]. Since one goal of this study was to compare our results with those of other ECG-PPG comparison studies, we computed the measures described in those studies. We collected the definitions of these parameters below (where N is the number of RR durations, RR_i is the ith RR duration in ms, P_i is the corresponding pulse value ($60{,}000/\mathrm{RR}_i$), and $\mathrm{DRR}_i = \mathrm{RR}_{i+1} - \mathrm{RR}_i$). The abbreviations have the following meanings: standard deviation of RR interval time series (SDRR), root mean square of successive differences (RMSSD), and probability of the successive differences of NN (or RR) intervals which differ by more than 50 ms (pNN50).

$$\begin{aligned}
\overline{\mathrm{RR}} &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{RR}_i,\\
\bar{P} &= \frac{1}{N}\sum_{i=1}^{N}p_i,\\
\mathrm{SDRR} &= \mathrm{std}(\mathrm{RR}_i),\\
\mathrm{RMSSD} &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\mathrm{DRR}_i^2},\\
\mathrm{pNN50} &= P(|\mathrm{DRR}_i| > 50\ \mathrm{ms}).
\end{aligned} \tag{1}$$

The definitions of the frequency-domain parameters contain the $f(\lambda)$ function, which is the power spectrum of the RR tachogram. The definitions of the abbreviations are the following. VLF stands for the power in the very low frequency range, LF represents the power in the low frequency range, and HF means the power in the high frequency range.

$$\begin{aligned}
\mathrm{VLF} &= \int_{0.003\ \mathrm{Hz}}^{0.04\ \mathrm{Hz}} f(\lambda)d\lambda,\\
\mathrm{LF} &= \int_{0.04\ \mathrm{Hz}}^{0.15\ \mathrm{Hz}} f(\lambda)d\lambda,\\
\mathrm{HF} &= \int_{0.15\ \mathrm{Hz}}^{0.4\ \mathrm{Hz}} f(\lambda)d\lambda.
\end{aligned} \tag{2}$$

In other studies, some of these parameters had a different name. For example, SDNN is the same as SDRR and the NN duration or the PP duration is equivalent to the RR duration. In different publications, AVNN (average of NN intervals) corresponds to the average RR (AVRR) or average PP (AVPP).

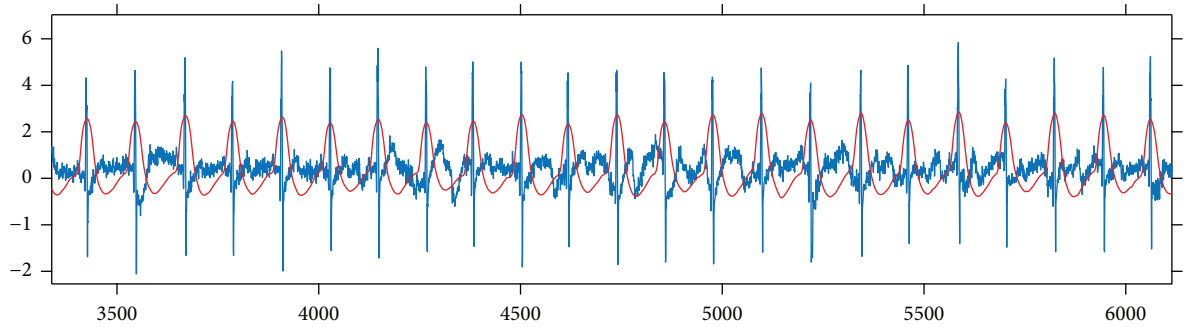


Figure 5: Illustration of the results of the synchronization process, with the ECG signal shown in blue and the PPG signal shown in red.

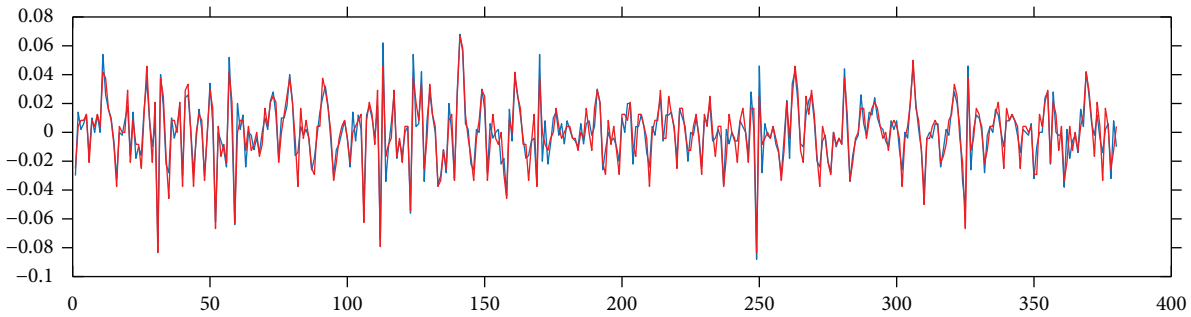


Figure 6: Delta RR duration series computed on ECG (blue) and PPG (red) signals.

2.5. Comparison Methods. Two kinds of comparison methodologies are commonly used in the literature. The first is the Pearson correlation coefficient (given below) with linear regression parameters computed on the two series [29]:

$$\mathrm{PC} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{3}$$

Because this correlation value was always close to the one in the experiments, but the differences of the PPG- and ECG-derived values displayed a clearly visible deviation, a more sophisticated plot and comparison method was introduced, called the Bland-Altman plot and analysis [30, 31]. The mathematical definitions of measurement values are the following:

$$\begin{aligned}
\mathrm{Bias} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i),\\
\mathrm{SD} &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - x_i - \mathrm{Bias})^2},\\
\mathrm{LOA} &= \mathrm{Bias} \pm 1.96\,\mathrm{SD},\\
\mathrm{AL} &= \pm\frac{1}{n}\sum_{i=1}^{n}\frac{y_i + x_i}{2},\\
\mathrm{BAR} &= \frac{1.96\,\mathrm{SD}}{1/n\sum_{i=1}^{n}(y_i + x_i/2)}.
\end{aligned} \tag{4}$$

Here, "bias" means an average shift in the values relating to the reference data (x), and SD denotes the standard deviation of the differences. Limit of agreement (LOA) stands for providing an agreement limit, when the distribution of differences is supposed to be a normal distribution. An acceptance limit (AL) is also introduced [12, 32], which is determined by the scale of the values of the reference and the ones examined (here, all the values are positive). The BAR (Bland-Altman ratio) parameter relates SD to AL, and it has been given a meaning [12, 33] that if the value is at most 10%, then, the agreement is ranked as good, and if the value is above, it is moderate ($10\% < \mathrm{BAR} \leq 20\%$) or insufficient ($\mathrm{BAR} > 20\%$).

Since both methods (correlation and Bland-Altman statistics) were used in different reports, we calculated all these statistical values for characterizing our measurements and for the sake of comparing our findings with those in the other studies.

3. Results and Discussion

Next, we will present our results of all the computed comparison parameters defined above. These parameters will be computed not just for the PPG-ECG signal pairs but also for the ECG channel pairs. Moreover, figures will be included to show the linear relationship between the indices and the Bland-Altman plots. Table 1 and the plots (Figures 7 and 8) show our HRV-PRV comparison results.

3.1. Results of Comparisons among ECG Channels. We mentioned previously that when comparing the parameters derived from PPG with those derived from ECG measurements, the ECG signal is treated to be a gold standard. However, a clinically used ECG device has more than one channel,

Table 1: Comparison values when PRV and HRV indices are compared with each other. Here, PC stands for the Pearson correlation coefficient (with the P values), m and b represent the coefficients for the linear regression on HRV (PRV) with the corresponding mean error (err), R^2 is the coefficient of determination, and bias, SD, and BAR values are the results of the Bland-Altman analysis. The definitions of the HRV indices were introduced earlier.

	PC	P	lin. m	lin. b	lin. err (MSE)	lin. R^2	Bias	SD	BAR
HR (beat/min)	1	$<10^{-23}$	1.00	−0.12	0.011	1	0.032	0.110	<0.001
Mean RR(ms)	1	$<10^{-23}$	1.00	−0.02	0	1	−0.002	0.009	<0.001
RMSSD (ms)	0.996	$<10^{-23}$	1.00	2.53	3.15	0.992	2.464	1.793	0.106
ln(RMSSD)	0.973	$<10^{-23}$	0.87	0.528	0.017	0.947	0.103	0.153	0.089
SDNN (ms)	0.999	$<10^{-23}$	1.01	1.06	0.582	0.998	1.271	0.776	0.035
pNN50 (%)	0.993	$<10^{-23}$	1.07	1.67	4.399	0.987	2.673	2.432	0.306
TP (total power, ms^2)	0.998	$<10^{-23}$	1.04	16.52	1439.3	0.997	50.15	46.05	0.100
LP (lf power, ms^2)	0.999	$<10^{-23}$	1.04	−1.14	86.63	0.999	15.96	16.03	0.089
HP (hf power, ms^2)	0.995	$<10^{-23}$	1.06	18.51	783.9	0.991	33.87	33.18	0.246
LP + HP	0.998	$<10^{-23}$	1.06	13.14	1124.4	0.996	49.83	45.21	0.144
LP/HP	0.941	$<10^{-23}$	0.68	0.341	0.326	0.885	−0.529	0.937	0.736

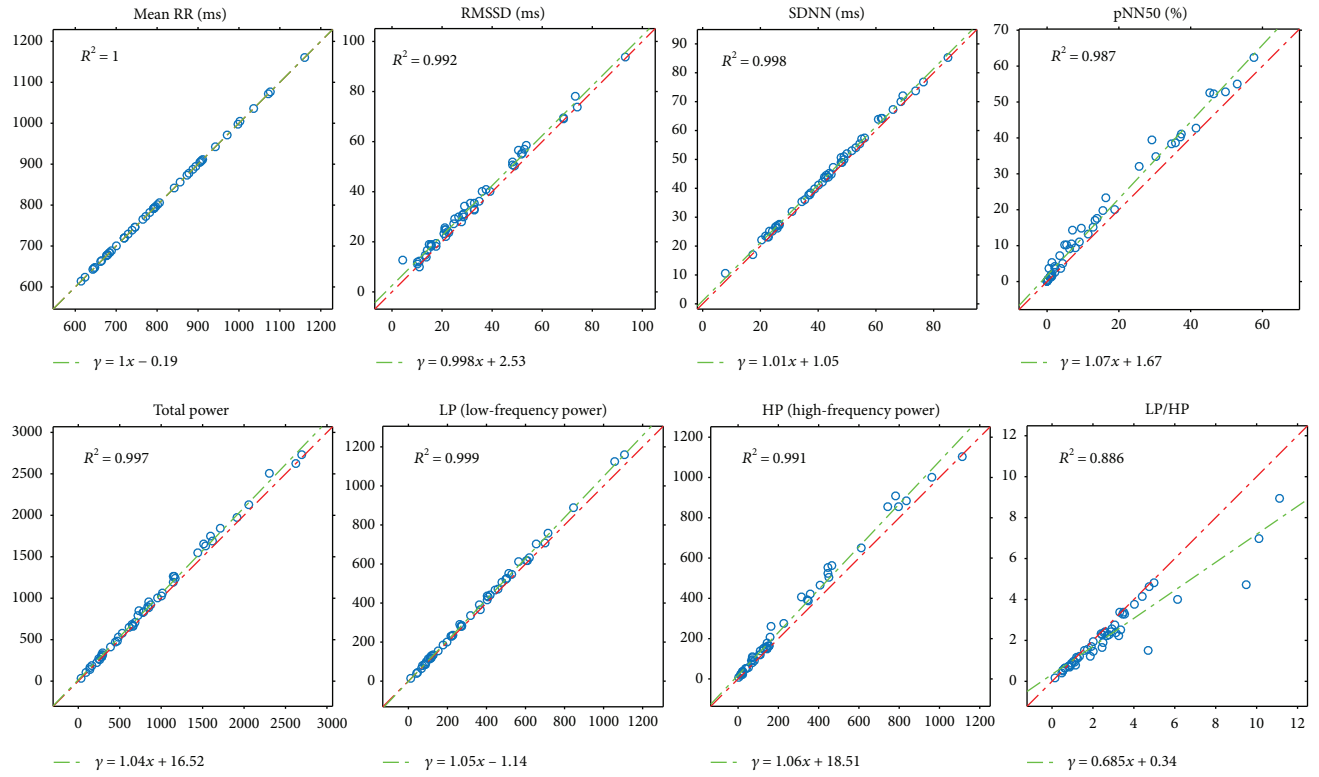


Figure 7: Plots of PRV indices related to HRV indices (horizontal axis) with R^2 and linear regression.

and the question arises, of which channel should be used as the basis of a comparison process. Moreover, what if, when comparing the HRV indices corresponding to different ECG channels with each other, we have similar properties, like when we compare HRV with PRV?

In the experiments, a Cardiax PC-ECG device was used that had four electrodes connected to the four limbs of the participants. This resulted in three channel data. Figures 9 and 10 show the same plots for the ECG(1)-ECG(2), as those for ECG-PPG (Figures 7 and 8). Figures 11 and 12 and Figures 13 and 14 show these results for the ECG(1)-ECG(3) and ECG(2)-ECG(3), respectively. Some key values are highlighted in the figures, and the relevant ones are listed in Table 2 in Discussion.

4. Discussion

Next, we will examine other studies to determine the position of our results relative to these. Furthermore, we will discuss the point that the ECG channels differ from each other, and this means that in an ECG-PPG comparison, this should be taken into account.

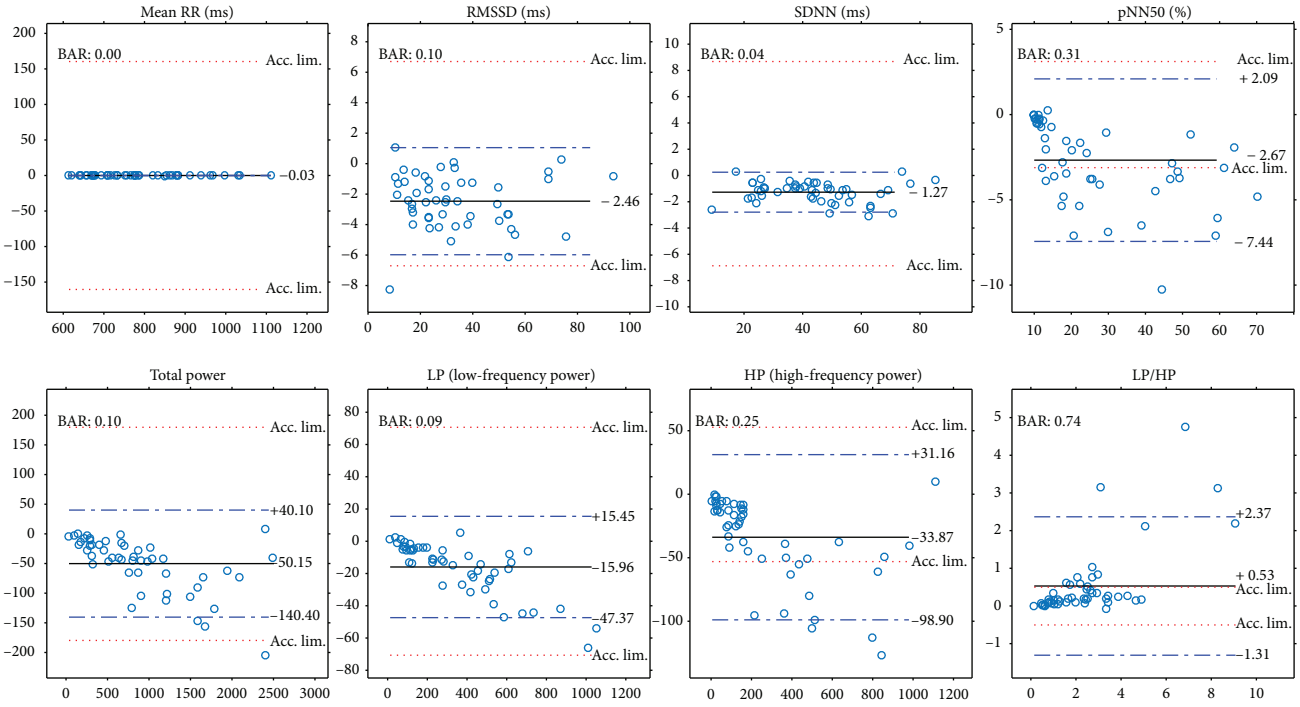


Figure 8: Bland-Altman plots for PRV and HRV indices with limits of agreement (blue dashed lines), bias (black lines), and acceptance limits (red dotted lines).

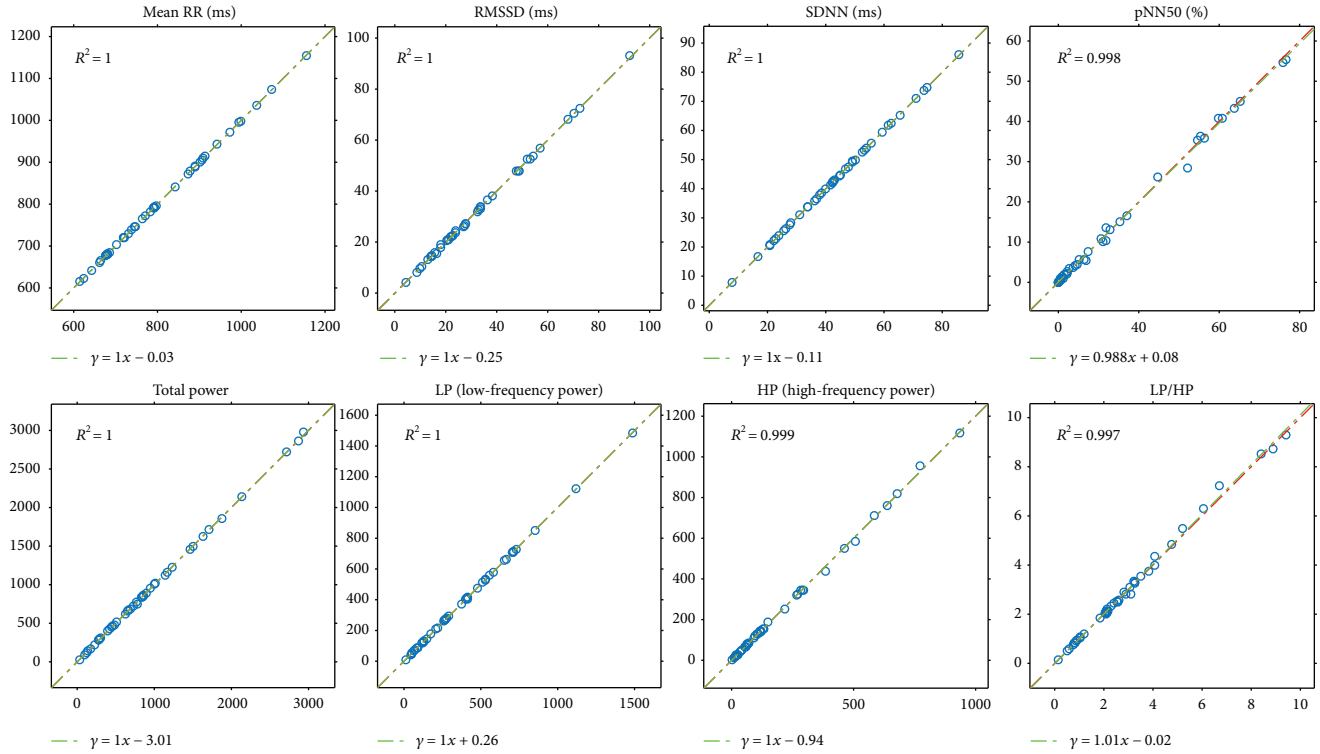


Figure 9: Plots of HRV indices calculated for ECG channel 1 and ECG channel 2 with R^2 and linear regression.

4.1. Our Result in Itself. Our results reveal a good correspondence between most indices of HRV and PRV (see Table 1 and Figures 7 and 8). Most of the correlations are above 0.99, and ln(RMSSD) and TP/HP have slightly lower correlation values. What is more, the Bland-Altman analysis also provides good results. The agreement is insufficient

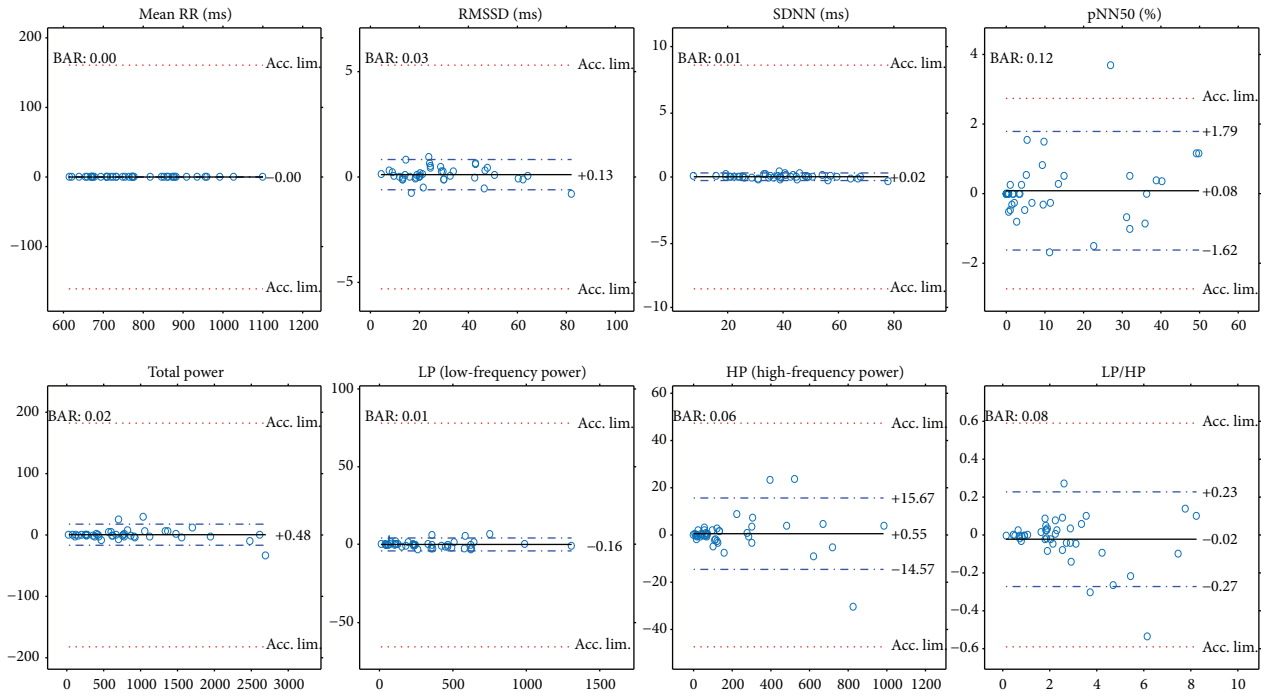


Figure 10: Bland-Altman plots of HRV indices calculated for ECG channel 1 and ECG channel 2 with limits of agreement (blue dashed lines), bias (black lines), and acceptance limits (red dotted lines).

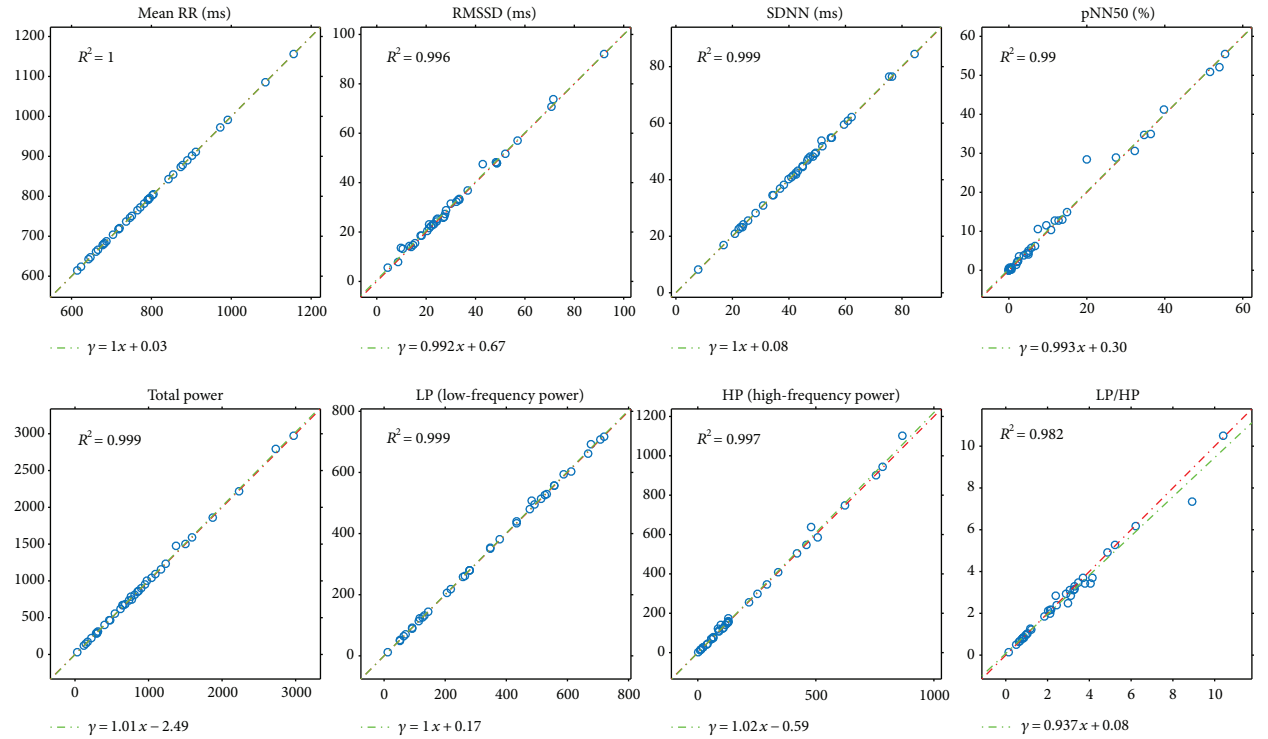


Figure 11: Plots of HRV indices calculated for ECG channel 1 and ECG channel 3 with R^2 and linear regression.

(BAR > 20%) just for pNN50, HP, and HP/LP parameters. The reason for this is the high bias, which is probably due to the influence of breathing on the high-frequency PRV components.

4.2. Comparison with Smartphone-Based PRV-HRV Correspondence Measurements. The authors of various studies have reported comparison results between the analyses of ECG and PPG signals. Among these studies, there are a few

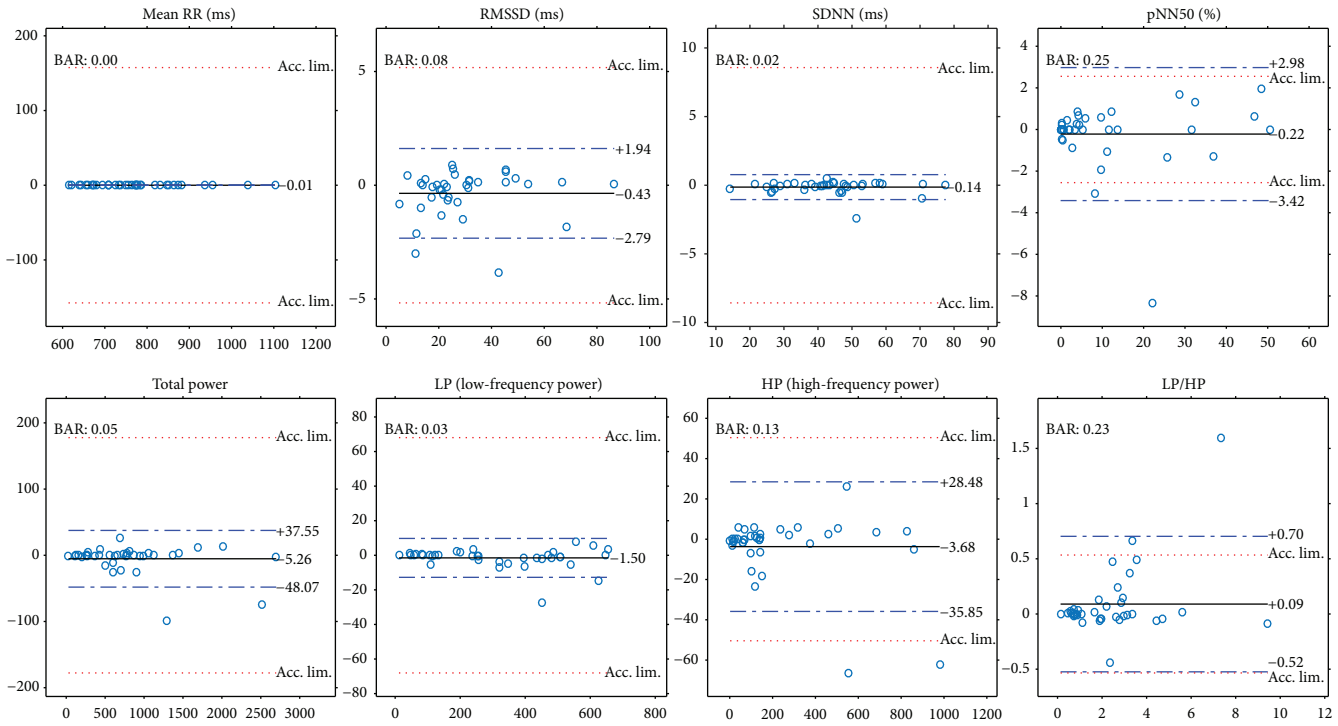


FIGURE 12: Bland-Altman plots of HRV indices calculated for ECG channel 1 and ECG channel 3 with limits of agreement (blue dashed lines), bias (black lines), and acceptance limits (red dotted lines).

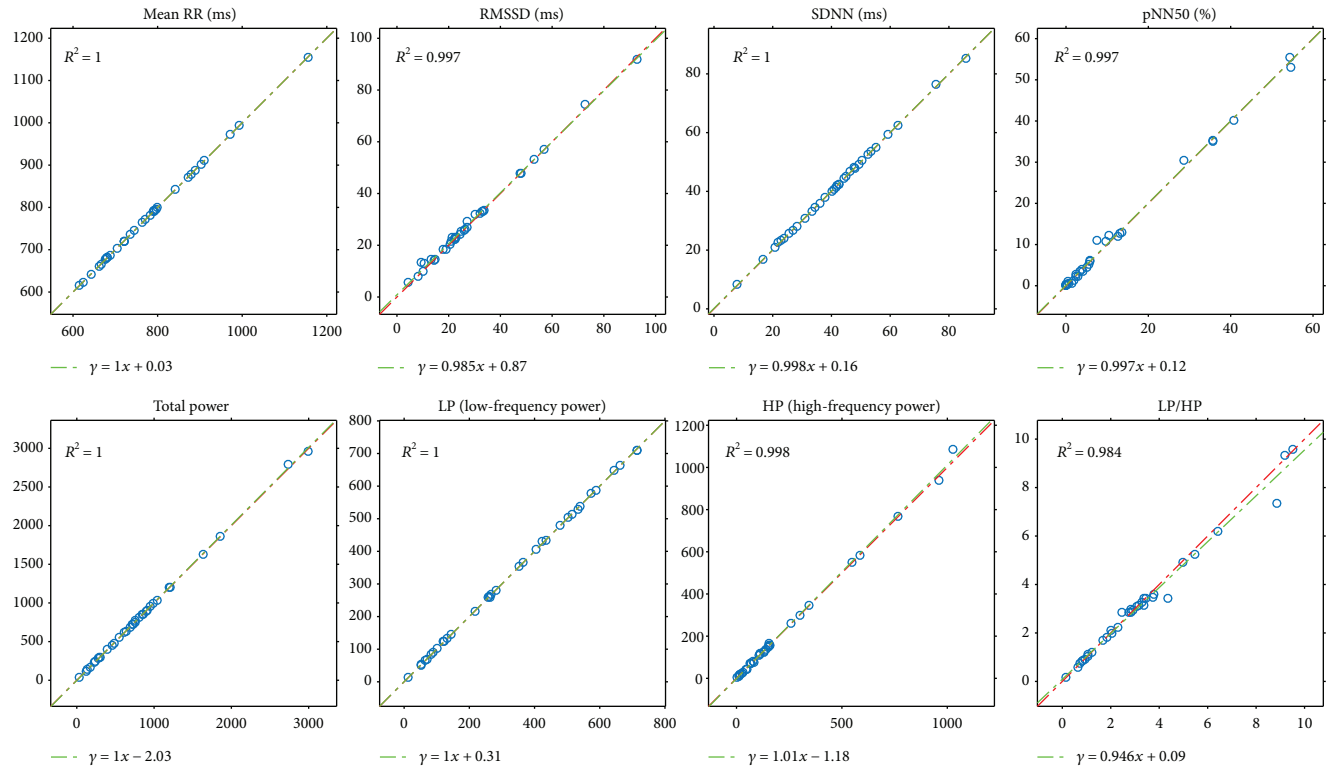


FIGURE 13: Plots of HRV indices calculated for ECG channel 2 and ECG channel 3 with R^2 and linear regression.

reports that describe measurements of the PPG signal using a smartphone. In a study [12], the authors used an HTC S510e smartphone to take PPG measurements (20–30 FPS) and a Finometer MIDI as an ECG data acquisition tool (200 Hz). The number of participants was 30, and the duration of the recordings was at least 5 minutes. They found a perfect

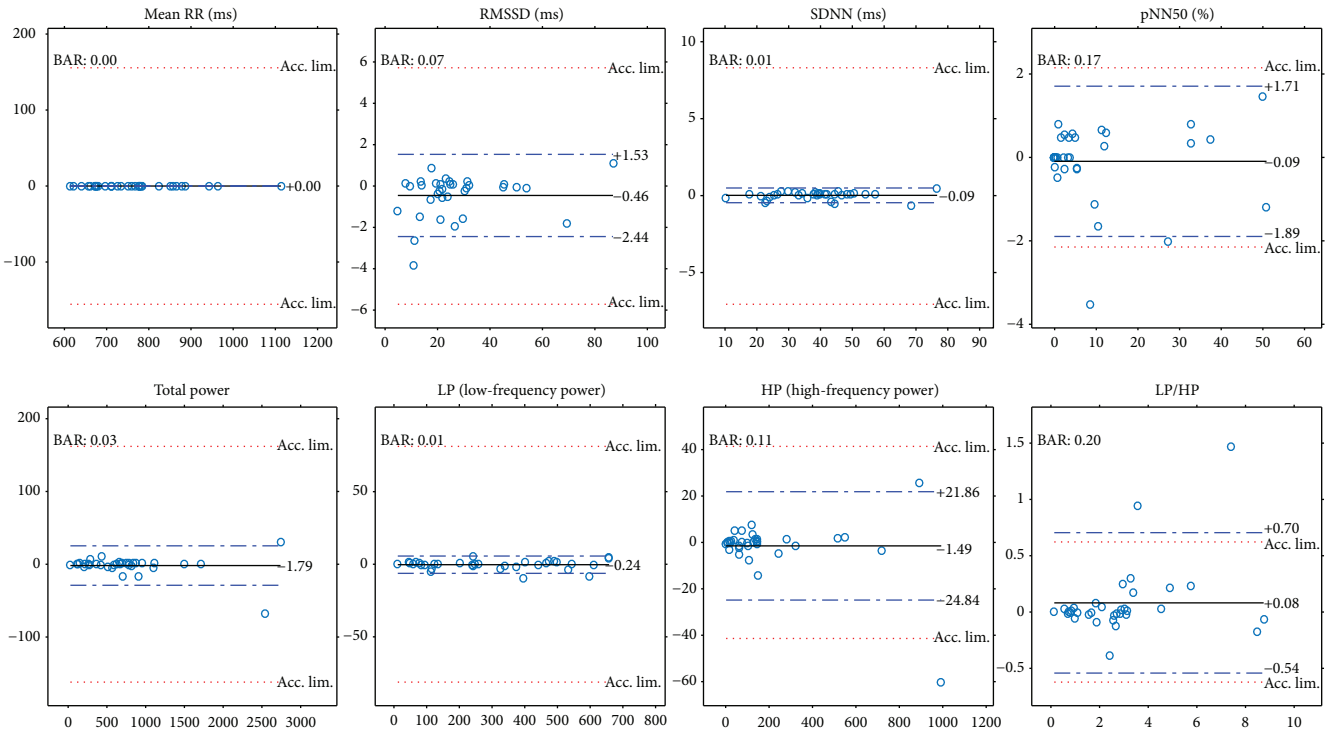


Figure 14: Bland-Altman plots of HRV indices calculated for ECG channel 2 and ECG channel 3 with limits of agreement (blue dashed lines), bias (black lines), and acceptance limits (red dotted lines).

Table 2: HRV and PRV index comparisons can be found in the corresponding literature (in parentheses). PC stands for the Pearson correlation coefficient (with the P values), and bias, SD, and BAR values are the results of the Bland-Altman analysis. The definitions of the HRV indices were introduced by us earlier.

Derived index comparison	Cited value	Our value
SDNN-PC ([12])	0.933	**0.999**
SDNN-BAR ([12])	19.17%	**3.5**%
RMSSD-PC ([12])	0.78	**0.996**
RMSSD-BAR ([12])	42.22%	**10.6**%
pNN50-PC ([12])	0.5	**0.993**
pNN50-BAR ([12])	79.91%	**30.6**%
LP-PC ([12])	0.996	**0.999**
LP-BAR ([12])	12.14%	**8.9**%
HP-PC ([12])	**0.996**	0.995
HP-BAR ([12])	**10.22**%	25.6%
LP/HP-PC ([12])	**0.982**	0.941
LP/HP-BAR ([12])	**19.3**%	73.6%
avg(PP)-PC ([13])	0.993	**1**
avg(PP)-bias ([13])	−0.05 beats/min	**0.032** beats/min
avg(PP)-SD ([13])	1.03 beats/min	**0.11** beats/min
ln(RMSSD)-bias ([21])	0.94	**0.103**
ln(RMSSD)-SD ([21])	1.77	**0.153**

correlation for just the AVNN time-domain parameter, but other correlations between time-domain indices were 0.933, 0.78, and 0.5 for the SDNN, RMSSD, and PNN50 indices, respectively. Our results for these correlations are 0.999, 0.996, and 0.993, respectively, which are much better results. The linear regression parameters display a much greater difference between the indices than those in our findings, which are summarized in Table 1. Surprisingly in the frequency domain, their HRV and PRV indices correlate better, but in the case of 5-minute measurements, the VLF power (power in the very low frequency range, 0.003–0.04 Hz) computation is not very useful (the authors gave this value in their study). The Bland-Altman analysis revealed similar findings in the time domain (they got worse results than ours) and in the frequency domain, as well. For example, their BAR value for SDNN is 19.17%; for RMSSD, 42.22%; and for PNN50, 79.91%, while our corresponding values are 3.5%, 10.6%, and 30.6%, respectively.

Another study [13] reported an experiment using iPhone 6 for PPG and a 12-lead ECG treadmill (GE Series 2000, GE Medical Systems Information Technologies Inc., Milwaukee, WI, USA) for HR measurements. They compared just the accuracy of heart rate estimates got from the two kinds of signals. In a resting position situation, they found a 0.993 correlation with a mean difference of −0.05 beats/min and a standard deviation of 1.03 beats/min. Our corresponding values for these parameters are 1 for the correlation, 0.032 beats/min for the bias, and 0.11 beats/min for the standard deviation.

In a third experiment [21], 30 participants were involved in measuring their ECG and PPG in parallel, using a Biopac ECG and an iPad2 combined with an infrared pulse sensor (ithlete™). They compared indices computed from ultra-short-term signals (of approximately one minute in length).

Table 3: Comparison values for the correspondence values among PRV and HRV indices, when comparing PPG to ECG, and the channels of ECG. Here, PC stands for the Pearson correlation coefficient (with P values smaller than 10^{-10}), m and b represent the coefficients for the linear regression, and bias, SD, and BAR values are for the Bland-Altman analysis. The worst values are in bold.

		ECG-PPG	ECG(1)-ECG(2)	ECG(1)-ECG(3)	ECG(2)-ECG(3)
Mean RR (ms)	PC	1	1	1	1
	m	1	1	1	1
	b	−0.019	−0.025	0.028	**0.032**
	Bias	−0.021	0.0015	**0.007**	−0.003
	SD	0.0093	0.021	**0.061**	0.026
	BAR	<0.001	<0.001	<0.001	<0.001
RMSSD (ms)	PC	0.996	1	**0.998**	0.999
	m	0.998	1.004	0.99	0.99
	b	2.53	−0.25	**0.67**	0.87
	Bias	2.64	−0.13	0.43	**0.46**
	SD	1.79	0.44	**1.21**	1.01
	BAR	0.104	0.027	**0.076**	0.07
SDNN (ms)	PC	0.998	1	1	1
	m	1.005	**1.002**	1.001	0.998
	b	1.055	−0.113	0.075	**0.162**
	Bias	1.27	−0.020	**0.136**	0.092
	SD	0.776	0.148	**0.464**	0.260
	BAR	0.035	0.0068	**0.021**	0.012
pNN50 (%)	PC	0.993	0.999	**0.995**	0.998
	m	1.07	0.988	**0.993**	0.997
	b	1.67	0.076	**0.305**	0.122
	Bias	2.67	−0.083	**0.221**	0.094
	SD	2.43	0.869	**1.631**	0.918
	BAR	0.306	0.125	**0.250**	0.168
TP (total power, ms^2)	PC	0.998	1	1	1
	m	1.04	1.003	**1.009**	1.005
	b	16.52	**−3.01**	−2.49	−2.032
	Bias	50.15	−0.484	**5.261**	1.788
	SD	46.15	8.768	**21.84**	13.83
	BAR	0.10	0.019	**0.048**	0.0335
LP (low-frequency power, ms^2)	PC	1	1	1	1
	m	1.05	1	**1.004**	1
	b	−1.14	0.26	0.167	**0.307**
	Bias	15.96	0.16	**1.50**	0.243
	SD	16.03	2.26	**5.74**	2.43
	BAR	0.089	0.012	**0.033**	0.015
HP (high-frequency power, ms^2)	PC	0.995	1	0.999	0.999
	m	1.06	1.002	**1.017**	1.013
	b	18.514	−0.94	−0.586	**−1.18**
	Bias	33.87	−0.55	**3.68**	1.49
	SD	33.18	7.71	**16.41**	11.91
	BAR	0.247	0.064	**0.13**	0.11

They got a bias of 0.94 and a standard deviation of 1.77 on the ln(RMSSD) index differences (when the measurements were taken in a seated position). These values are higher than ours (0.103, 0.153), which indicate a significantly worse results.

In Table 2, we collected all the data that could be accessed in previous publications on the topic of comparing smart device-based PPG measurements with ECG. The better values are shown in bold.

Table 4: Bland-Altman SD and BAR values for the ECG-PPG and the worst ECG(i)-ECG(j) correspondences.

		ECG-PPG	Worst ECG(i)-ECG(j)	(1)/(2)
Mean RR (ms)	SD	0.0093	0.061	0.15
	BAR	<0.001	<0.001	n/a
RMSSD (ms)	SD	1.79	1.21	1.48
	BAR	0.104	0.076	1.37
SDNN (ms)	SD	0.776	0.464	1.67
	BAR	0.035	0.021	1.67
pNN50 (%)	SD	2.43	1.631	1.49
	BAR	0.306	0.250	1.22
TP (total power, ms^2)	SD	46.15	21.84	2.11
	BAR	0.10	0.048	2.08
LP (low-frequency power, ms^2)	SD	16.03	5.74	2.79
	BAR	0.089	0.033	2.70
HP (high-frequency power, ms^2)	SD	33.18	16.41	2.02
	BAR	0.247	0.13	1.9

In Table 2, the results are an order of magnitude better when our measurements are compared to those in [13] or in [21]. Our results are significantly better compared to those of some important parameters examined in [12], but our high-frequency-domain parameters (HP, LP/HP) are much worse. We do not know the precise reason for this; perhaps, the authors of [12] described a special regulated breathing for the subjects of the experiments. The kind of breathing (spontaneous or regulated) during the experiments can influence the high-frequency-domain power spectrum.

4.3. Our Results in Relation to ECG-ECG Correspondence. Another topic in this study was not just to compare the parameters computed from a PPG signal with those computed from an ECG channel but also to investigate those values related to an ECG-to-ECG channel comparison. The results given in the previous sections (Figures 9–14) tell us that the HRV parameters (or indices) computed via an analysis of an ECG channel differ from each other for different channels. In our experiments here, the ECG(1)-ECG(2) channel comparison had the lowest standard deviation values on Bland-Altman difference plots, and the ECG(1)-ECG(3) differences were the highest. In Table 3, we list the correspondences for the most important indices for the ECG-PPG and the three ECG-ECG comparisons.

The results indicate a good agreement for the parameters mean RR, SDNN, TP, and LP. There is a moderate agreement for RMSSD in the PRV-HRV comparison, but the BAR value is not much higher than that for the ECG(1)-ECG(3) comparison. The agreements are insufficient for the PNN50 and HP values (PPG-ECG), but these are also insufficient in the ECG comparisons. In the PPG-ECG comparison, a significant bias was found for some HRV indices, which are not given in ECG(i)-ECG(j) comparisons. This means that PNN50 and the spectral parameters (TP, LP, and HP) are overestimated, especially when the reference values are large. This phenomenon is clearly visible in Figure 8. Other studies also mention this fact (for references, see [9]). In the latter study, the authors offer an explanation for this observation: "The fact that spontaneous breathing rates usually lie within the HF frequency band explains why many studies found that PRV overestimates HRV mostly in the HF domain or in variables reflecting short-term variability (HF, RMSSD, pNN50, etc.)."

In order to summarize the most important analysis values of the Bland-Altman method, we collected the SD (standard deviation) and BAR (Bland-Altman ratio) values for the various HRV indices corresponding to the ECG(i)-ECG(j) and ECG-PPG comparisons. From the ECG(i)-ECG(j) values, the worst were taken (which are in bold in Table 3). We also computed the ratio of the ECG-ECG and ECG-PPG values. Table 4 contains data concerning this comparison.

Earlier, we found that there is a significant bias between some ECG- and PPG-based variability indices. Table 4 tells us that for the time-domain indices, the standard deviation of the differences (SD) and the Bland-Altman ratio (BAR) corresponding to PPG indices are at most two times higher than those corresponding to ECG. This factor is slightly above two for the frequency-domain indices. We think that this correspondence between the HRV and PRV should suffice for an application if we wish to collect PRV data from a larger group worldwide.

5. Conclusions

In order to achieve our main goal, one of the first steps was to compare PRV with HRV. Our results indicate that almost all the indices computed from PRV may be an alternative to those computed from HRV, even for clinical use. This may be concluded from the results of our comparison among the PRV-HRV correspondences and HRV-HRV correspondences. However, there are some indices which show a bias related to the values computed from an HRV analysis (mainly pNN50 and high-frequency power). This phenomenon corresponding to biases was found in other earlier studies as well [9], so one might think that with some direct (possibly) linear transformation or by taking into account the rhythm of breathing, there should be a way to minimize the errors between the two kinds of rate variability indices.

In the future, we plan to validate PRV measured using a smartphone with HRV involving CAD (coronary artery disease) patients. Moreover, we are interested in whether there is any medical reason which explains the variability among the derived indices computed from different ECG channels. Also a short-term goal of ours is to make our smartphone application free to the public and get as many people involved in data collection as possible.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

The authors would like to thank Gábor Sipka and Tibor Szabó for their valuable contribution on the development of the previous version of this method. They are also grateful to Professors Ferenc Bari, Ph.D., D.Sc, and József Tolnai, Ph.D., for their support and for the Cardiax ECG device. This project GINOP-2.2.1-15-2017-00073 (Tigra) was supported by the European Union and cofinanced by the European Regional Development Fund.

References

[1] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996.

[2] R. Sassi, S. Cerutti, F. Lombardi et al., "Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society," *EP Europace*, vol. 17, no. 9, pp. 1341–1353, 2015.

[3] W. H. Organization, *World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals*, World Health Organization, Geneva, Switzerland, 2016.

[4] K. C. Bilchick and R. D. Berger, "Heart rate variability," *Journal of Cardiovascular Electrophysiology*, vol. 17, no. 6, pp. 691–694, 2006.

[5] M. Sosnowski, E. Clark, S. Latif, P. W. Macfarlane, and M. Tendera, "Heart rate variability fraction—a new reportable measure of 24-hour R-R interval variation," *Annals of Noninvasive Electrocardiology*, vol. 10, no. 1, pp. 7–15, 2005.

[6] B. Francesco, B. Maria Grazia, G. Emanuele et al., "Linear and nonlinear heart rate variability indexes in clinical practice," *Computational and Mathematical Methods in Medicine*, vol. 2012, pp. 1–5, 2012.

[7] J. Bolea, E. Pueyo, M. Orini, and R. Bailón, "Influence of heart rate in non-linear HRV indices as a sampling rate effect evaluated on supine and standing," *Frontiers in Physiology*, vol. 7, 2016.

[8] A. K. F. da Silva, D. G. D. Christofaro, A. F. B. Bernardo, F. M. Vanderlei, and L. C. M. Vanderlei, "Sensitivity, specificity and predictive value of heart rate variability indices in type 1 diabetes mellitus," *Arquivos Brasileiros de Cardiologia*, vol. 108, no. 3, pp. 255–262, 2017.

[9] A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram," *International Journal of Cardiology*, vol. 166, no. 1, pp. 15–29, 2013.

[10] V. Jeyhani, S. Mahdiani, M. Peltokangas, and A. Vehkaoja, "Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5952–5955, Milan, Italy, 2015.

[11] G. Leikan, E. Rossi, M. Sanz et al., "Evaluation of agreement between temporal series obtained from electrocardiogram and pulse wave," *Journal of Physics: Conference Series*, vol. 705, article 012038, 2016.

[12] R.-C. Peng, X.-L. Zhou, W.-H. Lin, and Y.-T. Zhang, "Extraction of heart rate variability from smartphone photoplethysmograms," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 516826, 11 pages, 2015.

[13] B. P. Yan, C. K. Chan, C. K. Li et al., "Resting and postexercise heart rate detection from fingertip and facial photoplethysmography using a smartphone camera: a validation study," *JMIR mHealth and uHealth*, vol. 5, no. 3, article e33, 2017.

[14] T. Nagy and Z. Gingl, "Low-cost photoplethysmograph solutions using the Raspberry Pi," in *2013 IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI)*, pp. 163–167, Budapest, Hungary, 19 November 2013.

[15] S. Bagha and L. Shaw, "A real time analysis of PPG signal for measurement of SpO_2 and pulse rate," *International Journal of Computer Applications*, vol. 36, no. 11, pp. 45–50, 2011.

[16] C. D. Galloway, D. E. Albert, and S. B. Freedman, "iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke," *International Journal of Cardiology*, vol. 165, pp. 193-194, 2013.

[17] J. B. Muhlestein, V. Le, D. Albert et al., "Smartphone ECG for evaluation of STEMI: results of the ST LEUIS Pilot Study," *Journal of Electrocardiology*, vol. 48, no. 2, pp. 249–259, 2015.

[18] L. Desteghe, Z. Raymaekers, M. Lutin et al., "Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting," *EP Europace*, vol. 19, no. 1, pp. 29–39, 2016.

[19] M. Kay, J. Santos, and M. Takane, "mHealth: new horizons for health through mobile technologies," *World Health Organization*, Geneva, Switzerland, 2011.

[20] J. A. Heathers, "Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research," *International Journal of Psychophysiology*, vol. 89, no. 3, pp. 297–304, 2013.

[21] M. R. Esco, A. A. Flatt, and F. Y. Nakamura, "Agreement between a smartphone pulse sensor application and electrocardiography for determining lnRMSSD," *Journal of Strength and Conditioning Research*, vol. 31, no. 2, pp. 1–385, 2017.

[22] G. H. Lin, Y. H. Chang, and K. P. Lin, "Comparison of heart rate variability measured by ECG in different signal lengths," *Journal of Medical and Biological Engineering*, vol. 25, no. 2, pp. 67–71, 2005.

[23] M. A. Peltola, "Role of editing of R–R intervals in the analysis of heart rate variability," *Frontiers in Physiology*, vol. 3, 2012.

[24] C. A. G. Martínez, A. O. Quintana, X. A. Vila et al., "Loading, plotting, and filtering RR intervals," in *Heart Rate Variability*

Analysis with the R Package RHRV. Use R!, pp. 15–28, Springer, Cham, 2017.

[25] G. D. Clifford, *Signal processing methods for heart rate variability, [Ph.D. thesis]*, University of Oxford, 2002.

[26] L. Drijkoningen, F. Lenaerts, J. Van der Auwera et al., "Validation of a smartphone based photoplethysmographic beat detection algorithm for normal and ectopic complexes," in *Computing in Cardiology 2014*, pp. 845–848, Cambridge, MA, USA, 2014.

[27] K. Matsumura, P. Rolfe, J. Lee, and T. Yamakoshi, "iPhone 4s photoplethysmography: which light color yields the most accurate heart rate and normalized pulse volume using the iPhysioMeter application in the presence of motion artifact?," *PLoS One*, vol. 9, no. 3, article e91205, 2014.

[28] S. A. Siddiqui, Y. Zhang, Z. Feng, and A. Kos, "A pulse rate estimation algorithm using PPG and smartphone camera," *Journal of Medical Systems*, vol. 40, no. 5, p. 126, 2016.

[29] M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.

[30] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[31] D. Giavarina, "Understanding Bland Altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.

[32] P. S. Myles and J. Cui, "I. using the Bland–Altman method to measure agreement with repeated measures," *British Journal of Anaesthesia*, vol. 99, no. 3, pp. 309–311, 2007.

[33] D. Stöckl, D. R. Cabaleiro, K. V. Uytfanghe, and L. M. Thienpont, "Interpreting method comparison studies by use of the Bland–Altman plot: reflecting the importance of sample size by incorporating confidence limits and predefined error limits in the graphic," *Clinical Chemistry*, vol. 50, no. 11, pp. 2216–2218, 2004.