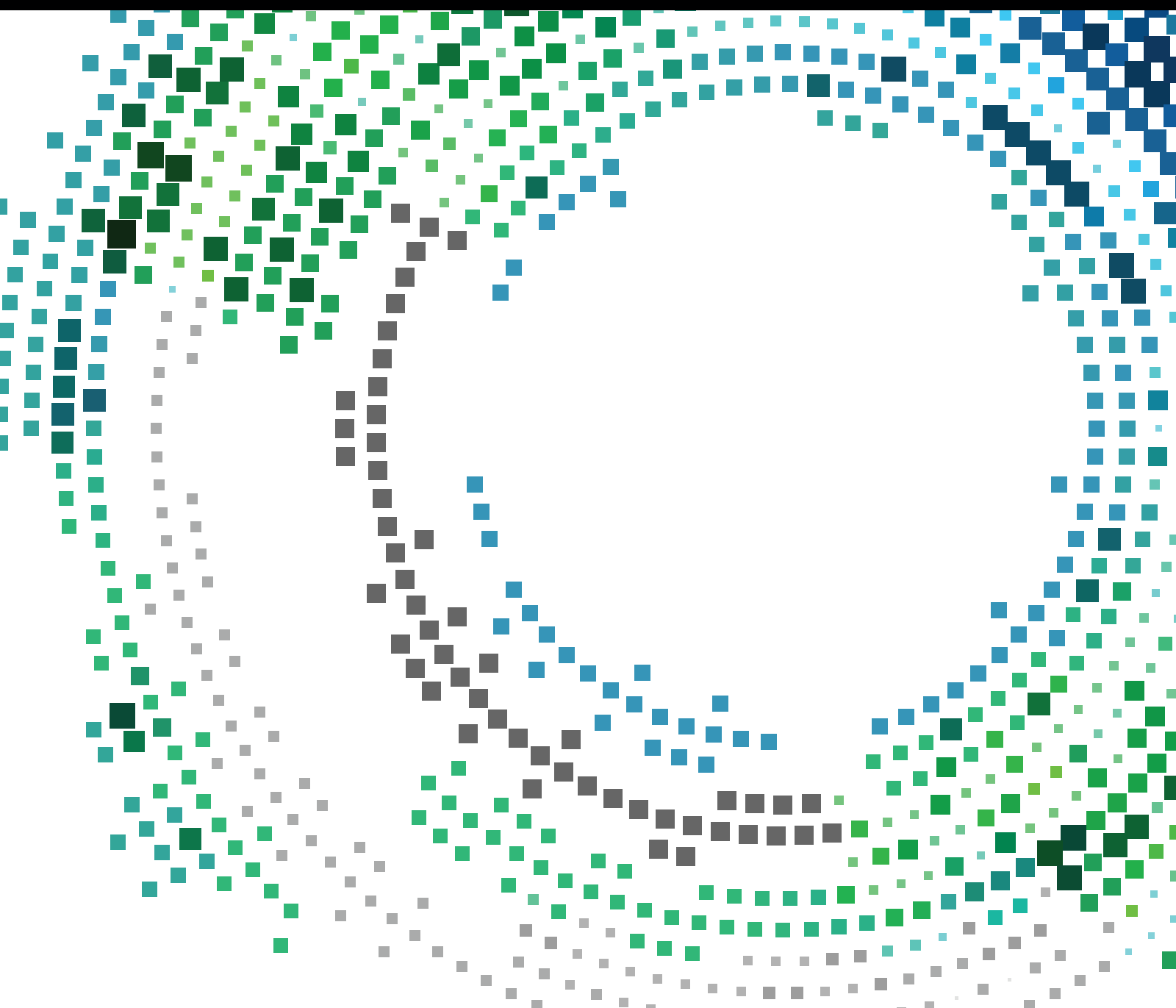


Artificial Intelligence for Mobile Health Data Analysis and Processing

Lead Guest Editor: Giovanna Sannino

Guest Editors: Nizar Bouguila, Giuseppe De Pietro, and Antonio Celesti





Artificial Intelligence for Mobile Health Data Analysis and Processing

Artificial Intelligence for Mobile Health Data Analysis and Processing

Lead Guest Editor: Giovanna Sannino

Guest Editors: Nizar Bouguila, Giuseppe De Pietro,
and Antonio Celesti



Copyright © 2019 Hindawi. All rights reserved.

This is a special issue published in “Mobile Information Systems.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Mari C. Aguayo Torres, Spain
Ramon Agüero, Spain
Markos Anastassopoulos, UK
Marco Anisetti, Italy
Claudio Agostino Ardagna, Italy
Jose M. Barcelo-Ordinas, Spain
Alessandro Bazzi, Italy
Luca Bedogni, Italy
Paolo Bellavista, Italy
Nicola Bicocchi, Italy
Peter Brida, Slovakia
Carlos T. Calafate, Spain
María Calderon, Spain
Juan C. Cano, Spain
Salvatore Carta, Italy
Yuh-Shyan Chen, Taiwan
Wenchi Cheng, China
Massimo Condoluci, Sweden
Antonio de la Oliva, Spain
Almudena Díaz Zayas, Spain

Filippo Gandino, Italy
Jorge Garcia Duque, Spain
L. J. García Villalba, Spain
Michele Garetto, Italy
Romeo Giuliano, Italy
Prosanta Gope, UK
Javier Gozalvez, Spain
Francesco Gringoli, Italy
Carlos A. Gutierrez, Mexico
Ravi Jhawar, Luxembourg
Peter Jung, Germany
Adrian Kliks, Poland
Dik Lun Lee, Hong Kong
Ding Li, USA
Juraj Machaj, Slovakia
Sergio Mascetti, Italy
Elio Masciari, Italy
Maristella Matera, Italy
Franco Mazzenga, Italy
Eduardo Mena, Spain

Massimo Merro, Italy
Jose F. Monserrat, Spain
Raul Montoliu, Spain
Mario Muñoz-Organero, Spain
Francesco Palmieri, Italy
José J. Pazos-Arias, Spain
Marco Picone, Italy
Vicent Pla, Spain
Amon Rapp, Italy
Daniele Riboni, Italy
Pedro M. Ruiz, Spain
Michele Ruta, Italy
Stefania Sardellitti, Italy
Filippo Sciarrone, Italy
Floriano Scioscia, Italy
Michael Vassilakopoulos, Greece
Laurence T. Yang, Canada
Jinglan Zhang, Australia

Contents

Artificial Intelligence for Mobile Health Data Analysis and Processing

Giovanna Sannino , Nizar Bouguila , Giuseppe De Pietro, and Antonio Celesti
Editorial (2 pages), Article ID 2673463, Volume 2019 (2019)

Computer-Assisted Diagnosis for Diabetic Retinopathy Based on Fundus Images Using Deep Convolutional Neural Network

Yung-Hui Li , Nai-Ning Yeh, Shih-Jen Chen , and Yu-Chien Chung
Research Article (14 pages), Article ID 6142839, Volume 2019 (2019)



ProMe: A Mentoring Platform for Older Adults Using Machine Learning Techniques for Supporting the “Live and Learn” Concept

Giorgos Kostopoulos, Katja Neureiter, Dragos Papatoiu, Manfred Tscheligi, and Christos Chrysoulas 
Research Article (8 pages), Article ID 9723268, Volume 2018 (2019)

User Evaluation of the Smartphone Screen Reader VoiceOver with Visually Disabled Participants

Berglind F. Smaradottir , Jarle A. Håland , and Santiago G. Martinez 
Research Article (9 pages), Article ID 6941631, Volume 2018 (2019)


Mobile Hardware-Information System for Neuro-Electrostimulation

Vladimir S. Kublanov , Mikhail V. Babich , and Anton Yu. Dolganov 
Research Article (7 pages), Article ID 2168307, Volume 2018 (2019)

WearableDL: Wearable Internet-of-Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future



Aras R. Dargazany , Paolo Stegagno , and Kunal Mankodiya 
Review Article (20 pages), Article ID 8125126, Volume 2018 (2019)

Artificial Intelligence to Prevent Mobile Heart Failure Patients Decompensation in Real Time: Monitoring-Based Predictive Model

Nekane Larburu , Arkaitz Artetxe, Vanessa Escolar, Ainara Lozano, and Jon Kerexeta
Research Article (11 pages), Article ID 1546210, Volume 2018 (2019)

Editorial

Artificial Intelligence for Mobile Health Data Analysis and Processing

Giovanna Sannino ¹, **Nizar Bouguila** ², **Giuseppe De Pietro**¹ and **Antonio Celesti**³

¹National Research Council of Italy (CNR), Naples, Italy

²Concordia Institute for Information Systems Engineering, Montreal, Canada

³University of Messina, Messina, Italy

Correspondence should be addressed to Giovanna Sannino; giovanna.sannino@icar.cnr.it

Received 30 December 2018; Accepted 31 December 2018; Published 22 January 2019

Copyright © 2019 Giovanna Sannino et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This special issue is devoted to the application of artificial intelligence for mobile health data analysis and processing. All papers in this issue went through the standard reviewing process of *Mobile Information Systems*.

This special issue is motivated by recent advances in mobile health. Indeed, Internet-of-things (IoT) is changing eHealth and especially mobile Health (m-Health) systems. Currently, more and more fixed and mobile medical devices are installed in patients' personal body networks and medical devices, and the surrounding clinical/home environments collect and send a huge amount of heterogeneous health data to healthcare information systems for their analysis. In this context, machine learning and data mining techniques are becoming extremely important in many real-life problems. Many of these techniques have been developed for health data processing and analysis on mobile devices. Several mobile applications based on these techniques have emerged as an essential technology for improving the quality of medical diagnosis and treatments of many illnesses as well as many health disorders.

Existing techniques used for processing health data can be broadly classified into two categories: (a) non-Artificial Intelligence (AI) systems and (b) Artificial Intelligence systems. Even though non-AI techniques are less complex in nature, most of the systems suffer from the drawbacks of inaccuracy and lack of convergence. Hence, these systems are generally replaced by AI-based systems which are much superior to the conventional systems. AI techniques are mostly hybrid in nature and include artificial neural

networks (ANNs), fuzzy theory, and evolutionary algorithms. Although most of the techniques are theoretically sound, their potential is not fully explored for practical applications. Many of the computational applications still depend on non-AI systems, which limit their practical usage.

The goal of this special issue is to present some applications of machine learning and data mining techniques on practical mobile health applications.

In N. Larburu et al.'s paper, the authors describe and analyze an approach to prevent mobile heart failure patients decompensation in real time via a monitoring-based predictive model. The proposed approach is based on mobile clinical data of 242 heart failure patients collected for a period of 44 months in the public health service of Basque Country (Osakidetza). The authors obtained the best predictive model as a combination of alerts based on monitoring data and a questionnaire with a Naïve Bayes classifier deploying Bernoulli distribution. This predictive model is shown to reduce significantly the false alerts.

In L. Yung-Hui et al.'s paper, the authors propose a computer-assisted diagnosis approach for diabetic retinopathy based on fundus images using a deep convolutional neural network (DCNN). Unlike the traditional DCNN approach, the authors replace the commonly used max-pooling layers with fractional max pooling. Using their approach, the authors achieved a recognition rate up to 86.17% which outperforms the state of the art. With the developed technique, the authors developed an app called "Deep Retina." Equipped with this app and a handheld

ophthalmoscope, an average person can take fundus image and obtain immediate result, calculated by the developed algorithm.

In their manuscript, V. S. Kublanov et al. describe organizational principles of a mobile hardware-informational system based on a multifactorial neuroelectrostimulation device. The system is implemented with two blocks. The first one forms spatially distributed field of low-frequency monopolar pulses between two multielement electrodes in the neck region and the second one which constitutes the specialized control interface performed by a smartphone.

In G. Kostopoulos et al.'s paper, the authors describe a mentoring platform for older adults developed using machine learning techniques to support the "live and learn" concept. This platform called "ProMe" supports different opportunities for informal communication through a number of functionalities such as video, text-chat, e-mail, blogs, and forums. It provides also to end users the opportunity to take different kinds of mentoring roles.

B. F. Smaradottir et al. present in their manuscript a user evaluation of VoiceOver, a built-in screen reader in Apple Inc. products, with a detailed analysis of the gesture interaction, familiarity and training by visually disabled users, and system response. The evaluation was done by six participants with prescribed visual disability in a usability laboratory under controlled conditions. The data collected via this evaluation have been analyzed using a mixed methods approach based on quantitative and qualitative measures.

In A. R. Dargazany et al.'s paper, the authors introduce the concept of wearable deep learning (WearableDL) which is a unifying conceptual architecture inspired by the human nervous system, offering the convergence of deep learning, Internet-of-things, and wearable technologies. In the proposed architecture, the brain represents deep learning for cloud computing and big data processing, the spinal cord represents Internet-of-things for fog computing and big data transfer, and the peripheral sensory and motor nerves represent wearable technologies as edge devices for big data collection.

We would like to thank the authors for their contributions and the reviewers of the papers for their help in bringing this issue to its current form.

Conflicts of Interest

The editors declare that there are no conflicts of interest regarding the publication of this special issue.

*Giovanna Sannino
Nizar Bouguila
Giuseppe De Pietro
Antonio Celesti*

Research Article

Computer-Assisted Diagnosis for Diabetic Retinopathy Based on Fundus Images Using Deep Convolutional Neural Network

Yung-Hui Li ¹, Nai-Ning Yeh,¹ Shih-Jen Chen ², and Yu-Chien Chung³

¹Department of Computer Science & Information Engineering, National Central University, Taoyuan 32001, Taiwan

²Department of Ophthalmology, Taipei Veterans General Hospital, School of Medicine, National Yang-Ming University, Beitou, Taipei, Taiwan

³Department of Ophthalmology, Fu Jen Catholic University Hospital, New Taipei City, Taiwan

Correspondence should be addressed to Yung-Hui Li; li.yunhui@icloud.com

Received 27 April 2018; Revised 28 June 2018; Accepted 27 November 2018; Published 2 January 2019

Guest Editor: Antonio Celesti

Copyright © 2019 Yung-Hui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diabetic retinopathy (DR) is a complication of long-standing diabetes, which is hard to detect in its early stage because it only shows a few symptoms. Nowadays, the diagnosis of DR usually requires taking digital fundus images, as well as images using optical coherence tomography (OCT). Since OCT equipment is very expensive, it will benefit both the patients and the ophthalmologists if an accurate diagnosis can be made, based solely on reading digital fundus images. In the paper, we present a novel algorithm based on deep convolutional neural network (DCNN). Unlike the traditional DCNN approach, we replace the commonly used max-pooling layers with fractional max-pooling. Two of these DCNNs with a different number of layers are trained to derive more discriminative features for classification. After combining features from metadata of the image and DCNNs, we train a support vector machine (SVM) classifier to learn the underlying boundary of distributions of each class. For the experiments, we used the publicly available DR detection database provided by Kaggle. We used 34,124 training images and 1,000 validation images to build our model and tested with 53,572 testing images. The proposed DR classifier classifies the stages of DR into five categories, labeled with an integer ranging between zero and four. The experimental results show that the proposed method can achieve a recognition rate up to 86.17%, which is higher than previously reported in the literature. In addition to designing a machine learning algorithm, we also develop an app called “Deep Retina.” Equipped with a handheld ophthalmoscope, the average person can take fundus images by themselves and obtain an immediate result, calculated by our algorithm. It is beneficial for home care, remote medical care, and self-examination.

1. Introduction

The global cost of treating adult diabetes and its induced chronic complications is USD 850 billion in 2017. Diabetic retinopathy (DR) is one of the most common and serious complications of diabetes mellitus and is a leading cause of low vision and blindness in working-age adults [1, 2]. The International Diabetes Foundation (IDF) estimated that the global population with diabetes in 2017 was 451 million and over one-third of the population had DR [3], representing a tremendous population at risk of visual impairment or blindness. By 2045, the worldwide prevalence of diabetes is expected to increase to 693 million people [3]. In addition, almost half (49.7%) of all people living with diabetes remain

undiagnosed for years because of silent symptoms [3]. However, long-term high blood sugar levels ultimately destroy blood vessels and nerves, leading to complications, such as cardiovascular disease and blindness. Detection and treatment of DR in the early stage will prevent its development or progression.

The diagnosis and severity of DR are based on retinal examination. Clinically, the classification of DR can be divided into two categories: (1) nonproliferative diabetic retinopathy (NPDR) with exudation and ischemia in different severity but without retinal neovascularization, and (2) proliferative diabetic retinopathy (PDR), which is characterized by neovascularization with or without its complications of traditional retinal detachment and the

initial appearance of vitreous hemorrhage. Microvascular diseases of NPDR include microaneurysms, retinal dot and blot hemorrhages, lipid exudates, venous beading change, and intraretinal microvascular abnormalities (IRMA). Based on the degree and extent of these lesions, NPDR can be divided into three levels: mild NPDR presents with microaneurysms or few retinal hemorrhages; moderate NPDR shows more severe microaneurysms, hemorrhage or soft exudate, but not reaching the level of severe NPDR, which is associated with marked retinal hemorrhage in 4 quadrants, venous beading in at least 2 quadrants and IRMA in at least 1 quadrant. Table 1 summarizes the DR category with its manifestation.

Manual grading by ophthalmologists has been the mainstay of DR screening in the past decades. However, due to the expanding population with diabetes and the recent advances in technology, automated detection of DR offers the potential to provide an efficient and cost-effective approach to screening. Current commercialized automated retinal image analysis systems (ARIAs), such as iGradingM, Retmarker, and EyeArt, focus on differentiating diseased/no disease, or detection of referable DR [5, 6]. Nonetheless, ARIAs are currently not sufficiently sophisticated to classify different levels of DR, which means that identifying the subtle change between levels is still a challenging task for the technique of medical image analysis. Figure 1 shows example fundus images for each lesion.

In addition to the accuracy of medical image processing, the mobility and portability of medical examination equipment are of equal importance. Currently, the acquisition of digital fundus images requires the cooperating patient to sit in front of the fundus camera in the room, with ambient lighting minimized or turned off. The patient needs to look forward at the camera at a fixed light and use infrared fundus imaging to focus on the area of interest. Many nonmydriatic cameras have software that automatically detects the posterior pole of the eye and takes a picture when it is focused behind the eye. The RGB image sensor still requires a flash to capture images in the visible light spectrum. However, the digital fundus imagers most popularly used in the clinics are bulky and expensive, as shown in Figure 2, which limit its capability for large-scale screening.

One of the major goals for this study, besides increasing the classification accuracy using artificial intelligence, is to come up with a new system framework for DR screening. The new framework combines the advantages of mobile computing, cloud computing, big data, and artificial intelligence. The components of the proposed framework can be described as following:

- (i) Mobile block: The fundus image acquisition is achieved using a hand-held fundus imager, coupled with a self-developed iPhone APP. The imager is small and light-weighted. It can be carried inside a backpack. The deployment of such devices is extremely convenient. The portable nature due to its small form factors and light-weight can benefit the medical service for remote rural areas.

- (ii) Cloud block: The proposed system does not sacrifice its computational performance for its portability. Thanks to the architecture of cloud computing, the core of the computational resources is moved to the cloud and can be scaled up flexibly as the request increases. We developed highly efficient deep learning algorithm which runs on cloud server and is able to respond to the diagnosis request within 10 seconds.

- (iii) Big data block: The cloud-based architecture also helps to collect big data. As more and more end devices (hand-held fundus camera) are being used, the number of fundus images that passed into the cloud will increase accordingly. By storing all of these fundus image data, we are able to make good use of such big dataset, such as machine learning model retraining, new feature exploration, or cross-domain data-mining for different types of ophthalmological diseases.

In summary, in this paper, we propose a new system framework for DR screening, based on artificial intelligence, mobile computing, cloud computing, and big data analytics. Figure 3 shows an illustration of the proposed system. Such system is a new paradigm for telemedical service and will benefit rural areas where the medical resources are insufficient.

In the following sections, we will gradually unveil our ideas and show the experimental results. In Section 2, we performed related literature review for important algorithms for the foundation of DR classification, which is retinal vessel segmentation. In Section 3, we performed related literature review for DR detection. In Section 4, we illustrated the proposed deep learning and machine learning algorithms in full details. We show the experimental results in Section 5 and discuss some important findings in Section 6. Finally, a conclusion is given in Section 7.

2. Literature Review of Retina Vessels Segmentation

In the process of identifying DR, it is pivotal to locate the retinal vessels. If the vessels position can be correctly known, we can determine whether the patient is suffering from DR based on information about the precise location and thickness of the vessels. However, vessel tracking is a complex process because of the many other substances besides vessels in fundus images. Numerous vessel segmentation methods have been proposed, which can be broadly divided into five categories: vascular tracking, matched filtering, morphological processing, deformation models, and machine learning.

2.1. Methods of Vascular Tracking. Methods of vascular tracking are based on the continuous structure of vessels, by starting at an initial point and following the vessels until no further vessels are found. The critical factor in this procedure is the setting of the initial point, as this will affect the

TABLE 1: Classification of diabetic retinopathy [4].

Category	Level	Manifestation
Nonproliferative diabetic retinopathy (NPDR)	1	Few microaneurysms or hemorrhage
	2	Microaneurysms, hemorrhages, soft exudates, venous beading
	3	Severe retinal hemorrhage in ≥ 4 quadrants, or venous beading change in ≥ 2 quadrants, or mild intraretinal microvascular abnormality in ≥ 1 quadrant
Proliferative diabetic retinopathy (PDR)	4	Neovascularization of retina or optic disc

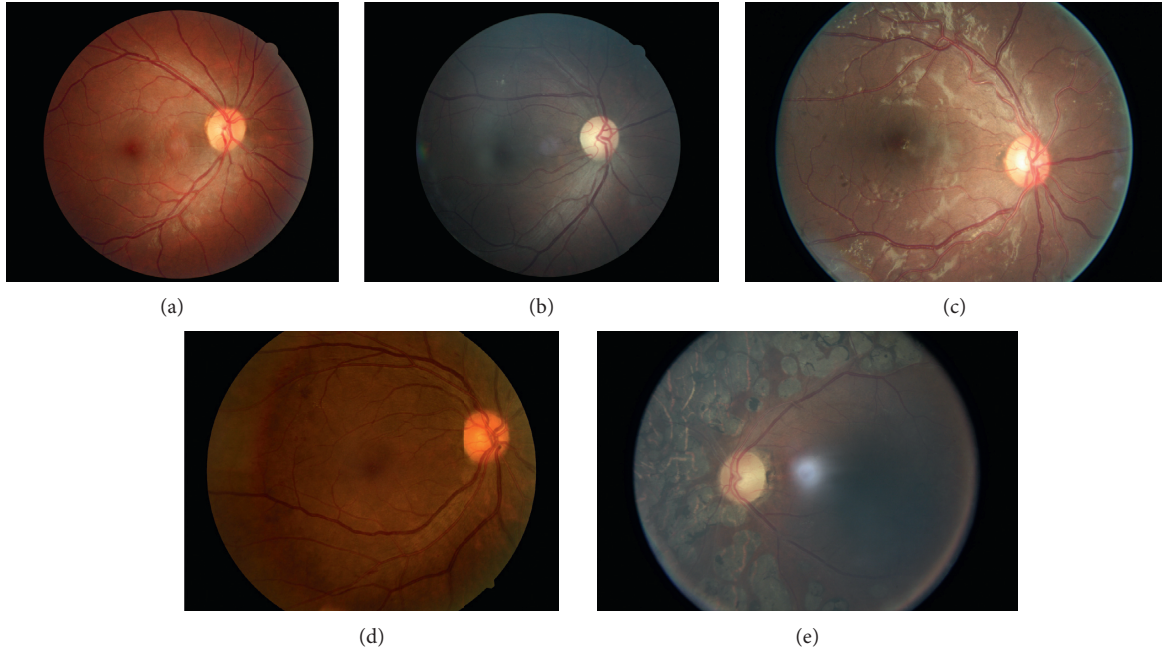


FIGURE 1: Examples of fundus images showing different lesions. (a) 0 level (normal), (b) level 1, (c) level 2, (d) level 3, and (e) level 4.

accuracy of vessel segmentation. Currently, setting the initial point can be done either artificially or automatically.

The earliest adaptive vascular tracking method was proposed by Liu and Sun [7] in 1993, which extracts the vasculature from X-ray angiograms. First, given an initial point and direction within a vessel, the authors apply an “extrapolation-update” scheme that involves estimating local vessel trajectories. Once a vessel fragment has been tracked, it is removed from the image. This procedure is repeated until the vascular tree has been extracted. The drawbacks of this strategy are that due to the algorithm used, the user must set the vessel starting points and that the approach does not seem adaptable to three-dimensional extraction. In 1999, an automatic vascular tracking method was developed by Can et al. [8]. This strategy mainly collects pixel wide vascular local minimum points (usually in the middle of a vessel) to perform tracking. Vlachos and Dermatas [9] suggested a multiscale line tracking method with morphological postprocessing. Yin et al. [10] proposed a retinal vascular tree extraction, based on iterative tracking and Bayesian method.

The advantage of vascular tracking is that it can provide local information about characteristics, such as the

diameter/width and direction of vessels. However, the vascular tracking performance can be easily affected by crossing or branching of vessels, which reduces the identification efficiency.

2.2. Methods of Matched Filtering. Matched filtering methods employ multiple matched filters for extraction, so designing proper filters is essential to detect vessels. Since the gray-scale distribution of fundus vessels is in keeping with Gaussian, an intuitive method exists that uses the maximum response of images after filtering to find vessel points. As the diameter/width of vessels is diverse, a multiscale Gaussian filter method is often used for vessel tracking.

In 1989, Chaudhuri et al. [11] pioneered the application of Gaussian filters in vessel tracking, by using some vascular characteristics, such as the fact that vessels are darker than the background, the width of the vessels ranges from 2 to 10 pixels, and the vessels grow from the optic disc into a radial shape. Therefore, Chaudhuri et al. [11] designed two-dimensional Gaussian filters that can detect vessels in 12 different directions. However, this method needs large computation, and some of the dark lesions are similar to the



FIGURE 2: Commercialized fundus camera used in the clinics. The machine is bulky and expensive, making it difficult for large-scale screening.

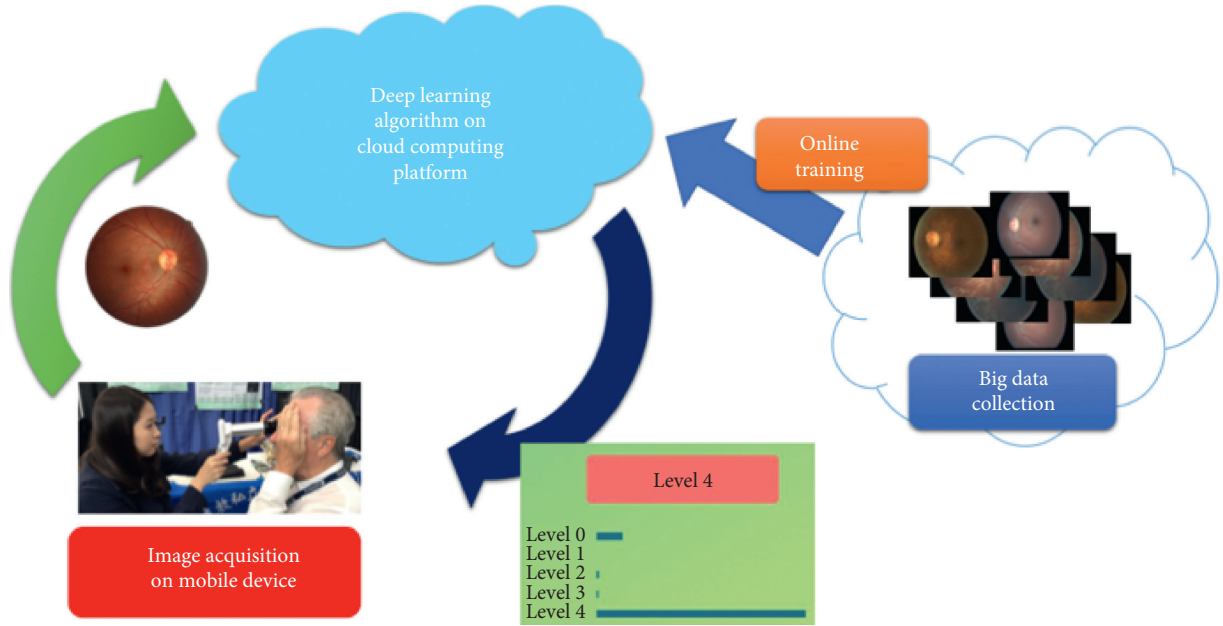


FIGURE 3: The proposed framework. It is composed of mobile block (image acquisition), cloud block (deep learning algorithm on cloud computing platform) and big data block (big data collection and training).

characteristic of vessels, causing tracking errors. Hoover et al. [12] described an improved method that considers local and regional characteristics of vessels to separate blood vessels in retinal images and iteratively determine whether the current point is a vessel point.

After such improvement, a large number of studies of reformed filters have been developed. Jiang and Mojon [13] promoted a generalized threshold method based on a multithreshold detection. Zhang et al. [14] improved the matching filtering method by applying a local vessel cross section analysis, using local bilateral thresholding. Li et al. [15] suggested a multiscale production of the matched filter, to enhance the extraction of tiny vessels.

2.3. Methods of Morphological Processing. Morphological processing facilitates the segmentation and identification of target objects by analyzing and processing structural elements in a binary image. Thus, linear and circular elements of blood vessels can be selected, isolating the desired structure instead of the background image. In addition,

morphological processing can also smooth and fill the image contour with the advantage of antinoise. However, this method overrelies on structural elements and does not make good use of characteristics of vessels.

According to vessel characteristics, Zana and Klein [16] introduced a mathematical morphology-based algorithm that allows separating the vessels from all possible undesirable patterns. Building on this approach, Ayala et al. [17] proposed using different average fuzzy sets. In Miri and Mahloojifar [18], fundus images were analyzed by the use of curvelet transform and morphological reconstruction of multistructural elements to enhance the boundaries and determine the vascular ridge. Karthika and Marimuthu [19] combined curvelet transform and morphological reconstruction of multistructural elements, with strongly connected component analysis (SCCA) to segment and identify vessels.

2.4. Methods of Deformation Models. First introduced by Kass et al. [20] in 1988, the key benefit of deformation

models is the ability to produce smooth parametric curves or surfaces. Two categories of deformation models are identified: parametric deformation and geometric deformation. Parametric deformation models are also called active contour or snake models (set of points each with an associated energy). Through the external and internal forces acting on the snake, the snake model can change its shape and smoothness toward the desired structure. In 2007, Espona et al. [21] used a parametric deformation model method on fundus images and promoted an improved method with morphological segmentation. With the assistance of morphological vessel segmentation, the snake model expands to the contour of the obtained vessels until the local energy function is minimal. Another deformation model method called ribbon of twins (ROT), which combines ribbon snakes and double snakes, was proposed by Al-Diri and Hunter [22]. Each twin consists of two snakes, one inside and one outside the vessel edges. The double snake model then attempts to integrate the pairs of twins on the vessel borders into a single ribbon and calculate the vessel width.

There are several shortcomings in parametric deformation models. For instance, the segmentation results depend on the initial contour, and difficulties arise when extending from low to high dimensions and in segmenting complex objects. Geometric deformation can well solve the problems caused by parametric deformable models. Geometric deformable models are based on deformation curve evolution theory and have no strict requirement on the position of the initial contour, which increases the robustness of the method and allows it to be extended to high dimensions. Zhang et al. [23] proposed an automatic vessel segmentation method, which uses nonlinear orthogonal projection to capture the characteristics of retinal blood vessels and obtained an adaptive local thresholding algorithm for blood vessel segmentation. Zhao et al. [24] suggested a retinal vessel segmentation method that employs a region-based active contour model with a level set implementation and a region growing model.

2.5. Methods of Deformation Model. Machine learning is an algorithm that teaches computers to learn to achieve goals automatically, by building generative or discriminative models from accumulated datasets. Machine learning can be divided into supervised learning and unsupervised learning. The supervised learning methods learn to achieve goals based on ground-truth, which means that during the training stage, the training data used to train the model come with a “label” that can be used by the machine learning algorithm to differentiate the data. Applying such paradigm in the problem of DR, it means that when using supervised learning, one needs to mark all of the pixels belonging to vessels in advance, whereas the unsupervised learning method does not need to mark them beforehand.

For supervised learning, Cesar and Jelinek [30] and Leandro et al. [31] proposed a supervised classification with two-dimensional Gabor wavelet. Each pixel has a feature vector that consists of the gray-scale feature and responses of distinct sizes of two-dimensional Gabor wavelet. Ricci and

Perfetti [25] proposed a segmentation method for retinal vessels based on online manipulation and support vector classification. Since the features are extracted by two orthogonal vertical lines, it reduces the features and training samples in supervised learning. A supervised method using neural network was proposed by Marin et al. [26], which has one input layer, three hidden layers, and one output layer. Each pixel in the image is represented by a seven-dimensional feature vector to train the network. Shanmugam and Banu [27] used an extreme learning machine (ELM) to detect retinal vessels by creating a seven-dimensional feature vector based on gray-scale features and invariant moments and using ELM to segment vessels. In 2015, Wang et al. [28] raised a new hierarchical retinal vascular segmentation, including three steps: preprocessing, hierarchical feature extraction, and integration classification. It involves using simple linear iterative clustering (SLIC) to perform super-pixel segmentation and randomly selecting a pixel to represent the entire super-pixel, as a more easy and efficient means of extracting features.

For unsupervised learning, in 1998, Tolia and Panas [32] created an automatic and unsupervised segmentation method based on blurred fundus images, which used fuzzy C-means (FCM) to find initial candidate points. Xie and Nie [33] proposed a segmentation method based on a genetic algorithm and FCM. Salazar-Gonzalez et al. [29] used methods of vector flow to segment retinal vessels.

Table 2 is a summary about the performance comparison between different existing methods.

3. Literature Review of Diabetic Retinopathy Detection

Although extracting vessels before detecting DR with machine learning can achieve high accuracy, it is time-consuming to create the marked ground-truth for retinal vessels. Another paradigm is to train the computer to automatically learn how to distinguish levels of DR by reading retinal images directly, without performing vessel segmentation. In 2000, Ege et al. [34] proposed an automatic analysis of DR by different statistical classifiers, including Bayesian, Mahalanobis, and k-nearest neighbor. Silberman et al. [35] introduced an automatic detection system for DR and reported an equal error rate of 87%. Karegowda et al. [36] tried to detect exudates in retinal images using back-propagation neural networks (BPN). Their features were decided by two methods: decision trees and genetic algorithms with correlation-based feature selection (GA-CFS). In their experiment, the best BPN performance showed 98.45% accuracy. Kavitha and Duraiswamy [37] did some research on automatic detection of hard and soft exudates in fundus images, using color histogram thresholding to classify exudates. Their experiments showed 99.07% accuracy, 89% sensitivity, and 99% specificity. In 2014, de la Calleja et al. [38] used local binary patterns (LBP) to extract local features and artificial neural networks, random forest (RF), and support vector machines (SVM) for detection. In using a dataset containing 71 images, their best result achieved 97.46% accuracy with RF.

TABLE 2: Performance comparison.

Method	Year	Reference	Sn	Sp	Acc	AUC	Database
Vascular tracking	2010	[9]	0.7468	0.9551	0.9285	—	DRIVE
	2012	[10]	0.6887	0.9562	0.9290	—	STARE
Matched filtering	2009	[14]	0.6611	0.9848	0.9497	—	STARE
	2012	[15]	0.7191	0.9687	0.9407	—	STARE
			0.7154	0.9716	0.9343	—	DRIVE
Morphological processing	2011	[18]	0.7352	0.9795	0.9458	—	DRIVE
	2014	[19]	0.7862	0.9815	0.9598	—	DRIVE
Deformation model	2007	[21]	0.6634	0.9682	0.9316	—	DRIVE
	2009	[23]	—	0.9736	0.9087	—	STARE
			—	0.9772	0.9610	—	DRIVE
			0.7187	0.9767	0.9509	—	STARE
	2014	[24]	0.7354	0.9789	0.9477	—	DRIVE
			—	—	—	—	—
Machine learning	2007	[25]	—	—	0.9584	0.9602	STARE
	2011	[26]	—	—	0.9563	0.9558	DRIVE
			0.6944	0.9819	0.9526	—	STARE
	2013	[27]	0.7067	0.9801	0.9452	—	DRIVE
			0.8194	0.9679	0.9725	—	DRIVE
	2014	[28]	0.8104	0.9791	0.9813	0.9751	STARE
			0.8173	0.9733	0.9767	0.9475	DRIVE
	2014	[29]	0.7887	0.9633	0.9441	—	STARE
			0.7512	0.9684	0.9412	—	DRIVE

Sn: sensitivity, Sp: specificity, Acc: accuracy, AUC: area under curve.

4. Material and Methods

We propose an automatic DR detection algorithm, based on DCNN, fractional max-pooling [39], SVM [40], and teaching-learning-based optimization (TLBO) [41]. Specifically, we train two DCNN networks with fractional max-pooling, combining their prediction results using SVM and optimizing the SVM parameters with TLBO. The reason for training two distinct networks is that different network architectures may have their unique advantages in feature space representation. By training two DCNNs and combining their features, the prediction accuracy can be further enhanced. Another important factor impacting the recognition rate is the parameter of classifiers. We propose to optimize the SVM parameters using TLBO. We illustrate the image preprocessing methods in Section 4.1 and present the fractional max-pooling, SVM, and TLBO, in Sections 4.2, 4.3, and 4.4, respectively.

4.1. Preprocessing. Given the vessels in the original fundus images are mostly not very clear, and the size of each fundus image may differ, it is essential to preprocess images so that they have the same size and the visibility of the vessels is improved. There are three steps in preprocessing. The first is to rescale images to the same size. Since the fundus images are circular, we rescale the input images so that the diameter of the fundus images becomes 540 pixels. After rescaling, the local average color value is subtracted from the rescaled images, and another transformation is performed so that the local average is mapped to 50% gray-scale in order to remove the color divergence caused by different ophthalmoscopes. Last but not least, because boundary effects may occur in some images, we remove the

periphery by clipping 10% from the border of the images. Figure 4 shows the original fundus image and the image after preprocessing.

4.2. Fractional Max-Pooling. Pooling is a procedure that turns the input matrix $M_{in} \times N_{in}$ into a smaller $M_{out} \times N_{out}$ output matrix. The purpose is to divide the input matrix into multiple pooling regions ($P_{i,j}$):

$$\begin{aligned}
 p_i &\subset \{1, 2, 3, \dots, M_{in}\} \quad \text{for each } i \in \{1, \dots, M_{out}\}, \\
 p_j &\subset \{1, 2, 3, \dots, N_{in}\} \quad \text{for each } j \in \{1, \dots, N_{out}\}, \\
 P_{i,j} &= p_i \times p_j.
 \end{aligned} \tag{1}$$

The pooling results are computed according to pooling type:

$$\text{Output}_{i,j} = \underset{(k,l) \in P_{i,j}}{\text{Oper}} \text{Input}_{k,l}. \tag{2}$$

In equation (2), “Oper” refers to a particular mathematical operation. For example, if max-pooling is used, the operation will be to take the maximum of the input region. For average pooling, the average of the input region is taken. For such a network that requires tremendous learning, it is preferable to use as many hidden layers as possible. In this work, the pooling layer used in our networks is fractional max-pooling instead of general max-pooling.

Fractional max-pooling is a pooling scheme that makes the size of the output matrix equivalent to fractional times that of the input matrix after pooling, i.e., $N_{in} = \alpha N_{out}$, $\alpha \geq 0$, $\alpha \notin \mathbb{Z}$. To describe the general

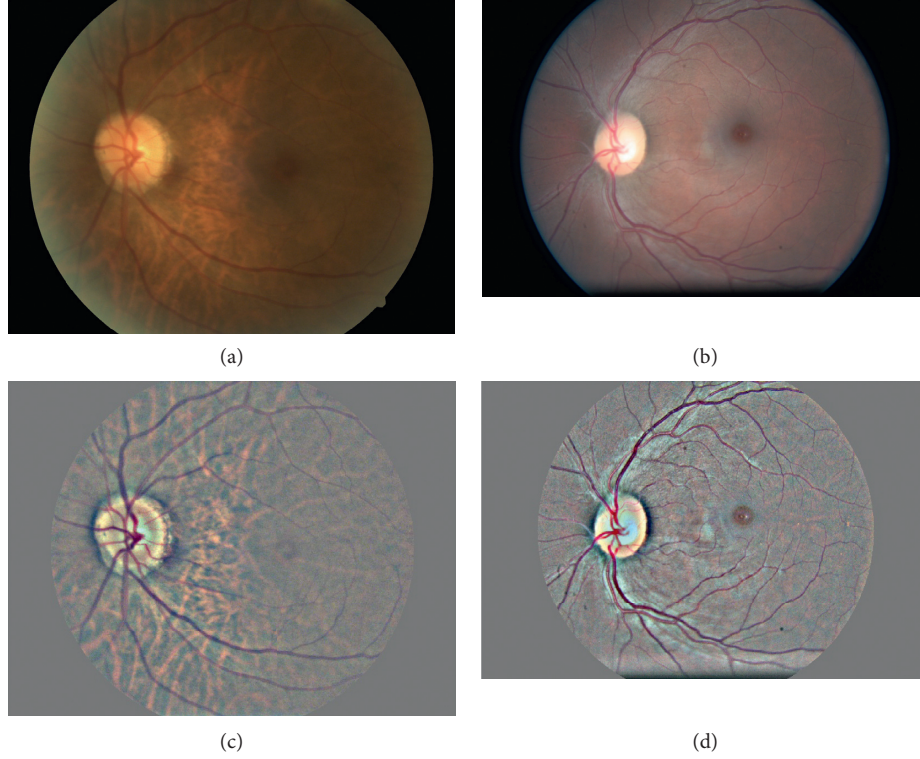


FIGURE 4: Image before and after the preprocessing stage. (a, b) images before preprocessing stage; (c, d) images after preprocessing stage.

pooling regions, let $(x_i)_{i=0}^{M_{\text{out}}}$ and $(y_j)_{j=0}^{N_{\text{out}}}$ be two increasing integer sequences starting with one and ending with M_{in} or N_{in} . These two sequences are used in pooling steps, as described in Figures 5 and 6.

$$P_{i,j} = [x_{i-1}, x_i - a] \times [y_{j-1}, y_j - b]. \quad (3)$$

The constants, a and b in equation (3), stand for the overlapping length and the width of the pooling window, respectively. Figure 5 is a simple example of overlapping pooling. Figure 6 illustrates different pooling region types.

After fractional max-pooling, the pooling window size is still integers, but the global pooling size will change. Namely, fractional max-pooling does not directly change the pooling window into a fractional scale. Instead, it uses windows of variable size to achieve fractional pooling. The generation of x and y sequences can be random or pseudorandom. Pseudorandom sequences generate more stable pooling regions than random sequences and can also achieve higher accuracy [39].

4.3. Support Vector Machine (SVM). SVM is a supervised learning method used for classification and regression analysis. SVM can find the hyperplane or decision boundary defined by the solution vector w , which not only separates the training vectors but also works well with unseen test data. To improve its generalization ability, SVM selects decision boundaries based on maximizing margins between classes.

Figure 7 illustrates the idea. Suppose there are n points in a binary dataset:

$$\begin{aligned} x &= \{1, 2, 3, 4, 5\}, y = \{1, 2, 4, 5\}, a = 0, b = 1 \\ P_{1,1} &= [x_{1-1}, x_1 - 0] \times [y_{1-1}, y_1 - 1] = [1, 2] \times [1, 2] \\ P_{1,2} &= [x_{1-1}, x_1 - 1] \times [y_{2-1}, y_2 - 1] = [1, 2] \times [2, 3] \end{aligned}$$

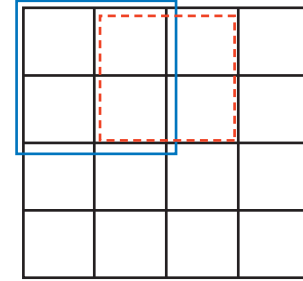


FIGURE 5: An overlapping example. The blue solid line indicates pooling region $P_{1,1}$ while the red dotted line shows the $P_{1,2}$.

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n), \quad (4)$$

where y_i is the data label, which can be 1 or -1, indicating the class to which \vec{x}_i belongs. We need to find the optimized hyperplane, such that the distance between the hyperplane and its nearest point \vec{x}_i is maximized. A hyperplane can be written as equation (5) based on \vec{x} :

$$\vec{w} \cdot \vec{x} - b = 0, \quad (5)$$

where \vec{w} is the normal vector of the hyperplane, and the value of $b/\|\vec{w}\|$ decides the margin of hyperplane from the training data point along the normal vector \vec{w} .

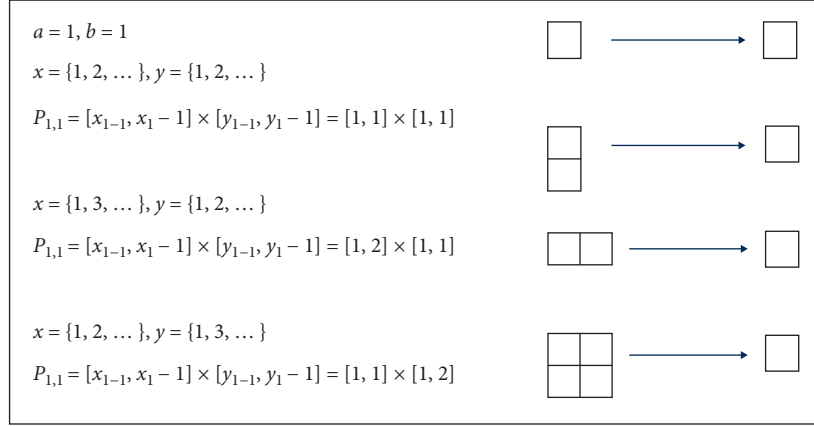


FIGURE 6: Different pooling types when both a and b equal to one. The left side shows the equations for pooling condition and computation of pooling region. Pictures on the right side indicate regions before and after pooling.

For y_i , whose value is 1, the data must satisfy $\vec{w} \cdot \vec{x} - b \geq 1$, and for y_i , whose value is -1, $\vec{w} \cdot \vec{x} - b \leq -1$ has to be satisfied. Combining these two conditions, we get

$$y_i (\vec{w} \cdot \vec{x} - b) \geq 1. \quad (6)$$

The goal is to maximize $b/\|\vec{w}\|$ according to the constrain of equation (6) in order to derive the optimized decision hyperplane for classification.

Sometimes, the training data might not be able to be perfectly separated using linear boundaries. Therefore, in the SVM formulation, we need to introduce the error metric ϵ and the cost parameter C , as shown in equation (7). The goal now becomes to minimize

$$\frac{1}{2} \vec{w} \cdot \vec{w} + C \sum_{i=1}^N \epsilon_i. \quad (7)$$

Subject to

$$\begin{aligned} \vec{w} \cdot \vec{x}_i + b &\geq 1 - \epsilon_i & \text{if } \vec{w} \cdot \vec{x}_i > b, \\ \vec{w} \cdot \vec{x}_i + b &\leq -1 + \epsilon_i & \text{if } \vec{w} \cdot \vec{x}_i < b, \\ \epsilon_i &> 0. \end{aligned} \quad (8)$$

The performance of SVM is influenced by two main parameters, the first one is C , which is a tunable parameter in equation (7). The other one is γ , which is used in the radial basis function (RBF) kernel to map data into a higher dimensional space before training and classification. The RBF kernel can be defined as

$$K(x_i, x_j) = e^{-\gamma |x_i - x_j|^2}, \quad (9)$$

where γ denotes the width of the Gaussian envelope in a high-dimensional feature space.

4.4. Teaching-Learning-Based Optimization (TLBO). TLBO, an evolution-based optimization algorithm, was proposed by Rao et al. [41], in 2011. The concept of TLBO is inspired by the evolution of the learning process when a group or a class of learners learn a target task. There are two

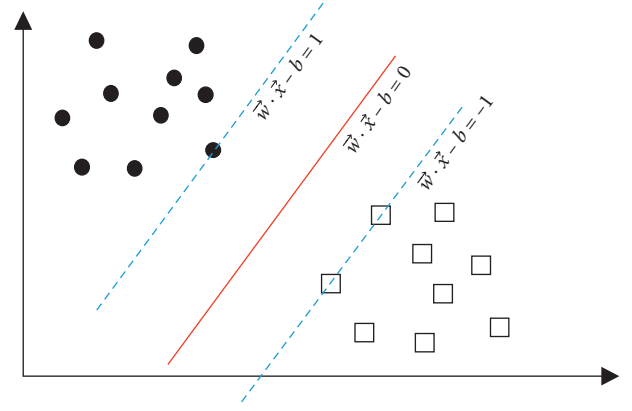


FIGURE 7: Example of decision boundary hyperplane with two classes of samples.

ways of learning in groups or classes: (1) learning from the guidance of the teacher and (2) learning from other learners. The procedure of TLBO can be divided into two phases, as described below in Sections 4.4.1 and 4.4.2.

4.4.1. Teacher Phase. In the whole population, the teacher (X_{teacher}) can be considered as the best solution. Namely, learners learn from the teacher in the teacher phase. In this phase, the teacher strives for enhancing the results of other individuals (X_i) by increasing the mean result of the classroom. This can be described as adjusting X_{mean} to approximate X_{teacher} . In order to maintain a stochastic nature during the optimization process, two randomly generated parameters, r and T_F , are applied in each iteration for the solution X_i as

$$X_{\text{new}} = X_i + r_i (X_{\text{teacher}} - T_F X_{\text{mean}}), \quad (10)$$

$$T_F = \text{round}[1 + \text{rand}(0, 1)]. \quad (11)$$

In equation (10), r_i is a randomly selected number in the range of 0 and 1. Moreover, X_{new} and X_i are the new and existing solutions at iteration i , respectively. T_F in equation (11) is a teaching factor which can be either 1 or 2.

4.4.2. Learner Phase. The learners gain their knowledge by interacting with each other. Therefore, an individual learns new information if other individuals have more knowledge than him or her. In this phase, the student X_i interacts randomly with another student X_j ($i \neq j$) in order to enhance his or her knowledge. Equation (12) shows that if X_i is better than X_j (i.e., $f(X_j) > f(X_i)$ for minimization problems), X_j is moved toward X_i . Otherwise, it is moved away from X_i .

$$X_{\text{new}} = \begin{cases} X_i + r_i(X_i - X_j), & f(X_j) > f(X_i), \\ X_i + r_i(X_j - X_i), & f(X_i) > f(X_j), \end{cases} \quad (12)$$

If the new solution X_{new} to the problem is better than the old ones, the new solution X_{new} will be recorded as the best solution. After updating the status of each learner, a new iteration begins. A stop criterion, based on the iteration number or the difference of the cost function, can be set to stop the iteration properly. The flowchart of TLBO is shown in Figure 8.

5. Result

Our fundus image data is from the database provided by one of the Kaggle contests; entitled “Identify signs of diabetic retinopathy in eye images” [42]. In this database, there are about 90,000 images. We separate 1000 images from the training dataset to be the validation dataset. The detailed information of each dataset is shown in Table 3, and our two network architectures are shown in Figure 9.

Our proposed method uses two DCNNs with fractional max-pooling layers. For every input fundus image, the two DCNN will output a vector of size 1×5 , representing the probability distribution of the prediction for each lesion (category). The probability distribution, together with other values, forms a feature with dimensionality 24. The 24 features are described as follows:

- (i) DCNN probabilities of each lesion, respectively (5 features)
- (ii) Averages of R, G, and B channel values within 50% * 50% center cropped image (3 features)
- (iii) Widths and heights of 50% * 50% center cropped image (2 features)
- (iv) Overall standard deviation of the original image and 50% * 50% centered cropped image, Laplacian-filtered image (2 features)
- (v) In total, there are 12 features for one fundus image. We then append another 12 features from the fundus image of the other eye of the same subject. Therefore, the overall length of the feature vector is 24 for one fundus image. The 24 feature vectors of dimensionality are used as input vectors of SVM

The 24-dimensional vector is used to train a multiclass SVM (five classes), whose parameters are optimized using the TLBO method. We implemented the method described

in [39] and used it as the baseline. The baseline system uses similar features with a scheme of ensemble classifier (RF).

We used the validation set data to optimize the parameter set (C , γ) in SVM using TLBO. The upper and lower bounds of the parameter are set within $[0, 100]$. We ran 50 iterations with 50 students.

Our final accuracy for five-class classification task of DR is 86.17% and the accuracy for the binary class classification task is 91.05%. Labels for five-class classification are normal, NPDR level 1, NPDR level 2, NPDR level 3, and PDR while labels for binary class classification are normal and abnormal. For binary classification, its sensitivity is 0.8930 while the specificity is 0.9089. Except counting accuracy, we also do a T -test for our binary class classification. The T -test is also called the Student’s t -test. It is a statistical hypothesis test, in which the test statistic follows a Student’s t -distribution. Usually, the t -test is used to compare whether there is a significant difference between two groups of data and assists in judging the data divergence. In doing a paired samples t -test with results from binary class classification and random judgment, its outcome is 1 for the hypothesis test result, zero for the p value and $[0.3934, 0.4033]$ for the confidence interval, under null hypothesis at the 5% significant level.

The hypothesis test result is an index that tells whether two data come from the same distribution or not. If the data come from the same distribution, the value of the hypothesis test result will be close to 0. On the contrary, if the data resources are distinct, the result will be close to 1, which means there is a differentiation between the data. The p value is the probability of accepting the assumption that there is a difference between two data may be wrong. The smaller the p value, the more reason that there is a disparity between data.

Also, we designed an app called “Deep Retina,” providing personal examination, remote medical care, and early screening. Figure 10 shows our app interfaces. After choosing a fundus image that the user wants to check, it will send the image to our server and use our designed machine learning method. It takes about 10 s (depends on network speed) to get the result, which will be presented as the probability of each lesion. With a handheld device, individuals can do the initial examination at the district office or even at home. More importantly, it can benefit some remote areas that lack medical resources.

6. Discussion

6.1. Accuracy Improvement. Table 4 shows the accuracy comparison when using different classifiers and parameter optimization methods on each dataset. Using the default parameters with SVM (without optimization), accuracies in both validation and test sets are higher than that of the RF [39]. If we optimize the parameters using the default parameter searching method provided in the LIBSVM software package, though it achieves very high accuracy in the five-fold cross validation experiment, the

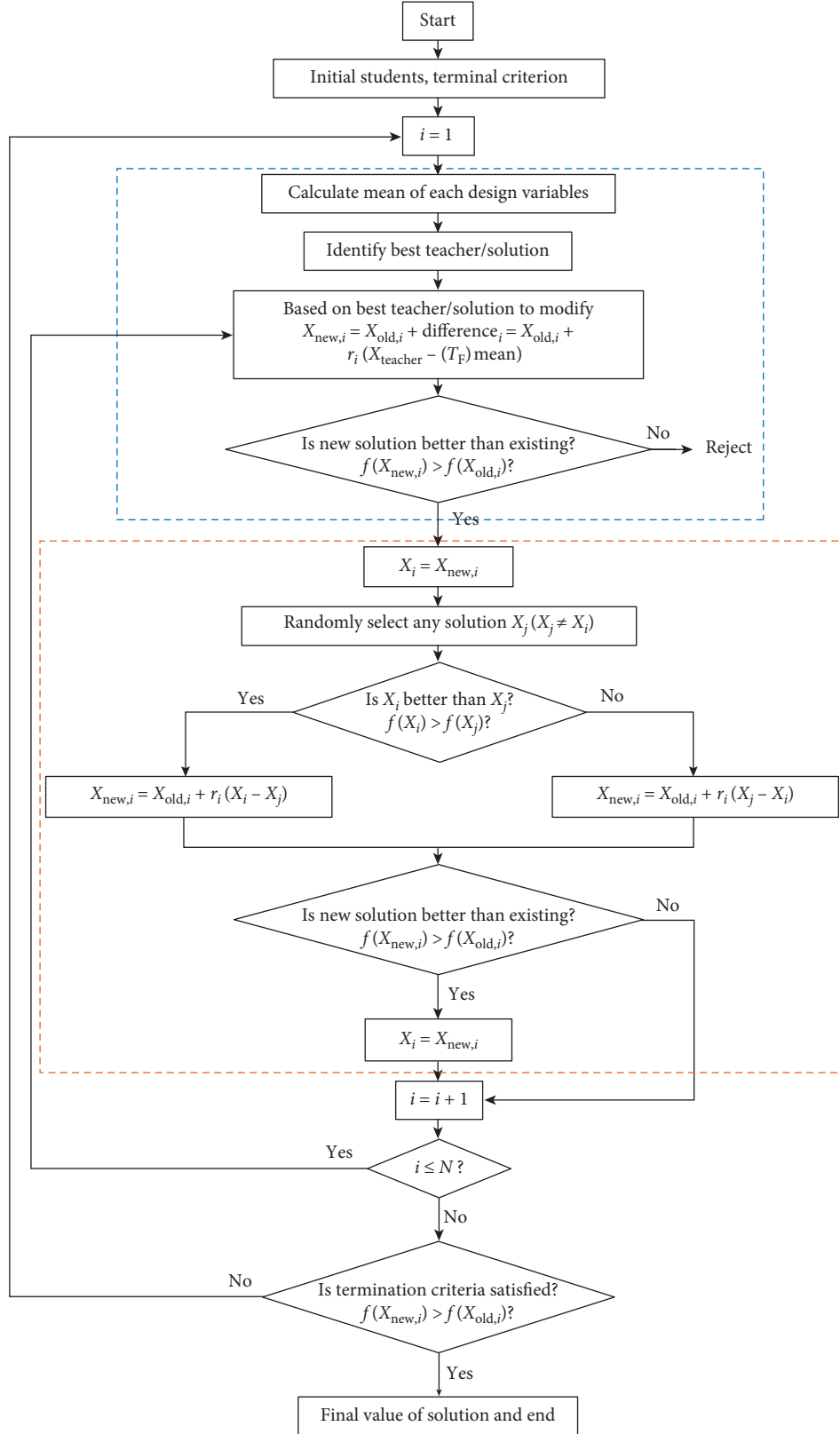


FIGURE 8: TLBO flowchart.

validation and test accuracies are even lower than the default one. From this result, we believe that overfitting arises when optimizing parameters in SVM.

Table 5 shows the confusion matrix of the classification results from the two DCNN networks (before performing SVM classification). Network 1 is the architecture shown on

TABLE 3: Detailed information of each dataset.

Dataset	# of images	DR lesions				
		0	1	2	3	4
Train	34124	73%	6%	15%	3%	3%
Validation	1000	70%	8%	15%	4%	3%
Test	53572	74%	7%	15%	2%	2%

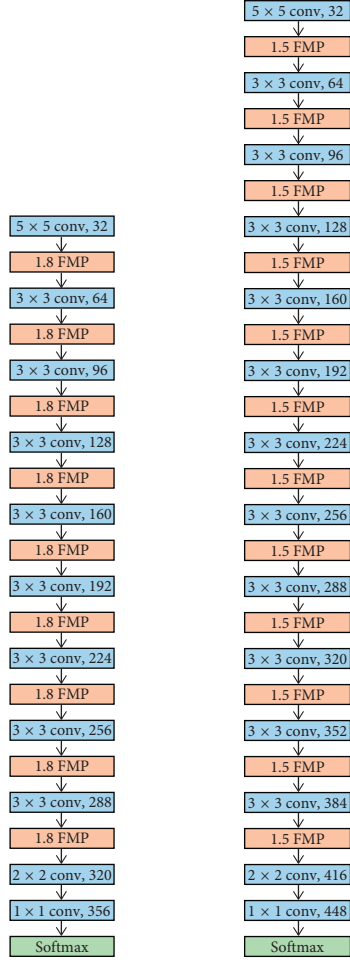


FIGURE 9: Architecture of the two DCNN networks that we used.

the left side of Figure 9, and network 2 is the one on the right side. From Table 5, it shows that the lesion classifications of 0 and 2 are better than the other categories. For lesion 1, most of the prediction results are incorrect. Also, for lesions 3 and 4, the majority of the results are misclassifications that are classified into lesion 2.

Table 6 shows the confusion matrix of the classification results using the full procedure of the proposed method (using SVM with TLBO). Table 7 displays the difference between Tables 5 and 6, which serves as a performance comparison between the two methods (using DCNN only and DCNN + SVM + TLBO). From the table, every class, except class 0 and overall accuracy, is increased in network 1. For network 2, each accuracy, except class 3, is increased. The decline in accuracy of class 3 is mainly caused by

misclassification of class 2. Table 8 shows the confusion matrix of the classification results using the baseline method as reported in [39], for comparison purposes.

6.2. Deep Learning vs Traditional Classification Methods. Many traditional classification methods try to solve the problem of DR detection by (1) using image processing to capture symptoms in fundus images and then (2) building a classifier to make decisions based on the detected symptoms (1). The shortcoming of image processing methods is that the manifestations of the symptoms are random across different images; therefore, it is extremely time-consuming and requires intense efforts to label the locations of the symptoms. Abiding by the new philosophy that comes with the emergence of the deep learning technology, our proposed method is trying to learn how to make decisions directly from the image data itself. Different than the former approaches, our images only need to be labeled with lesion number instead of labeling symptom locations. Consequently, it saves considerable time during the database preprocessing stage. On top of the classification results by the two DCNN networks, we use SVM optimized by TLBO to generate an improved outcome, and we achieve 86.17% accuracy. Our result is better than the first-place winner in the Kaggle competition. It shows that our research result is the state-of-the-art.

6.3. Limitation. In our current datasets, the number of images of lesions 3 and 4 is not sufficient to train a network, which is a limitation of the proposed method. Therefore, one of our future works is to develop deeper collaborative relations with hospitals and clinics to acquire more data of lesions 3 and 4. With more data, we believe the classification accuracy will be further increased. In addition, from our result, we found that it is hard to differentiate the images between lesions 0 and 1. Therefore, when we collect new data, it is desirable to collect more images belonging to lesions 0 and 1. Also, we can attempt to use a different network architecture for this problem.

7. Conclusion

It is feasible to train a deep learning model for automatic diagnosis of DR, as long as we have enough data for statistical model training. Furthermore, the database preparation stage only needs a categorical label for each training image. It does not require detailed annotation for retinal vessel tracking in every image. Hence, it is time-efficient compared to the traditional machine learning-based method for automatic diagnosis of DR. The final accuracy can achieve 86.17% and 91.05% for five-class and binary class classifications, respectively.

The sensitivity and specificity of binary classification are 0.8930 and 0.9089, respectively, which is a satisfactory result. Furthermore, we developed an automatic inspection app that can be used in both personal examination and remote medical care. With more image data collected, we expect the accuracy can be even more enhanced, further improving our system.

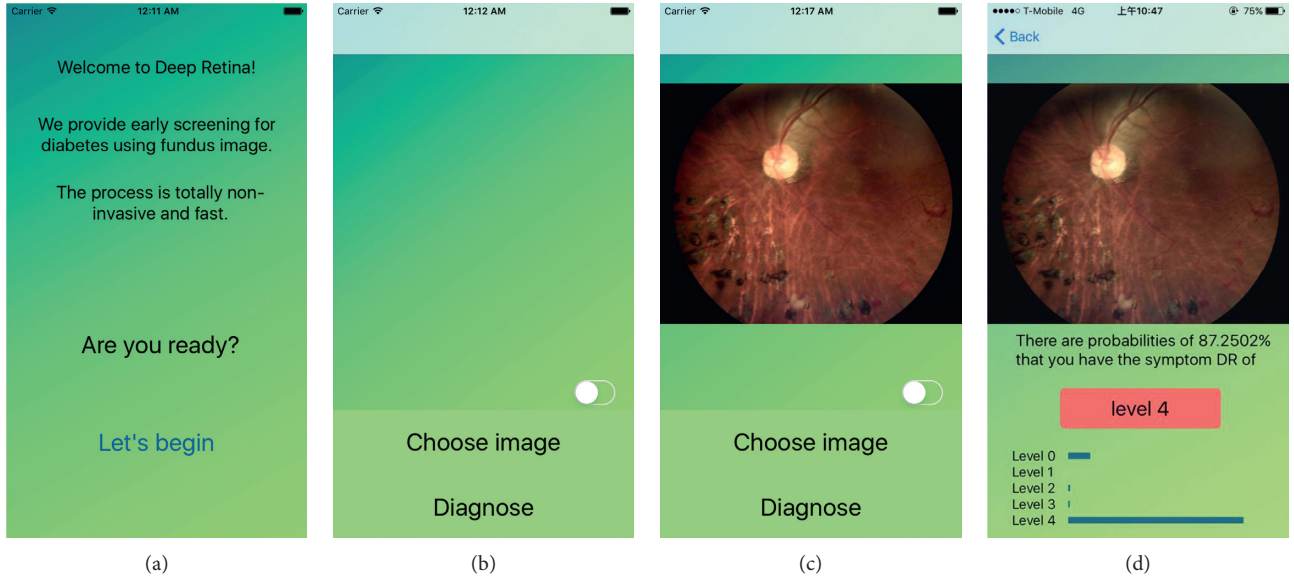


FIGURE 10: App interfaces.

TABLE 4: Accuracy comparison when using different classifiers and different parameter optimization methods on each dataset.

Dataset	Method			RF [39]
	Without optimization	SVM Default parameter searching	TLBO	
Five-fold cross validation (using training data)	88.744%	96.5724%	—	—
Validation data	85.4%	79.6%	—	84.7%
Test data	86.1177%	81.0573%	86.17%	86.02%

TABLE 5: Confusion matrix of the classification results from the two DCNN networks, before performing SVM classification.

Determined level	Ground-truth level									
	Network 1					Network 2				
	0	1	2	3	4	0	1	2	3	4
0	39031	2693	2361	75	160	38310	2233	1617	51	78
1	118	438	185	0	1	327	735	290	1	3
2	339	626	5058	738	392	790	786	5509	620	345
3	0	0	170	332	102	4	1	338	486	185
4	43	5	85	69	551	100	7	105	56	595
Accuracy (%)	98.74	11.64	64.36	27.35	45.69	96.91	19.54	70.01	40.03	49.34
Overall accuracy (%)			84.76					85.18		

TABLE 6: Confusion matrix of the classification results using the full procedure of the proposed method (using SVM with TLBO).

Determined level	Ground-truth level				
	SVM optimized with TLBO				
	0	1	2	3	4
0	38611	2233	1513	44	80
1	312	831	349	1	1
2	570	696	5706	701	393
3	3	0	215	408	123
4	35	2	76	60	609
Accuracy (%)	97.67	22.09	72.60	33.61	50.50
Overall accuracy (%)			86.17		

TABLE 7: Differences between Tables 5 and 6.

Determined lesion	Ground-truth lesion									
	SVM-network 1					SVM-network 2				
	0	1	2	3	4	0	1	2	3	4
0	+420	-460	-848	-31	-80	+310	0	-104	-43	+2
1	+194	+393	+164	+1	0	-15	+96	+59	0	-2
2	+231	+70	+648	-37	+1	-220	-90	+197	+81	+48
3	+3	0	+45	+76	+21	+1	-1	-123	-78	-62
4	+3	-3	+45	+76	+58	-65	-5	-28	+4	+14
Accuracy increment (%)	-1.08	89.77	12.80	22.88	10.52	0.78	13.05	3.69	-16.03	2.3
Overall accuracy increment (%)			1.6					1.16		

TABLE 8: Confusion matrix of the classification using the baseline method [39] and its overall accuracy.

Determined	Original Random forest				
	0	1	2	3	4
0	38228	2001	1331	38	64
1	546	1026	434	3	2
2	710	732	5726	635	358
3	2	1	279	465	146
4	45	2	89	73	636
Accuracy (%)	96.70	27.27	72.86	38.30	52.74
Overall accuracy (%)			86.01		

Data Availability

All experiment data come from a Kaggle contest called “Identify signs of diabetic retinopathy in eye images” (website: <https://www.kaggle.com/c/diabetic-retinopathy-detection>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Mr. Chia-Ming Hu, for his help in preparing computational instruments and performing parts of the experiments in this paper.

References

- [1] N. H. Cho, J. E. Shaw, S. Karuranga et al., “IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045,” *Diabetes Research and Clinical Practice*, vol. 138, pp. 271–281, 2018.
- [2] American Diabetes Association, *Eye Complications*, American Diabetes Association, Arlington, VA, USA, 2013, <http://www.diabetes.org/living-with-diabetes/complications/eye-complications/?referrer=https://www.google.com.tw/>.
- [3] International Diabetes Foundation (IDF), *Diabetes Atlas*, International Diabetes Federation, Brussels, Belgium, 7th edition, 2015.
- [4] C. P. Wilkinson, F. L. Ferris, R. E. Klein et al., “Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales,” *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.
- [5] A. Tufail, C. Rudisill, C. Egan et al., “Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders,” *Ophthalmology*, vol. 124, no. 3, pp. 343–351, 2017.
- [6] A. Tufail, V. V. Kapetanakis, S. Salas-Vega et al., “An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness,” *Health Technology Assessment*, vol. 20, no. 92, pp. 1–72, 2016.
- [7] I. Liu and Y. Sun, “Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme,” *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, pp. 334–341, 1993.
- [8] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, “Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 2, pp. 125–138, 1999.
- [9] M. Vlachos and E. Dermatas, “Multi-scale retinal vessel segmentation using line tracking,” *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 213–227, 2010.
- [10] Y. Yin, M. Adel, and S. Bourennane, “An automatic tracking method for retinal vascular tree extraction,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, Kyoto, Japan, March 2012.
- [11] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, “Detection of blood vessels in retinal images using two-dimensional matched filters,” *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 263–269, 1989.

- [12] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.
- [13] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 131–137, 2003.
- [14] L. Zhang, Q. Li, J. You, and D. Zhang, "A modified matched filter with double-sided thresholding for screening proliferative diabetic retinopathy," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 528–534, 2009.
- [15] Q. Li, J. You, and D. Zhang, "Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7600–7610, 2012.
- [16] F. Zana and J. C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1010–1019, 2001.
- [17] G. Ayala, T. León, and V. Zapater, "Different averages of a fuzzy set with an application to vessel segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 384–393, 2005.
- [18] M. S. Miri and A. Mahloojifar, "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1183–1192, 2011.
- [19] D. Karthika and A. Marimuthu, "Retinal image analysis using contourlet transform and multistructure elements morphology by reconstruction," in *Proceedings of the World Congress on Computing and Communication Technologies*, IEEE, Tiruchirappalli, India, 2014.
- [20] M. Kass, A. Witkin, and D. Terzopoulos, "Snake: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [21] L. Espona, M. J. Carreira, M. Ortega et al., "A snake for retinal vessel segmentation," in *Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis*, Springer-Verlag Berlin and Heidelberg GmbH & Co., KG, Girona, Spain, June 2007.
- [22] B. Al-Diri and A. Hunter, "A ribbon of twins for extracting vessel boundaries," in *Proceedings of the 3rd European Medical and Biological Engineering Conference*, Prague, Czech Republic, November 2005.
- [23] Y. Zhang, W. Hsu, and M. L. Lee, "Detection of retinal blood vessels based on nonlinear projections," *Journal of Signal Processing Systems*, vol. 55, no. 1–3, pp. 103–112, 2009.
- [24] Y. Q. Zhao, X. H. Wang, X. F. Wang, and F. Y. Shih, "Retinal vessels segmentation based on level set and region growing," *Pattern Recognition*, vol. 47, no. 7, pp. 2437–2446, 2014.
- [25] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1357–1365, 2007.
- [26] D. Marin, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 146–158, 2011.
- [27] V. Shanmugam and R. S. D. W. Banu, "Retinal blood vessel segmentation using an extreme learning machine approach," in *Proceedings of the 2013 IEEE Point-of-Care Healthcare Technologies*, IEEE, Bangalore, India, January 2013.
- [28] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, 2015.
- [29] A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu, "Segmentation of the blood vessels and optic disk in retinal images," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1874–1886, 2014.
- [30] R. M. Cesar Jr. and H. F. Jelinek, "Segmentation of retinal fundus vasculature in nonmydriatic camera images using wavelets," in *Angiography and Plaque Imaging: Advanced Segmentation Techniques*, J. S. Suri and S. Laxminarayan, Eds., pp. 193–224, CRC Press, Boca Raton, FL, USA, 2003.
- [31] J. J. G. Leandro, J. V. B. Soares, R. M. Cesar et al., "Blood vessels segmentation in nonmydriatic images using wavelets and statistical classifiers," in *Proceedings of the XVI Brazilian Symposium on Computer Graphics and Image Processing*, IEEE, São Carlos, Brazil, October 2003.
- [32] Y. A. Tolias and S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 263–273, 1998.
- [33] S. Xie and H. Nie, "Retinal vascular image segmentation using genetic algorithm plus FCM clustering," in *Proceedings of the 2013 Third International Conference on Intelligent System Design and Engineering Applications*, IEEE, Hong Kong, China, January 2013.
- [34] B. M. Ege, O. K. Hejlesen, O. V. Larsen et al., "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Computer Methods and Programs in Biomedicine*, vol. 62, no. 3, pp. 165–175, 2000.
- [35] N. Silberman, K. Ahrlich, R. Fergus et al., "Case for automated detection of diabetic retinopathy," in *Proceedings of the AAAI Spring Symposium: Artificial Intelligence for Development*, Stanford, CA, USA, March 2010.
- [36] A. G. Karegowda, A. Nasiha, M. A. Jayaram, and A. S. Manjunath, "Exudates detection in retinal images using back propagation neural network," *International Journal of Computer Applications*, vol. 25, no. 3, pp. 25–31, 2011.
- [37] S. Kavitha and K. Duraiswamy, "Automatic detection of hard and soft exudates in fundus images using color histogram thresholding," *European Journal of Scientific Research*, vol. 48, pp. 493–504, 2011.
- [38] J. de la Calleja, L. Tecuapetla, M. A. Medina et al., "LBP and machine learning for diabetic retinopathy detection," in *Proceedings of the 2014 International Conference on Intelligent Data Engineering and Automated Learning*, Springer, Salamanca, Spain, September 2014.
- [39] B. Graham, "Fractional max-pooling," 2015, <https://arxiv.org/pdf/1412.6071v4.pdf>.
- [40] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.
- [41] R. V. Rao, V. J. Savsani, and D. P. Vakharia, "Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems," *Computer-Aided Design*, vol. 43, no. 3, pp. 303–315, 2011.
- [42] Kaggle contests, *Identify Signs of Diabetic Retinopathy in Eye Images*, Kaggle, San Francisco, CA, USA, 2015, <https://www.kaggle.com/c/diabetic-retinopathy-detection>.

Research Article

ProMe: A Mentoring Platform for Older Adults Using Machine Learning Techniques for Supporting the “Live and Learn” Concept

Giorgos Kostopoulos,¹ Katja Neureiter,² Dragos Papatoiu,³ Manfred Tscheligi,⁴ and Christos Chrysoulas⁵ 

¹Gluk Advice B. V., Maastricht, Netherlands

²AIT Austrian Institute of Technology GmbH, Seibersdorf, Austria

³Siveco SRL, București, Romania

⁴Center for Human-Computer Interaction, University of Salzburg, Salzburg, Austria

⁵Computer Science & Informatics Dep., London South Bank University, London, UK

Correspondence should be addressed to Christos Chrysoulas; chrysouc@lsbu.ac.uk

Received 25 June 2018; Accepted 11 November 2018; Published 2 December 2018

Guest Editor: Giovanna Sannino

Copyright © 2018 Giorgos Kostopoulos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is well known what an important role employment plays in our lives and how it influences our everyday life. With the help of employment, all of us as individuals manage to create our personal and social net that not only gives meaning and financial security but also heavily affects our social status, contributes to self-esteem, and plays a vital role on the way our social relationships at work and (not only) are formed. Being work active even after the retirement time has been found to have mostly useful and positive effects. Age groups that are in the after retirement age group (65 and more) and keep working can significantly improve their mental health since they feel that there are still things to offer to others. ProMe aims to provide a unique framework for the people being in this transitional phase of their life, from work to retirement, thus playing a vital role in supporting their mental health and well-being. ProMe provides a platform for the younger generation to benefit from the accumulated professional and formal knowledge and experience of the older. In this way, ProMe supports professional intergenerational cooperation and mentoring, bringing together older adults with younger generations.

1. Introduction

Employment offers an opportunity to use one's skills and experience and to maintain social contacts beyond retirement, enhancing self-esteem by providing one's expertise to and for others. Activities such as voluntary work (e.g., helping others) and programs that bridge employment and retirement (i.e., any kind of paid occupation between a person's career and the withdrawal from labor force) have been shown to enhance one's well-being and quality of life. The effects of demographic aging show that constant decrease on the share of people of working age compared to an increasing relative number of those who are retired. This demographic shift requires society to reconsider the role

older adults' play in a society, even after retirement. In the Organisation for Economic Co-operation and Development (OECD) countries, it is common for workers to take early retirement, leaving the labor market before the standard pension eligibility age. This, in turn, leads to an increasing number of economically inactive people relative to those who are economically active. ProMe [1] aims to change perceptions on aging and the vocational activities of older adults. It offers an innovative solution that is designed to serve intergenerational dialogue, lifelong learning, and, on top of those, creating value among generations. It will also play a vital role on the mental health of the elderly since there are numerous studies specifically pointing out the importance of work in the elderly (after retirement) [2–7].

Based on these reflections, ProMe aims in providing meaningful work opportunities in the life of older adults, in order to smoothly pass from work to retirement and what after that. During their working life (or no), adults accumulated a lot of knowledge in domains related with their work and beyond. Our vision is to harness this and to achieve professional intergenerational cooperation and mentoring via our proposed platform, bringing together older adults with younger generations. Especially in the work context, mentoring is considered a valuable approach, e.g., to enhance the diffusion of expert knowledge within a working group or an organisation.

ProMe is based on theoretical concepts for mentoring [8, 9]. modern (professional) social networks (e.g., Xing, ResearchGate, and LinkedIn) try to promote strengthening social relations among people who share interests and activities through their mechanisms. Apart from the above, ProMe offers the opportunity for older adults' meaningful occupation on a voluntary basis through taking an active role as a mentor. This is feasible by encouraging them to share their knowledge and experience with younger generations who are eager to learn, trying to include the ones that not only seeking a career but also a more general life advice.

Through ProMe, older adults will be able to continue working thus allowing them to be and play an active role in social and working life, even after their retirement. In this way, added value for the society and economy is created. We consider mentoring as "walking alongside" a person—sharing expertise, values, and skills. It entails informal communication, including voluntary engagement, mutual respect, shared responsibility, and empowerment of the mentee, resulting in a mutually beneficial relationship. It can be described as a process for the informal transmission of knowledge, ranging from accompanying a person over a longer period (e.g., preparing a project or settling in a new company) to simply providing advice or information regarding specific topics (e.g., when applying for a new job).

ProMe provides an effective way of bringing together people in a long term and effectively relationship they need for developing a shared and transparent psychological contract they both commit to. "The term psychological contract is used to describe a set of individual beliefs or set of assumptions about promises voluntarily given and accepted in the context of a voluntary exchange relationship between two or more parties" [10]. "Psychological contract theory suggests that we shift the focus from what one expects to gain from the relationship to what one feels he or she is obligated to provide in the relationship. (...) understanding these obligations might provide valuable insight into why some specific functions are provided and others are not, especially with regard to structural characteristics of the relationship, such as the level of formality" [9]. In such collaborative relationships, you have two and sometimes even more, if we are talking for instance for a learning network multiple, psychological contracts: a psychological mentoring contract of a "requester of support" contains the obligations each of the involving parts owe each other [11]. For an effective collaborative relationship, it is important that both psychological contracts are compatible and preferably

transparent for both parties (so forming a shared contract). Even though, a possible lack of agreement on the mutual obligations does not mean that there is no psychological contract. An important point to remember is that psychological contracts are conceptualized as an individual's perceptions of mutual obligations and that actual agreement on the contract terms is not a requirement for a psychological contract to exist [12]. There is the possibility for an individual to form a type of psychological contract with someone he/she considers to be a supporter (or protégé) without the other party to develop a similar psychological contract with them. In this scenario, there is the possibility that the collaborative relationship will not be productive and long lasting [11]. Associated with the promises each party makes with another are mutual obligations and expectations, and depending on each party's beliefs about these promises, a psychological contract is subject to variations in expectations about that contract, i.e., matches and mismatches [13], which may affect the potential for each party's expectations being met. If beliefs and assumptions and their underlying promises are clear for all involved parties, it is more likely the expectations will be met. We highly believe that valid "information" about what each part expects and what to be expected of them, as they get involved in one of the three collaborative relationships, will play a vital role to the quality and the sustainability of that collaborative relationship.

Based on these considerations and the pivotal role of mentoring, the ProMe framework/platform is supporting different opportunities for informal communication through a number of functionalities, like video/text-chat, email, blogs, and forums. Apart from the aforementioned means of informal communication, ProMe provides potential-end users with the opportunity to take over different kinds of "mentoring roles." Recognizing the diverse aspects of mentoring, different types/roles of mentors and mentees, have been considered based on someone's experience (please also check Section 4, regarding the system architecture).

The remainder of this paper is organized as follows. Section 2 surveys related work in the area. Section 3 presents a representative use case of the proposed framework. Section 4 provides an insight on the internal architecture of the system. In Section 5, the proposed recommendation system is presented, while Section 6 concludes the paper and provides future directions.

2. State of the Art

There are numerous web-based platforms currently available that aim to match and guide mentors and protégés, for example, connecting graduates with a mentor to contemplate on one's career or to support students to achieve their goals. Most of these services also have partnerships, with colleges, universities, or professional societies to increase resources at a local level and expand mentoring programming and opportunities. In contrast to these solutions, ProMe does not only provide an opportunity to match mentor and mentee, but facilitates contact by means of the different functionalities (e.g., video conferencing and text

chat) that are offered via the platform on different devices. It provides a flexible communication environment and allows mentoring independent of time and space, which is currently only realized by a small number of online e-mentoring services, like Horseshmouth [14] and Senior Engage [15].

The Aware [16] project proposes a telematics platform for older adults to prepare them for the transition to retirement by allowing sharing knowledge to maintain an active role after retirement. It also provides a training module to acquire expertise in using information and communication technologies (ICTs) and services over Internet. ProMe approach offers different opportunities for maintaining meaningful occupation not only in the transition to retirement but also beyond. Besides sharing knowledge ProMe aims at supporting intergenerational learning so that both, older adults and younger generations, benefit from the solution.

E-mentoring [17] provides Internet-based education, social, and corporate email-based mentoring, while ProMe provides a variety of different mentor roles (according to users' capabilities) and suitable communication opportunities for the mentor and the mentee (e.g., text/video-chat, forum, and blog).

The Hear Me [18] project connects retirees with young adults who have not completed their secondary education and lack the qualifications for meaningful employment. Also provides cognitive and practical education for older adults to become mentors for youths at risk. Our approach not only focuses on a specific group of mentees but also addresses interested mentees from all generations. ProMe is not intended to be an education tool but considers older adults as competent to get active on the platform, focusing on supporting them to share knowledge and skills.

The Horseshmouth [14] project aims in the creation of a public value through the exchange of experiences and knowledge concerning all areas of life (e.g., work and learning). ProMe from the other side focuses on facilitating the exchange of professional knowledge supporting the older adults in continuing and managing an occupational lifestyle.

Horizons Unlimited [19] platform provides mentoring and coaching programs for personal and professional development through matching mentors and mentees. ProMe supports autonomy and independence by allowing older adults to get active as a mentor according to one's own expertise. Tailored specifically to the needs of older adults creates value for the target group by addressing a variety of user experience and acceptance factors (curiosity, motivation, usefulness, accessibility, flexibility, etc.).

The Mentor [20] project aims for young people between the ages of 6 and 18, supports and guides them to build productive and as much as possible meaningful lives, and also matches mentor and mentees. The ProMe approach goes beyond matching mentor and mentee by providing different roles and means of communication. Mentor Net [21] aims at connecting students in engineering and science with mentors in industry. It helps students to achieve their goals through e-mentoring. ProMe encourages mentoring between all age groups. It not only focuses on supporting the mentee but aims at fostering mutually beneficial relationships.

The Senior Engage [15] approach offers an opportunity for retired professionals to share knowledge and experiences with younger professionals to reduce isolation. It uses different means of communication such as video/text chat, and email. The ProMe approach from the other side not only offers an opportunity to exchange knowledge and experiences but also is based on a concept of mentoring, providing a variety of different mentoring roles and continuously supporting the mentoring relationship. It is also worth mentioning that it is not built from scratch and uses existing social structures people are embedded in (e.g., Xing and LinkedIn).

3. Use Case Scenario

Susan is 58 years old in the transition to her retirement. She has more than thirty years of experience working in a marketing department of a large company and has gained a significant amount of knowledge and expertise throughout the years. Although Susan is looking forward to having more time to spend with her two grandchildren and to pursue her hobbies, she is also thinking about potential opportunities for future occupation to maintain purpose in her life. Susan does not see herself as being an "old person" but as vital and active, and she wants to continue working to some extent, probably on a voluntary basis. Her greatest fear is that she is no longer needed and that all the knowledge she gained throughout the years is lying fallow.

Recently, Susan was invited by her company to give a presentation on marketing strategies based on her knowledge and experience. She enjoyed being involved and would like to continue this kind of activity. At this event, Susan heard about the ProMe platform and its use in sharing professional knowledge with younger generations. She became curious and had a look at the platform. As she could easily access the platform with her LinkedIn account, she logged in and explored the different opportunities and functionalities it provides. She extended her profile, indicating information about her professional knowledge and experience and what kind of mentoring roles she could imagine taking over. The intelligent expert system (agent) provided suggestions for potential mentees, searching for expertise in the area of marketing and she started looking at the several interesting profiles. She did not initially want to take over a specific mentoring role, and she simply started posting some ideas on marketing strategies on the ProMe platform. She obtained interesting feedback from the young students; they were enthusiastic about the blog posts. Following this initial success, she decided to take over the role of a leader after a few weeks. With the easy to use functionalities of the ProMe platform, she started to offer training via video conferencing where she discusses business concepts and ideas with young people. Susan really enjoys it and notices that, in addition to sharing her thoughts and experience, she is also learning new ideas from the young mentees (a two-way process).

Susan wanted to find out more about the different mentoring roles she could become involved with and took the online "Mentor-Check." Apart from her interests,

expertise, and knowledge she already indicated when extending her profile, the Mentor-Check included questions about the time she could imagine investing being a mentor on the ProMe platform.

The results indicated that she had the skill set required to be the perfect buddy. She really liked the idea of accompanying a younger person, e.g., a marketing graduate starting their first job out of college. Again, the agent supported her in easily finding mentees looking for a mentor in the area she had expertise in. She reviewed the suggested profiles of young people who were searching for a buddy and decided to get in contact with a young woman who was going to start her first big project on her own. They worked together to establish the management concept, using working tools (e.g., calendar to coordinate appointments), concepts, and communication and collaboration tools such as video-chat, email, or Whiteboard, allowing brainstorming together on a first concept. Susan has been working together with her mentee for several months and was perfectly supported by the functionalities provided on the ProMe platform according to the specific needs of this mentoring relationship. Also, after her mentee had successfully started her project, they stayed in contact via the platform for further support in everyday businesses.

4. System Architecture

The actual ProMe platform users are considered users with different characteristics on the systems. Figure 1 provides an overview of the different roles in the ProMe system.

- (i) Guest/anonymous users: guest/anonymous users are users that can access publicly available sections of the portal without any need for authentication/authorization. For example, they can have access to the homepage where they can find what the project is about and success stories.
- (ii) Registered users: as registered users, we consider those users that have to follow the normal sign-in procedure when trying to access the platform. Registered users are in position to access more personalized (restricted) content and make use the tools that the ProMe platform can offer, based on the role they serve. Three types of registered users exist, considering their roles and attributes in the project. In the scope of the ProMe project, we have identified two types of users: one that acts as a receiver and another that acts as a provider. Those two pairs of user were considered best for the scope of the project in user testing and experts evaluation.
- (iii) Provider user: provider users are those users that based on their previous professional experience can choose a suitable role(s) from the list of roles or by doing a “Mentor-check,” i.e., based on its profile (expertise, interests, etc.).
 - (a) Mentor
 - (b) Advisor

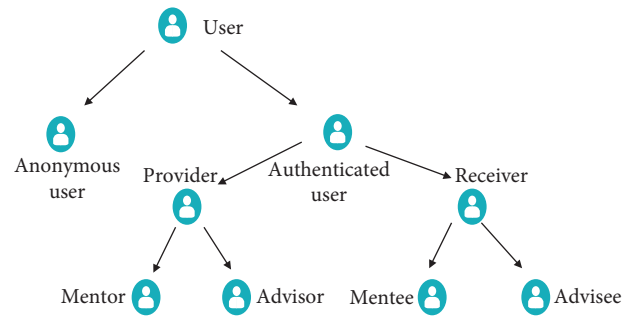


FIGURE 1: ProMe platform users.

- (iv) Receiver user: receiver user has the role of a participant in the activities initiated by the providers.
 - (a) Mentee
 - (b) Advisee
- (v) Administrators: administrators are users that can perform administrative tasks in the system.

The ProMe system is in position to adequately provide users a modern way of communication; thus easily exchange useful learning materials. In Figure 2, we can see all the different levels that the ProMe platform is based on.

Right in the middle and high up, there is the presentation level. The presentation level contains those components needed so as the ProMe platform to be in position to deliver services to the users. These components deliver the graphic interface for the users to access the platform, and platform information that will allow any external systems to gain access to provided services. The described graphical interface is heavily based on up-to-date frameworks like HTML5 or CSS3. Web services are exposed using REST [22].

The service level contains services for implementation of the business flow. The aforementioned set of services can be used by the integration and presentation service blocks in order to provide the needed for the end user functionality. (i) Search in portal: it provides an easy and user friendly way for navigation throughout the platform. (ii) Posting content: it provides an easy way to place/post comments on the forum, and/or there is the option to upload extra educational material. (iii) Creating an account: a new user will be in position to create an account and become a registered user. (iv) Calendar of events: users are in position to organize an event and even invite others. (v) Alerts: a user can make use of the alerts feature so as to receive useful notifications. (vi) Evaluation tests: it provides the option for the users to take by using the platform. (vii) Video conference: the platform users have the opportunity to communicate with each other using audio and/or video means. The suggested solution for the platform, also meeting the identified need, is the BigBlueButton.

External services consist of applications which ProMe uses in order to deliver personalized services to any external users. (i) Web interface: a web interface is considered as the mediator between the actual user and any software that exists on a web server. (ii) Web services: web service is

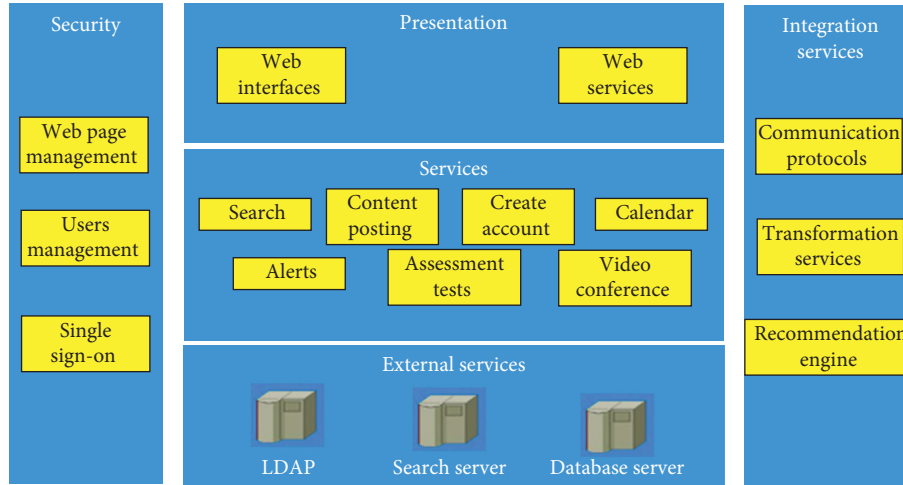


FIGURE 2: System architecture.

considered to be a software system, supporting interoperable machine-to-machine interaction over a network. (iii) Integration services: integration services include all those extra services that the ProMe platform needs to communicate with any external services. (iv) Communications protocol: communication protocols deal with all the exchanging messages. (v) Transformation services: data transformation services (or DTS) is a set of objects and utilities to allow the reading and writing to the system's databases. The sub-components of the DTS objects are called DTS tools. (vi) Management of web pages: administrators of the systems have the right to manage the content of the web page. (vii) Single sign-on (SSO): Single sign-on is supported thus allowing the users of the platform to make use of the same credentials to access different applications supported by the platform. (viii) Users profile management: it provides the rights for administrators to enable or disable a user account.

5. Recommendation System

Building a highly adaptive recommendation system is crucial for our system. The system will be in position to provide recommendations to the users based on historical data, taking into account each user's personal characteristics and needs. Some information about adaptive recommendation systems can be found in [23]. The very basic ingredients for the development of a more precise recommendation system are as follows: trust, similarity, and reputation. When we are talking about trust, what in essence we are trying to capture is the user's trustworthiness. When dealing with similarity, capturing the level of proximity amongst users is the target. And when discussing about reputation, what other users are thinking about a specific use is targeted. A system which can make use of those three ingredients will be in position to provide accurate recommendations about services that someone using the platform is interested for; for example, a user that is in position to give insight to others based on his/her experience. The three aforementioned ingredients can possibly give answers to the next questions. (i) How valid is a review provided by a user?

(ii) Can, for example, a mentor address mentee's needs? (iii) What the online community thinks of him/her?

ProMe is in position to provide accurate (to the platform user) recommendations. This is possible by implementing algorithms to adequately cope with the vast amount of data coming from Internet, by implementing customized algorithms for analyzing the incoming data, and by designing and implementing the needed for the user behavior analysis and recommendation algorithms. Figure 3 provides an insight on the abstract architecture/functionality of ProMe's recommendation engine. By simply following the data flow, anyone easily understands the interactions taking place among the various subcomponents. Initially, a user is asking for possible recommendations with the use of a query, which is then passed to the actual recommender component. Based on historical user activities, the recommender component "calculate" the following ingredients: (i) trust, (ii) reputation, and (iii) similarity. Based on the above "calculations," the ProMe platform is in position to provide the end user with the needed recommendations.

Community structure is the base for ProMe's recommendation engine. The central idea of the community structure is that a community can be depicted as a graph. In this graph, users and items are considered to be the nodes, and on the edges, their relationships are placed. In this way, we can easily weight the edges of the graph, and we can also add several features to them, for example, creation time and number of interactions with other nodes.

ProMe is proposing the "walking" across the edges of the graph. So, the level of their trust is in essence the actual length of a walk between two nodes. In the case of no path among two nodes, recommendation can be based on reputations. Figure 4 presents the actual model.

A trust relation from node i to node j indicates how much trust i places in j . Considering the recommendation problem in trust networks, such as ProMe, is how can we recommend (predict) a trustworthy mentor (i.e., i) to a mentee (i.e., j). Let us start with a simple case where all nodes (edges) have the same trust value 1. For example, Figure 5 illustrates a trust network (TN1), which contains the

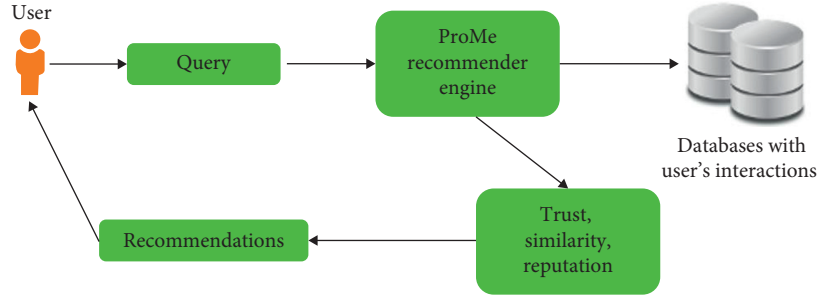


FIGURE 3: ProMe recommendation system.

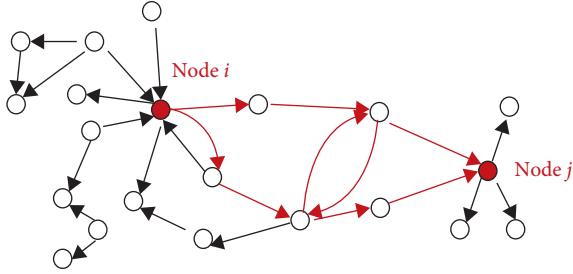


FIGURE 4: The model of ProMe recommendation engine.

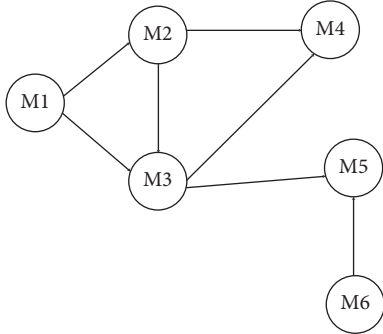


FIGURE 5: A trust network with no edge weights.

neighbours of mentor M1. Among all the agents except M1's neighbours M2 and M3, M4 is the most possible candidate because it is connected by two neighbours of M1.

Figure 6 presents an example of a trust network (TN2), which shares the same topology as TN1 except TN2 has trust values (rather than 1) at its nodes (edges). Unlike the result in Figure 5, although M5 has fewer connections with M1's neighbours than M4, M5 has the highest similarity score because the trust value of its only connection is much stronger than the trust values of M4's connections. Note that M3 (not considered as a recommendation because it is already a neighbour of M1) also has a high similarity score because it is connected by M2 (M1's neighbour) with a high trust value.

5.1. Mining and Machine Learning Aspects. In the machine learning field, complexity and diversity can be handled using the "Black Box" principle. Based on this principle every machine learning method is expected to fit a simple mold.

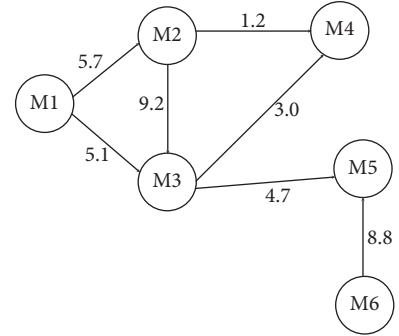


FIGURE 6: A trust network where edge weights represent trust values.

The ProMe recommendation engine/component will be responsible to gather any data coming out of the queries and constructing the actual dataset to be used for training and testing the machine learning algorithm. The recommendation engine/component supports the following three procedures:

- (1) To establish a safe connection with the actual databases
- (2) To query the databases and receive the data
- (3) To save the data in a proper, for the machine learning algorithm, format

Despite the fact that each algorithm is different from the others, the following five steps should be followed throughout the development and application phase:

- (1) Data collection: data collection refers to the actual method to be followed in order to collect the data. Data can be collected with the use of APIs, RSS feed, through (sensor) devices that can collect and transmit the data, etc.
- (2) Data preparation: data preparation is one of the most important steps. Data should be in a suitable format to be fed to an algorithm. Algorithms may need the provided features to follow a specific format. Some work only with features and/or variables in string format, while others ask for data to be integers.
- (3) Training the algorithm: in the training the algorithm step, we provide the algorithm with the data from the previous step, and as an outcome, we expect to gain a

much deeper and extended insight into the provided data. In the case of unsupervised learning and since we do not have a specific value to target, no training step takes place.

- (4) Testing the algorithm: In the testing the algorithm, the evaluation of the algorithm is taking place. When we are referring to supervised learning, someone usually has known values to evaluate the algorithm. For example, ground truth data to be checked against the performance of the algorithm. Support and confidence should be used in unsupervised learning, in order to check if it is successful or not.
- (5) Usage: in this final step, we check the algorithm's implementation in practice. Of course, there is always the need to continuously check if the aforementioned four steps are delivering as expected.

5.2. Recommendation System Testing. The recommendation system is the system responsible for the three from the end steps in the previously described sequence of actions (steps 3 to 5). The recommendation system is providing the whole framework with the needed functionality in order to (i) receive the data, (ii) to feed the machine learning algorithm, and (iii) to provide a better insight for them. The ultimate target for a machine learning algorithm is to spot any interesting relationships when working with the dataset. But how do we quantify interesting relationships? First, we should find a frequent itemset, and second, we should measure any interesting relationships in association rules.

Algorithms, like Apriori [24], can be used to enhance the recommendation system with extra association rules, thus providing an even more robust system. Apriori is based on the principle that if an item is not frequent, that automatically suggests that supersets containing that specific item are not frequent too (Algorithm 1). What plays a vital role on how to test and measure is the user's population. The bigger the dataset, the bigger the quality which can be achieved.

6. Conclusions

The basic scope of the ProMe project was to design and eventually bring to life a platform which will be in position to provide the needed tools so as mentoring and of course knowledge sharing among older and younger generations to be feasible. Another target of the framework is to assist older adults throughout the transition phase, from full-time employment to retirement. Thus supporting their mental health and quality of life. Last but not least, to provide equal importance is the continuity in knowledge and experience passing from the older ones to the younger generation. Results till now show that users consider the platform useful and easy to use.

ProMe is in position to provide accurate (to the platform user) recommendations. This is possible by implementing algorithms to adequately cope with the vast amount of data coming from Internet, by implementing customized algorithms for analyzing the incoming data, and by designing and implementing the needed for the user behavior analysis

```

(1)  $C_k$ : Candidate itemset of size  $k$ 
(2)  $L_k$ : frequent itemset of size  $k$ 
(3)  $L_1 = \{\text{frequent items}\}$ ;
(4) for ( $k = 1; L_k \neq \emptyset; k++$ ) do begin
(5)    $C_{k+1} = \text{Candidate generated from } L_k$ ;
(6)   for each transaction  $t$  in database do
(7)     increment the count of all candidates in
(8)      $C_{k+1}$  that are contained in  $t$ 
(9)      $L_{k+1} = \text{candidates in } C_{k+1} \text{ with min\_support}$ 
(10)  end
(11)  return  $\cup_k L_k$ ;

```

ALGORITHM 1: The Apriori algorithm.

and recommendation algorithm. Initially a user is asking for possible recommendations with the use of a query, which is then passed to the actual recommender component. Based on historical user activities, the recommender component “calculate” the following ingredients: (i) trust, (ii) reputation, and (iii) similarity. Based on the above “calculations” the ProMe platform is in position to provide the end user with the needed recommendations.

ProMe is now a working online service platform for mentoring, which focuses on intergenerational cooperation. It enables adults after retirement to pass their professional formal and tacit knowledge to younger adults, who can benefit from this valuable asset, e.g., when starting their own career. Mentoring is a win-win situation for all parties involved. The mentor has the opportunity to help changing another person's life by sharing knowledge and experience. The mentee can be sure to be supported in walking towards a self-defined goal by someone experienced and professional. The ProMe platform provides a matching system that connects mentors/mentees accordingly to their needs and expectations. A variety of communication and collaboration tools are offered that support mentor/mentee to successfully work together.

Data Availability

The data are coming from an insurance company in Greece, which was really happy to assist in the validation of the proposed architecture/platform. Unfortunately, the data cannot be available due to their nature. The data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, (6/12 months) after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

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

Acknowledgments

This research was enabled by the ProMe project. The financial support by the AAL JP and the support of all involved parties in the project are gratefully acknowledged.

References

- [1] ProMe website, 2018, <http://pro-me.eu/>.
- [2] R. Schaap, A. de Wind, P. Coenen, K. Proper, and C. Boot, "The effects of exit from work on health across different socioeconomic groups: a systematic literature review," *Journal of Social Science & Medicine*, vol. 198, pp. 36–45, 2018.
- [3] P. Eibih, "Understanding the effect of retirement on health: mechanisms and heterogeneity," *Journal of Health Economics*, vol. 43, pp. 1–12, 2015.
- [4] H. Schmitz, "Why are the unemployed in worse health? The causal effect of unemployment on health," *Journal of Labour Economics*, vol. 18, no. 1, pp. 71–78, 2011.
- [5] M. Bertoni, G. Brunello, and G. Mazzarella, "Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers," *Journal of Health Economics*, vol. 58, pp. 215–227, 2018.
- [6] T. Müller and M. Shaikh, "Your retirement and my health behavior: evidence on retirement externalities from a fuzzy regression discontinuity design," *Journal of Health Economics*, vol. 57, pp. 45–59, 2018.
- [7] O. Shai, "Is retirement good for men's health? Evidence using a change in the retirement age in Israel," *Journal of Health Economics*, vol. 57, pp. 15–30, 2018.
- [8] H. T. M. Nguyen, "Theories of mentoring," in *Models of Mentoring in Language teacher Education*, Springer Nature, Basel, Switzerland, 2017.
- [9] T. Kerry and A. S. Mayes, *Issues in Mentoring*, The Open University, Milton Keynes, UK, 2014, ISBN: 9781136159343.
- [10] D. Rousseau, "Psychological contracts in organizations," in *Understanding Written and Unwritten Agreements*, Sage Publications, Thousand Oaks, CA, USA, 1995.
- [11] D. L. Haggard and D. B. Turban, "The mentoring relationship as a context for psychological contract development," *Journal of Applied Social Psychology*, vol. 42, no. 8, pp. 1904–1931, 2012.
- [12] D. M. Rousseau, "Psychological and implied contracts in organizations," *Employee Responsibilities and Rights Journal*, vol. 2, no. 2, pp. 121–139, 1989.
- [13] J. P. Kotter, "The psychological contract: managing the joining-up process," *California Management Review Spring*, vol. 15, no. 3, pp. 91–99, 1973.
- [14] Horseshmouth Website, 2018, <http://www.horseshmouth.co.uk/index.publisha>.
- [15] Senior Engage Website, 2018, <http://seniorengage.eu/>.
- [16] Aware Website, 2018, <http://aware.ibv.org/en/home/>.
- [17] E-Mentoring Website, 2018, <http://www.e-mentoring.org/>.
- [18] Hear Me Website, 2018, <http://www.viauc.com/projects/>.
- [19] Horizons Unlimited Website, 2018, <http://horizonsunlimited.com.au/>.
- [20] Mentor Website, 2018, <http://www.mentoring.org/>.
- [21] Mentor Net Website, 2018, <http://www.mentornet.net/>.
- [22] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architecture," *ACM Transactions on Internet Technology (TOIT)*, vol. 2, no. 2, pp. 115–150, 2002.
- [23] C. Chrysoulas and M. Fasli, "Building an adaptive E-learning system," in *Proceedings of 9th International Conference on Computer Supported Education*, Porto, Portugal, April, 2017.
- [24] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings 20th International Conference on Very Large Data Bases, VLDB*, pp. 487–499, Santiago de Chile, Chile, September 1994.

Research Article

User Evaluation of the Smartphone Screen Reader VoiceOver with Visually Disabled Participants

Berglind F. Smaradottir ¹, Jarle A. Håland ² and Santiago G. Martinez ³

¹Department of Information and Communication Technology, University of Agder, Grimstad N-4879, Norway

²Kongsgård School Centre, Kristiansand N-4631, Norway

³Department of Health and Nursing Science, University of Agder, Grimstad N-4879, Norway

Correspondence should be addressed to Berglind F. Smaradottir; berglind.smaradottir@uia.no

Received 3 August 2018; Revised 30 October 2018; Accepted 11 November 2018; Published 2 December 2018

Guest Editor: Giuseppe De Pietro

Copyright © 2018 Berglind F. Smaradottir et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Touchscreen assistive technology is designed to support speech interaction between visually disabled people and mobile devices, allowing hand gestures to interact with a touch user interface. In a global perspective, the World Health Organization estimates that around 285 million people are visually disabled with 2/3 of them over 50 years old. This paper presents the user evaluation of VoiceOver, a built-in screen reader in Apple Inc. products, with a detailed analysis of the gesture interaction, familiarity and training by visually disabled users, and the system response. Six participants with prescribed visual disability took part in the tests in a usability laboratory under controlled conditions. Data were collected and analysed using a mixed methods approach, with quantitative and qualitative measures. The results showed that the participants found most of the hand gestures easy to perform, although they reported inconsistent responses and lack of information associated with several functionalities. User training on each gesture was reported as key to allow the participants to perform certain difficult or unknown gestures. This paper also reports on how to perform mobile device user evaluations in a laboratory environment and provides recommendations on technical and physical infrastructure.

1. Introduction

Since the last decade, touchscreen technology has been increasingly used not only across multiple types of devices, such as smartphones and tablets [1–3], but also in photocopying machines, automated teller machines (ATMs), and ticket machines in bus, railway stations, and airports. Reviews from the perspective of human factors and ergonomics and studies of people with developmental disabilities pointed out the relevance of the specific context of system interaction in order to maximize safety, performance, and user satisfaction [4] and the need for more research [5]. Touchscreens require the use of fingers and a choreography of gestures for interaction between the user and the device's user interface (UI) [6, 7]. However, this type of screen interaction can represent a challenge for visually disabled users where the screens are designed for a visual feedback while using the system [8].

The World Health Organization (WHO) estimates that the number of people with visual disability is around 285 million globally and that about 2/3 of them are older than 50 years [9, 10]. Traditionally, visually disabled people have used different assistive technology devices, such as an external keyboard, a braille terminal, or a screen reader that provides speech feedback related to the visual elements on the screen. Mobile phones with physical buttons are still functional for many visually disabled people because of the surface and the rugosity of the buttons that provide palpable guidance when using the device. However, this type of communication device has become less popular in favour of smartphones with touchscreens that currently dominate the market. Smartphones with touchscreen interaction do mainly incorporate visual and sound feedback for communication with the user. This type of communication represents a challenge for the UI navigation to visually disabled people who do not see the screen with sufficient details and buttons without tactile

feedback [11]. Several solutions are available in the market to improve the accessibility of smartphone technology for visually disabled people [12–14]. Some of these solutions are standalone products, and others are used in conjunction with other technology. One of the products available is VoiceOver [12], the integrated screen reader in Apple Inc. products. VoiceOver allows users to interact with the UI through gestures and with speech feedback to guide the navigation. The screen reader has been included in Apple Inc. products since April 2005 in Mac OS X 10.4, since June 2009 in iPhone 3GS OS 3.0, and in iPad OS 3.2 since its introduction in April 2010. VoiceOver has to be activated in the device's settings, and when activated, the device provides a speech feedback when a user interacts using hand gestures on the touchscreen. There are different gestures that can be performed on the UI, and they provide immediate feedback interpreted by the screen reader. For instance, tap with one finger and drag will read the item in the cursor (selected), and four-finger tap near the top of the screen will read the first item at the top. The gestures must be made with the fingers, and the screen reader does not respond to voice commands or sense motion.

In this context, the research project “Visually impaired users touching the screen—A user evaluation of assistive technology” aimed at evaluating the accessibility and usability of a screen reader for touchscreens in smartphones [15]. This paper presents the results from the evaluation of the usability and the accessibility of the screen reader VoiceOver (iOS 7.1.2), which is an integrated functionality in iPhone mobile devices. In addition, the paper provides recommendations on technical and physical infrastructure to perform an evaluation of mobile devices in a laboratory environment.

The three research questions (RQs) targeted by this study were as follows:

RQ1: What is the user experience of visually disabled users when interacting with the VoiceOver?

RQ2: How is the VoiceOver screen reader response to a set of 16 performed hand gestures during a user evaluation?

RQ3: What technical infrastructure can be suitable for an evaluation of mobile assistive technology with visually disabled users?

Following this introduction, the research methodology and the technical test infrastructure are described. The results are presented based on the user evaluation outcomes and experience related to the test infrastructure. Furthermore, a discussion of the main results is provided followed by a summary of the research contributions and conclusions.

2. Materials and Methods

A mixed methods research approach was employed in the evaluation of the screen reader [16–18], with quantitative and qualitative measures. The evaluation was conducted in three phases: (1) individual user training at the participant's home and introduction to the gestures a few days before the test, supplied with a written instruction sent by e-mail;

(2) a usability test in a controlled laboratory environment including a pretest interview for collecting participant background information; and (3) a posttest interview for qualitative analysis of the test output. The research team had three members whose background was health technology, educational training with assistive technology, and clinical practice. All research team members had professional experience in working with people with visual disabilities.

In the initial preparation of the study, phone interviews were made with three key informants with expertise in visual disabilities, who worked at the Norwegian State Agency for Special Needs Education Service (StatPed) [19]. The goal of the interviews with the key informants was to gather insights on assistive technology for visually disabled people. Based on the interviews, a pilot test of the evaluation was prepared with a comparison of Android and Apple tablet devices. Two voluntary members from the Norwegian Association of the Blind and Partially Sighted [20] participated in the pilot test, running several tasks. Afterwards, a focus group interview was conducted in order to better understand the interactions and any of the problems that the users found. In the phone interviews and also in the pilot test, the informants explained that their experience was that the smartphone iPhone was the most commonly used and preferred device among their peers, also visually disabled people. Based on that information, an iPhone 4 (iOS 7.1.2) device was chosen for the study (the device can be seen in Figure 1) because it was widely available and had the VoiceOver screen reader integrated. The tasks were inspired by the standard gestures' descriptions in the VoiceOver guide manual [21].

2.1. Recruitment of Participants. The recruitment of participants was made in collaboration with the Norwegian Association of the Blind and Partially Sighted [20]. In addition, the professional network of one of the researchers with expertise in teaching and user training of assistive technology was used to support the recruitment process. The first contact made with the participants was a phone conversation to inform them about the study. The second contact was an e-mail with information about the study and a consent form to be signed by each participant. Six visually disabled people were recruited to participate in the user evaluation, see Table 1 for distribution of participants. They had a mean age of 42.8 years and an average of 1.9 years of user experience with VoiceOver. All the participants had previous experience with using a screen reader for desktop and/or laptop computers.

2.2. Test Procedure. In the first phase of the evaluation, each participant had individual user training at home (Figure 2) on 16 specific hand gestures for screen interaction. The individual user training lasted 15–30 minutes (with an average of 21.7 minutes), led by a member of the research team. The gestures that a user knew in advance and which ones were learned during the training were registered during the training session.

The second phase was executed in a usability laboratory. One of the researchers acted as the moderator and sat down



FIGURE 1: The smartphone used in the test.

TABLE 1: The background of the test participants.

Participants $n = 6$	Age	Gender	Device familiarity	Years of VoiceOver use	Self-graded skill
1	60	Female	iPhone	1.5	Medium
2	31	Male	iPhone	1.5	Advanced
3	27	Female	iPhone	1	Medium
4	30	Male	iPod	3	Advanced
5	48	Female	iPhone	2	Medium
6	61	Male	iPhone	2.5	Advanced

beside the test participant. The participants were informed about the subsequent test and signed a consent form before the test began. Demographic information and user experience with specific technical devices were also collected. Each user evaluation followed the same test plan, with a set of 16 tasks related to the use of gestures for touchscreen interaction. The moderator guided through the tasks and asked the participants to speak out loudly during the task solving (Figure 3) following a think aloud protocol [22–24].

The task solving was followed by a posttest individual interview (third phase). The participants were asked to score the gesture performance and task solving, choosing among three categories: “easy,” “medium,” or “difficult.” In addition, problems or obstacles observed or reported were discussed. The interviews also covered the general user experience with the smartphone and the first-time use of the VoiceOver.

Each test session (second and third phases) lasted between 90 and 120 minutes, and a total of six test sessions were run across three separate days.

2.3. Technical and Physical Test Infrastructure. The evaluation was executed in the usability laboratory at the Centre for eHealth of the University of Agder, Norway [25]. The usability laboratory consisted of two rooms; one test room and one control room, connected through a one-way mirror with visualisation towards the test room. In the test room, the moderator was placed together with a test participant, and in the control room, two observers followed the test from monitors and directly through the one-way mirror. The technical and physical infrastructure is described in Figure 4.



FIGURE 2: User training of VoiceOver gestures at a participant's home.



FIGURE 3: The moderator (left) guiding a participant (right) through the task solving in the test room.

For replicability and information purposes, the technical material and equipment used during the study are presented below grouped by rooms.

Test room:

- (i) Apple Inc. iPhone 4 MD128B/A iOS 7.1.2 with VoiceOver activated
- (ii) Fixed camera: Sony BRCZ330 HD 1/3 1CMOS P/T/Z 18x optical zoom (72x with digital zoom) colour video camera
- (iii) Portable camera: Sony HXR-NX30 series
- (iv) Apple Inc. iPad MD543KN/A iOS 8.1 for additional sound recording
- (v) Sennheiser e912 condenser boundary microphone
- (vi) Landline phone communication

Control room:

- (i) Stationary PC: HP Z220 CMT workstation, Intel Core i7-3770. CPU @ 3.4 GHz, 24 GB RAM, Windows 7 Professional SP1 64 bit
- (ii) Monitor: 3x HP Compaq LA2405x
- (iii) Remote controller: Sony IP Remote Controller RM-IP10
- (iv) Streaming: 2x Teradek RX Cube-455 TCP/IP 1080p H.264
- (v) Software Wirecast 4.3.1
- (vi) Landline phone communication

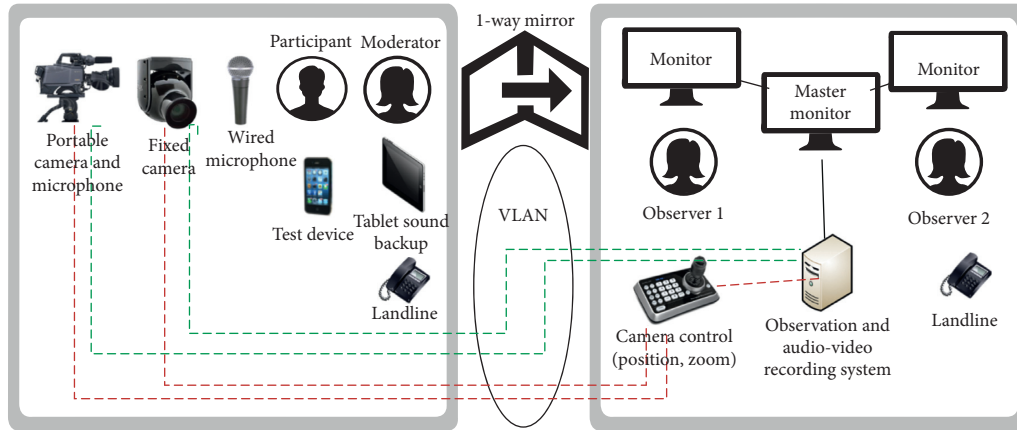


FIGURE 4: The technical and physical test infrastructure.

2.4. Data Collection. The test sessions were audio-visually recorded in a F4V video file format. The recordings from two audio-visual sources were merged into one video file using the software Wirecast v.4.3.1 [26], with multiple video perspectives and one single audio channel. The files were exported to the Windows Media Video (WMV) format and then imported to the qualitative software tool QSR NVivo 10 [27]. The recordings were transcribed verbatim and categorized for a qualitative content analysis [28]. Quantitative measurements of the time and number of attempts in the task solving were made as a part of the analysis of the recordings. In addition, the research team made annotations during the test sessions that were included in the data collection (Figure 5).

2.5. Ethical Approval. The Norwegian Centre for Research Data [29] approved this study with the project number 40636. All participants received verbal and written information about the project and confidential treatment of their collected data. They were informed that their participation was voluntary, and each participant signed a consent form. The participants were aware that they could withdraw at any time without reason. In that case, their data would be consequently withdrawn and deleted. For health and safety reasons, each test participant was thoroughly informed about the physical environment before entering the test room and the participants were never left alone in the laboratory facilities.

3. Results

All six participants went through the laboratory test. The test results are presented divided into three categories: user training, quantitative metrics from the user tests, and qualitative outcome of the posttest interviews.

3.1. Pretest User Training. The familiarity with the VoiceOver gestures registered in the user training is presented in Table 2. The registration showed that all participants knew the double tap gesture (number 4) and three-finger flick to the left or right (number 10). 5 out of 6 were familiar with the



FIGURE 5: The control room showing the visual access to the test room through the one-way mirror.

one-finger tap gestures (numbers 1–3). For gesture numbers 6 and 7, the four-finger tap at the top or the bottom of the screen, 5 out of 6 participants did not know them in advance.

3.2. User Evaluations. The quantitative measurements from the user evaluations are presented in Table 3, separated in six columns. The first column describes the 16 VoiceOver standard gestures that were used to solve the associated task. The tasks are described in the second column. The third column displays the average number of attempts needed for the task solving. The fourth column shows the task solving average time that was used, measured in seconds. The fifth column presents the system response to the gesture interaction differentiated in the categories “consequent” and “inconsequent” speech feedback. Consequent speech feedback refers to sufficient and adequate information in the system response and inconsequent feedback to insufficiency or lack of information in the system response. In usability studies, the task accuracy is often categorized into completed or not completed task [23, 30]. In this particular test, there was an additional variable related to the task performance, which was the feedback that the system provided when a participant performed a specific action. The categories chosen were therefore “consequent feedback” or “inconsequent feedback” to the specific hand gesture performed. The “consequent feedback” referred to the system appropriately

TABLE 2: Familiarity per participant with the VoiceOver gestures in the pretest user training.

Gesture	P#1	P#2	P#3	P#4	P#5	P#6
(1) Tap with one finger, lift and tap again	Yes	Yes	Yes	Yes	Yes	Yes
(2) Tap with one finger and drag	Yes	Yes	Yes	Yes	Yes	Yes
(3) Tap with one finger and swipe to right or left	Yes	Yes	No	Yes	Yes	Yes
(4) One-finger double tap	Yes	Yes	Yes	Yes	Yes	Yes
(5) Split-tap: touch one finger and then tap with a second finger	No	Yes	No	Yes	No	No
(6) Four-finger tap at the top of the screen	No	No	No	No	No	Yes
(7) Four-finger tap at the bottom of the screen	No	No	No	No	No	Yes
(8) Two-finger flick up	Yes	Yes	No	Yes	No	No
(9) Two-finger flick down	Yes	Yes	No	Yes	No	No
(10) Three-finger flick to the left or right	Yes	Yes	Yes	Yes	Yes	Yes
(11) Three-finger tap	No	Yes	No	Yes	No	No
(12) Two-finger rotate	Yes	Yes	No	Yes	Yes	Yes
(13) Flick up and down with one finger	Yes	Yes	No	Yes	No	Yes
(14) Three-finger double tap	Yes	No	Yes	Yes	Yes	Yes
(15) Three-finger triple tap	Yes	Yes	No	Yes	Yes	Yes
(16) Two-finger double tap	Yes	Yes	No	Yes	Yes	Yes

P = participant.

TABLE 3: Quantitative metrics of the user evaluations.

Gesture	Task	Average no. of attempts	Average time for task solving in seconds	System speech feedback	Specification
(1) Tap with one finger, lift and tap again	Speak the item in the cursor: find the app Map	24.2	28	Consequent	
(2) Tap with one finger and drag	Speak the item in the cursor: find the app Clock	14.3	13.8	Consequent	
(3) Tap with one finger and swipe to right or left	Speak the item in the cursor: find the app Calendar	22	16.5	Consequent	
(4) One-finger double tap	Open the app Calendar	1	1	Consequent	
(5) Split-tap: touch one finger and then tap with a second finger	Open the app Weather	1	1	Consequent	
(6) Four-finger tap at the top of the screen	Read the item at the top	3.7	16.8	Consequent	
(7) Four-finger tap at the bottom of the screen	Read the item at the bottom	9.2	27	Consequent	
(8) Two-finger flick up	Read the current page starting at the top	1.2	1.2	Consequent, except once	The screen reader did not read
(9) Two-finger flick down	Read from the cursor to the end of the current page	1	1	Consequent	
(10) Three-finger flick to the left or right	Change to the next page in the start screen and back	2.8	6.7	Inconsequent	The screen reader did not consequently read the next page
(11) Three-finger tap	Read where the cursor is	1.8	3.5	Inconsequent	The screen reader did not read the cursor (application's name)
(12) Two-finger rotate	Rotor: find the setting for the speed of the speech feedback	10.3	13.3	Consequent	
(13) Flick up and down with one finger	Rotor: adjust the speed of the speech feedback	1.5	2.8	Consequent	
(14) Three-finger double tap	Mute VoiceOver	1.2	1.7	Consequent	
(15) Three-finger triple tap	Turn the screen curtain on	1.3	2.8	Consequent	
(16) Two-finger double tap	Terminate a phone call	3.6	1.8	Inconsequent	The system did not terminate the phone call

providing feedback that corresponded to the hand gesture performed by a participant. The “inconsequent feedback” referred to a system feedback that did not correspond to the

hand gesture performed by a participant or absence of any feedback. The sixth column specifies the type of inconsequent response occurred.

TABLE 4: The grading of the task solving made by the participants in the posttest interview ($n = 6$).

Gesture	Associated task	Easy	Medium	Difficult
(1) Tap with one finger, lift and tap again	Speak the item in the cursor: find the app Map	4	1	1
(2) Tap with one finger and drag	Speak the item in the cursor: find the app Clock	5	1	
(3) Tap with one finger and swipe to right or left	Speak the item in the cursor: find the app Calendar	6		
(4) One-finger double tap	Open the app Calendar	6		
(5) Split-tap: touch one finger and then tap with a second finger	Open the app Weather	4	2	
(6) Four-finger tap at the top of the screen	Read the item at the top	3	1	2
(7) Four-finger tap at the bottom of the screen	Read the item at the bottom	4	1	1
(8) Two-finger flick up	Read the current page starting at the top	3	2	1
(9) Two-finger flick down	Read from the cursor to the end of the current page	6		
(10) Three-finger flick to the left or right	Change to the next page in the start screen and back	5	1	
(11) Three-finger tap	Read where the cursor is	5	1	
(12) Two-finger rotate	Rotor: find the setting for the speed of the speech feedback	5	1	
(13) Flick up and down with one finger	Rotor: adjust the speed of the speech feedback	5	1	
(14) Three-finger double tap	Mute VoiceOver	6		
(15) Three-finger triple tap	Turn the screen curtain on	6		
(16) Two-finger double tap	Terminate a phone call	3	1	2

The performance of three different one-finger tap gestures (tasks 1–3) for speaking the item in the cursor required many attempts to succeed. The system response was consequent. The double tap and slit-tap gestures (tasks 4–5) were easy and fast to perform for the participants. The gesture four-finger tap at the top and bottom of the screen (tasks 6–7) were reported as technically difficult to perform by the participants, which was also indicated by the time for the task solving. The gestures two-finger flick up and down to read the page from top or bottom (tasks 8–9), were easy to perform and showed consequent speech feedback. The three-finger flick and tap gestures (tasks 10–11) were reported as easy to perform, but there was inconsequent system response related to insufficiency in the speech feedback when trying to inform about the current page. For the rotor-related tasks, 12 and 13, two of the participants needed several attempts (7 and 41) for finding the rotor settings, but adjusting the speed of the speech feedback was easier. The gestures three-finger double and triple tap (tasks 14–15) were easy to perform and with a quick task solving. The two-finger double tap in task 16, to terminate a phone call, was easy to perform but there was inconsequent feedback from the system and the phone call was not terminated in three out of six tests.

3.3. Posttest Interviews. The participants graded the performance of gestures and task solving (Table 4) during the individual posttest interview.

Five of the gestures in the task solving were categorized “easy” to perform, such as the one-finger double tap and the three-finger double and triple taps. Six gestures were categorized as “easy” or “medium,” such as the one-finger flick up and down and three-finger tap. There were gestures that were categorized as “difficult” by two participants, such as the four-finger tap at the bottom and the top of the screen and the two-finger double tap. The task for the two-finger double tap was termination of a phone call, and in the interviews, the participants confirmed that during the test

but also in general, the gesture was associated with inconsistency from the system. For the rotor-related gestures, one participant emphasised the importance of user training to succeed with the specific use of the rotor function.

Regarding the first-time user experience, all participants needed user training to be able to start using the smartphone and for activation of the screen reader VoiceOver. Three had family or friends that helped them with the first-time use: one went to a course organized by the Norwegian Association of the Blind and Partially Sighted and two found it out by themselves explaining that VoiceOver as such provides user training and guidance by informing about which gesture to perform for an action. Four participants stated: *It was a bit complicated with first-time set up of the new phone with apple-id and activation of VoiceOver, besides that it is easy to use. [...] After user training, when I understood how the system worked, I found it easy to use. [...] The functions make sense, and there is a logical structure. [...] It was terrible in the beginning, because I knew none of the gestures and I wanted to throw the phone away, but the price stopped me from doing it ... now I find it fantastic!*

Two participants highlighted the benefits of the smartphone: *I like that I can buy it myself in the store, I did not need to apply for and receive assistive technology from the municipal services. [...] This is the first device I use with built-in accessibility, as the screen reader is included.*

Two participants described how the use of the screen reader had increased their self-management: *I feel more included in the society, now I can use the Internet and check the same apps as other people do, such as Facebook, weather forecast and reading news. [...] It is a feeling of freedom when the phone can read messages for you when you are outdoors, before I had to ask people I did not know about reading from the screen if I received a message, I can now manage it myself and that is a new world for me.* In addition, one participant expressed: *VoiceOver has made my life much easier and I have become much more independent. Everyone with a visual impairment should use a phone with it.*

However, user text input with the VoiceOver keyboard was reported as complicated by four participants, and, for this reason, those participants preferred to use an external keyboard. Another participant stated: *It was hard in the beginning with the virtual keyboard, but with some training I overcame the difficulties.* Five participants told that they preferred to use at home a desktop or laptop computer with reading list because the text input was quicker than in the smartphone and relying on the latter when they were out of home. Two participants expressed that it was easier for them to navigate on a small screen when compared to a larger tablet screen.

4. Discussion

This paper has presented a user evaluation of the Apple screen reader VoiceOver (iOS 7.1.2) with six visually disabled participants. The aim was to identify challenges related to the performance of the standard VoiceOver gestures and evaluate the associated system response. Considering the sensory limitation of the target user group, the screen reader was expected to be intuitive with an optimal presentation of the functionality and distribution of the UI. The study showed that most of the gestures were easy to perform for the participants; however, some gestures were unfamiliar to the participants, especially those connected to the rotor function. The possibility of receiving individual user training before the evaluation was an advantage to succeed with the practical use of those gestures. The system appropriately responded to the users' hand gestures, but inconsistent responses and lack of information were reported in the two-finger flick up, three-finger flick to the left or right, three-finger and double-finger taps. The three research questions (RQs) formulated at the beginning of this paper are answered below based on the results from the study.

RQ1 asked about the user experience when interacting with the VoiceOver. The user experience with VoiceOver in general was positive, as the function was described to increase the self-management and support independence. Most of the gestures were both reported and observed as easy to perform, with some exceptions. The two most difficult ones reported by the participants were the four-finger tap and the two-finger double tap gestures. The gesture made using four-finger tap on the bottom or on the top of the screen to, respectively, read the content of the UI from either side was explicitly reported as difficult to perform.

RQ2 asked about the system response to the 16 hand gestures made on the touchscreen mobile device. The speech feedback appropriately responded during the test with useful information for participants to navigate through the UI, but a few inconsistent responses on correctly performed gestures were registered such as with the two-finger double tap to terminate a phone call. The phone call was terminated correctly only in 3 out of 6 tests and can be considered as a weakness in the system with a negative consequence for the users since speaking on the phone is one of the most frequently used functions. Other user problems identified were related to the gesture made by three-finger flick to the left or

right for swiping between screens where the speech feedback was inconsistent and lacked information.

RQ3 asked about recommended technical infrastructure in evaluations of mobile assistive technology with visually disabled users. A suitable infrastructure would be the one that optimizes the data collection and allows an effective retrospective analysis under more demanding conditions than other user evaluations. In addition, the comfort, safety, and trust of the visually disabled test participants are crucial to avoid interference and distortion with the test results. The described technical and physical infrastructure in Figure 4 serves as an example of a controlled scenario for an evaluation with the same type of technology and participants. The video recordings require a sufficient quality allowing us to zoom in the user interface and the finger interactions in details. A professional software video program is needed to substantially reduce the speed for optimal viewing and retrospective analysis. In addition, the data should be collected with synchronized audio and video signals because streaming over a network usually incorporates latency. The synchronization is of high importance for the retrospective analysis, as the gestures and finger interactions with a mobile device's screen are often made at high speed. Another issue experienced and specific for tests with visually disabled participants was that the sound from the VoiceOver interfered and overlapped with the sound from the test participant and the moderator in the recordings from the table microphone unit. This might complicate the retrospective analysis, and based on that experience, we recommend using several microphones to record the sound sources separately.

This study of the screen reader VoiceOver had some limitations such as the number of test participants ($n = 6$) and tests were conducted only in a usability laboratory setting. However, the number of the participants with a distribution in their ages and smartphone skills meaningfully represented the user group of visually disabled users of smartphones. Other studies have shown that a small number of participants in usability studies can be sufficient for having valid results [31–33]. The laboratory setting allowed the collection of detailed research data under controlled conditions. The collected data material was thoroughly analysed in detail to study the interaction between the visually disabled user and the UI touchscreen. Furthermore, the application of mixed method research, combining laboratory tests with detailed interviews, provided insights into the user experiences, as well as benefits and barriers of using the VoiceOver function.

5. Conclusions

This study was made as a part of the project “Visually impaired users touching the screen—A user evaluation of assistive technology” that aimed at evaluating the usability and accessibility of the screen reader VoiceOver. The main contribution of this study lies in the detailed analysis of the interaction with gestures between the visually disabled participants and the screen reader, preceding the responses from the system. In general, most of the hand gestures were

easy to perform for the participants, although user training played a key role for the understanding and successful performance of specifically complex gestures. Without training, participants could not have been able to perform such gestures. The system response and speech feedback were in most cases correct, but some functionalities of the system might be improved. The results presented are in line with other studies on assistive technologies and visually disabled users [34–36]. The methodological procedures with the use of mixed methods, combining quantitative laboratory test with qualitative interviews and observations, can be recommended to other studies of similar characteristics. The test procedure with user training on the specific hand gestures in advance reduced the memory load in the laboratory test situation, as all the participants were familiar with the gestures and could focus on performing the tasks. The application of a think aloud protocol in the usability laboratory together with posttest interviews is strongly recommended for other studies related to touchscreen assistive technology because they may provide a more comprehensive result.

In terms of future work, it is proposed to validate the laboratory results in the field and address research with a larger sample size focusing on text input and navigation using VoiceOver on a smartphone or tablet device. A comparison between the screen readers VoiceOver from Apple Inc. and TalkBack, which is mainly developed for Android devices, could illustrate differences across different platforms. The integration of VoiceOver in the Apple Watch provides new opportunities of studying user-friendliness and accessibility for visually disabled users. A comparison of the use of VoiceOver on a desktop or laptop computer which are generally more command based could be easily made in a similar usability laboratory. Finally, newer models of iPhone to date, such as 8 and Xs, provide more tactile feedback through vibration during interactions than previous versions and the impact of those functions for visually disabled users would be interesting to evaluate.

Data Availability

The video recordings data used to support the findings of this study have not been made available because of national regulations regarding the privacy of the test participants.

Conflicts of Interest

The authors declare that there are no conflicts of interest with any of the participants, private organizations, or vendors regarding the publication of this paper.

Acknowledgments

The authors would like to thank all the participants in the study and the informants prior to the study for their disinterested contribution. Thanks are due to Hildegunn Mellesmo Aslaksen for professional photographing and also to the Centre for eHealth at the University of Agder, Norway, for close collaboration and adaptation of the test infrastructure facilities.

References

- [1] R. Ling, *The Mobile Connection: The Cell Phone's Impact on Society*, Morgan Kaufmann, Burlington, MA, USA, 2004.
- [2] S. L. Jarvenpaa and K. R. Lang, "Managing the paradoxes of mobile technology," *Information Systems Management*, vol. 22, no. 4, pp. 7–23, 2005.
- [3] G. Goggin, *Cell Phone Culture: Mobile Technology in Everyday Life*, Routledge, Abingdon, UK, 2012.
- [4] A. K. Orphanides and C. S. Nam, "Touchscreen interfaces in context: a systematic review of research into touchscreens across settings, populations, and implementations," *Applied Ergonomics*, vol. 61, pp. 116–143, 2017.
- [5] J. Stephenson and L. Limbrick, "A review of the use of touchscreen mobile devices by people with developmental disabilities," *Journal of Autism and Developmental Disorders*, vol. 45, no. 12, pp. 3777–3791, 2015.
- [6] P. A. Albinsson and S. Zhai, "High precision touch screen interaction," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 105–112, Fort Lauderdale, FL, USA, April 2003.
- [7] A. Butler, S. Izadi, and S. Hodges, "SideSight: multi-touch interaction around small devices," in *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, pp. 201–204, Monterey, CA, USA, October 2008.
- [8] L. Hakobyan, J. Lumsden, D. O'Sullivan, and H. Bartlett, "Mobile assistive technologies for the visually impaired," *Survey of Ophthalmology*, vol. 58, no. 6, pp. 513–528, 2013.
- [9] World Health Organization (WHO), October 2018, <http://www.who.int/mediacentre/factsheets/fs282/en/>.
- [10] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.
- [11] S. K. Kane, J. P. Bigham, and J. Wobbrock, *Fully Accessible Touch Screens for the Blind and Visually Impaired*, University of Washington, Tacoma, WA, USA, 2011.
- [12] Voiceover, October, 2018, <https://www.apple.com/accessibility/iphone/vision/>.
- [13] Window-Eyes, October, 2018, <http://www.windoweyesforoffice.com>.
- [14] TalkBack, October, 2018, https://support.google.com/accessibility/android/answer/6283677?hl=en&ref_topic=3529932.
- [15] B. F. Smaradottir, S. G. Martinez, and J. A. Håland, "Evaluation of touchscreen assistive technology for visually disabled users," in *Proceedings of IEEE Symposium on Computers and Communications (ISCC)*, pp. 248–253, Heraklion, Greece, July 2017.
- [16] J. W. Creswell and V. L. P. Clark, *Designing and Conducting Mixed Methods Research*, SAGE Publications Inc., Thousand Oaks, CA, USA, 2007.
- [17] R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner, "Toward a definition of mixed methods research," *Journal of Mixed Methods Research*, vol. 1, no. 2, pp. 112–133, 2007.
- [18] C. Teddlie and A. Tashakkori, "Mixed methods research," in *The SAGE Handbook of Qualitative Research*, N. K. Denzin and Y. S. Lincoln, Eds., vol. 4, pp. 285–300, SAGE Publications Inc., Thousand Oaks, CA, USA, 2011.
- [19] StatPed, October, 2018, <http://www.statped.no/Spraksider/In-English/temp/Professional-services-and-areas-of-expertise/Expertise-within--the-specialisation-felds-for-special-needs-education/#6.6>.
- [20] The Norwegian Association of the Blind and Partially Sighted, October, 2018, <https://www.blindeforbundet.no/om-blindeforbundet/brosjyrer/muligheter-til-et-aktivt-liv-engelsk>.

- [21] Learn VoiceOver gestures on iPhone, October, 2018, <https://help.apple.com/iphone/12/?lang=en#/iph3e2e2281>.
- [22] M. W. M. Jaspers, "A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence," *International Journal of Medical Informatics*, vol. 78, no. 5, pp. 340–353, 2009.
- [23] A. W. Kushniruk and V. L. Patel, "Cognitive and usability engineering methods for the evaluation of clinical information systems," *Journal of Biomedical Informatics*, vol. 37, no. 1, pp. 56–76, 2004.
- [24] K. A. Ericsson and H. A. Simon, "Verbal reports as data," *Psychological Review*, vol. 87, no. 3, pp. 215–251, 1980.
- [25] Centre for eHealth at University of Agder, October 2018, <https://www.uia.no/en/centres-and-networks/centre-for-ehealth>.
- [26] Wirecast, October 2018, <http://www.telestream.net/wirecast/overview.htm>.
- [27] QSR NVIVO 10, October 2018, http://www.qsrinternational.com/products_nvivo.aspx.
- [28] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, John Wiley & Sons, Hoboken, NJ, USA, 2010.
- [29] Norwegian Social Science Data Services, October 2018, http://www.nsd.uib.no/personvernombud/en/about_us.html.
- [30] J. M. C. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," *International Journal of Medical Informatics*, vol. 79, no. 4, pp. e18–e23, 2010.
- [31] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of ACM Conference on Human Factors in Computing Systems*, pp. 206–213, Amsterdam, Netherlands, April 1993.
- [32] C. W. Turner, J. R. Lewis, and J. Nielsen, "Determining usability test sample size," *International Encyclopedia of Ergonomics and Human Factors*, vol. 3, no. 2, pp. 3084–3088, 2006.
- [33] J. Nielsen, "Why you only need to test with 5 users," Alertbox, October 2018, <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>.
- [34] K. Park, T. Goh, and H. J. So, "Toward accessible mobile application design: developing mobile application accessibility guidelines for people with visual impairment," in *Proceedings of HCI Korea*, pp. 31–38, Hanbit Media, Inc., Seoul, Republic of Korea, 2014.
- [35] B. Leporini and M. C. Buzzi, "Interacting with mobile devices via VoiceOver: usability and accessibility issues," in *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pp. 339–348, ACM, Melbourne, VIC, Australia, November 2012.
- [36] D. McGookin, S. Brewster, and W. Jiang, "Investigating touchscreen accessibility for people with visual impairments," in *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, pp. 298–307, ACM, Lund, Sweden, October 2008.

Research Article

Mobile Hardware-Information System for Neuro-Electrostimulation

Vladimir S. Kublanov , Mikhail V. Babich , and Anton Yu. Dolganov 

Research Medical and Biological Engineering Centre of High Technologies, Ural Federal University, Mira 19, 620002 Yekaterinburg, Russia

Correspondence should be addressed to Vladimir S. Kublanov; kublanov@mail.ru

Received 2 August 2018; Accepted 8 October 2018; Published 21 November 2018

Guest Editor: Giuseppe De Pietro

Copyright © 2018 Vladimir S. Kublanov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The article describes organizational principles of the mobile hardware-informational system based on the multifactorial neuro-electrostimulation device. The system is implemented with two blocks: the first block forms the spatially distributed field of low-frequency monopolar current pulses between two multielement electrodes in the neck region. Functions of the second block, specialized control interface, are performed by a smartphone. Information is exchanged between two blocks through a telemetric channel. The mobile hardware-informational system allows to remotely change the structure of the current pulses field, to control its biotropic characteristics and to change the targets of the stimulation. Moreover, it provides patient data collection and processing, as well as access to the specialized databases. The basic circuit solutions for the neuro-electrostimulation device, implemented by means of microcontroller and elements of high-level hardware integration, are described. The prospects of artificial intelligence and machine learning application for treatment process management are discussed.

1. Introduction

What has the twenty-first century brought to the humanity? Scientific and technological progress in modern society has led to an increase in the duration and improvement of the quality of human life, as well as maintenance of high efficiency and intellectual activity. These processes are taking place at a time of growing mental stress caused by unstable economic development and unpredictable crisis situations, local wars, interethnic conflicts, and natural disasters. The health of the population, which is the basis of the well-being and harmony of human civilization, is deteriorating. The most disturbing is the growth of chronic stress and mental disorders, personality disorders. As a result, a person is losing the ability of efficient information processing, cognitive control, and decision-making, the basic mechanisms of the social version are violated. In the field of neurology and psychiatry, there has been a catastrophic increase in the number of lost years due to movement, coordination, sensitivity, speech, intelligence, and memory disorders [1].

As noted by World Health Organization, among 56.9 million deaths worldwide in 2016, ischemic heart disease and stroke are the world's biggest killers, accounting for a combined 15.2 million deaths. These diseases have remained the leading causes of death globally in the last 15 years. Deaths due to dementias more than doubled between 2000 and 2016, making it the 5th leading cause of global deaths in 2016 compared to 14th in 2000 [2].

Every year, more than 795,000 people in the United States have a stroke. About 87% of all strokes are ischemic strokes, in which blood flow to the brain is blocked. Stroke is a leading cause of serious long-term disability. Stroke reduces mobility in more than half of stroke survivors age 65 and over. Stroke costs the United States an estimated \$34 billion each year. This total includes the cost of health care services, medicines to treat stroke, and missed days of work [3].

The most common approach for treating such diseases is a neuroprotective therapy, which helps normalize and strengthen the physiological activity of brain tissue. During

neuroprotective therapy, medicines are predominantly used. But physiotherapeutic methods of restorative medicine can be also applied [4].

Of all the variety of physical fields and methods, the most promising are spatially distributed fields of monopolar low-frequency current pulses whose structure and characteristics are adequate to endogenous processes in the human body [5–8].

2. Materials and Methods: Multifactorial Neuro-Electrostimulation of Neck Nervous Structures and Organization of Control Process

Personalized multidisciplinary approach to the organization of the patient's treatment process is promising for increase of the neuro-electrostimulation effectiveness. It implies to actively use neuro-electrostimulation in addition to various procedures of neurorehabilitation for triggering mechanisms of neuroplasticity in management of the brain functional processes [9].

The choice of the neck as a target for neuro-electrostimulation is determined by the location in it: centers of segmental control for vital functions (*cervical sympathetic ganglia*) and the conducting pathways of the suprasegmental centers of the homeostasis regulation (*glossopharyngeal* and *vagus nerves* and their branches, as well as the *cervical plexus* of the spinal nerves) [10]. In the deep muscles of the neck there are nodes of the *sympathetic trunk*, formed by the nervous processes of the *autonomic nuclei* of the spinal cord. The *upper*, *middle*, and *lower (stellate) sympathetic nodes* have numerous branches that enable sympathetic innervation of glands, meninges, vessels of the head, neck, and spine. The afferent fibers of the *spinal plexus* located on the posterior surface of the neck pass through the *posterior horns* of the spinal cord and end in the sensitive nuclei of the brainstem and the reticular formation. Near the main arteries of the neck lies the *vagus nerve*. The nuclei of the *vagus nerve* are located in the brainstem and are common to the *glossopharyngeal nerve*. They have extensive connections with the *hypothalamus*, olfactory system, and reticular formation. Together, the *glossopharyngeal* and *vagus nerves* activate parasympathetic innervation of most organs. Nerve formations in the neck are closely related to the *brainstem*, through which they have two-way links to the *pons*, *middle brain*, *cerebellum*, *thalamus*, *hypothalamus*, and *cerebral cortex*. The presence of these relations ensures the involvement of the neck nervous structures in the analysis of sensory stimuli, the regulation of the muscle tone, and autonomic and higher integrative functions [11].

With electrostimulation of the *cervical spinal plexus*, branches of the *vagus nerve*, *nervus accessorius*, and *glossopharyngeal nerve*, the gray matter of the brainstem can be stimulated along the afferent pathways. Through the reticular formation, the effect in this case extends to the *thalamic structures* and the *cerebral cortex*. The stimulation of the nodes of the *sympathetic trunk* makes it

possible to influence both the vascular tone of the cerebral arteries and the autonomic nuclei of the spinal cord. As a result of these actions, it is possible to influence various functional processes in the brain tissues, modulate autonomic processes, and influence motor control and cognitive functions.

The next step in creating a promising neuro-electrostimulator suitable for providing comprehensive rehabilitation is the selection of the best solutions for organizing the architecture of the neuro-electrostimulator, taking into account the characteristics of the conducting pathways of the neck nervous formations.

3. Results: Selection of Basic Circuit, Engineering, Hardware, and Software Solutions

The analysis of tasks that are implemented in modern physiotherapy devices for recovery medicine shows that, as a rule, two tasks are performed in them:

- (1) Formation of a physical field in the problem area of the body
- (2) Regulation of characteristics of the physical field, which form a biological effect

Generally, such devices are constructed as the single block units and tend to have relatively high mass-dimensional characteristics [12–15].

Let us note that operationally, the first task is functionally “tied” to the patient and the second to the physician. In our case, we divide the neuroelectrostimulation device into two blocks, one of which will solve only the first task; the second block will only solve the second task. The information exchange between them will be provided by a telemetric communication channel, like Bluetooth. Then, according to this principle, a new architecture of the promising neuro-electrostimulator can be organized, which will make it compact and mobile [16]. This applies equally to the first block and to the second.

The implementation of the first block as compact and mobile one is quite realistic, as only the following components are mandatory:

- (i) Two multielement electrodes, between which a spatially distributed field of current pulses is formed
- (ii) Multichannel impulse current source, whose functions are performed by two multiplexers and a controlled current source
- (iii) Accumulator
- (iv) Bluetooth transmitter
- (v) Flash memory
- (vi) Microcontroller unit

Core features of the first block:

- (a) Number of partial cathodes: 13
- (b) Number of anodes: 13
- (c) Mass, 200 g

- (d) Dimensions, $90 \times 50 \times 18.5$ mm
- (e) Current pulse amplitude, 0–15 mA
- (f) Partial pulse duration, 15–60 μ s
- (g) Modulation frequency, 5–150 Hz
- (h) Minimal time of autonomic work, 24 h
- (i) Accumulator charging socket, USB Mini B

Diagram of the neuro-electrostimulator is presented on Figure 1, photo of the first block on Figure 2, and photo of the first block's printed circuit board on Figure 3.

The flowchart of the first block functioning algorithm is shown on Figure 4.

The program of the algorithm application is implemented as a set of tasks. Tasks that are not critical to the launch period are performed in the main program loop. Such tasks include first block unit testing, synchronizing cathodes pattern, and stimulation targets between first and second blocks, calculating amplitude for each cathode.

The tasks that are critical to the launch period include starting a new pulse packet and starting a new partial pulse inside pulse packet. This tasks forms current pulses field structure, sets up biotropic parameters, and determines stimulation targets. The critical to the launch period tasks has a higher priority and their starts are initiated by interrupt signals from the built-in microcontroller peripherals.

The current pulses field structure changing is possible only in a determined time points.

$$t = \frac{a}{\nu} + n * \tau, \quad (1)$$

where $a \in N$, $n \in N$, $0 \leq n \leq K$, K is the number of partial cathodes participating in field structure, τ is the partial pulse duration, and $T = 1/\nu$ is the current pulses field structure modulating period.

When the current pulses field structure changing time point t occurs, microcontroller in the first block performs the following steps:

- (1) Switching off the current partial cathode and switching on the new one according to neuro-electrostimulation program. If the current partial cathode is the last one in accordance with neuro-electrostimulation program, then a new partial cathode will not be connected.
- (2) If the anode needs to be changed according to the neuro-electrostimulation program or by the physician's command from the second block, the current connected anode will be disconnected and a new anode will be connected.
- (3) If current amplitude and time characteristics of current pulses field structure (such as modulating period T , partial impulse duration τ , and partial impulse amplitude A) differ from the target ones, then the characteristics is changing according to following equations:

$$\begin{aligned} A_{i+1} &= \begin{cases} A_{\text{target}}, & |A_i - A_{\text{target}}| < \Delta_A, \\ A_i + \Delta_A, & A_i + \Delta_A < A_{\text{target}}, \\ A_i - \Delta_A, & A_i - \Delta_A > A_{\text{target}}, \end{cases} \\ T_{i+1} &= \begin{cases} T_{\text{target}}, & |T_i - T_{\text{target}}| < \Delta_T, \\ T_i + \Delta_T, & T_i + \Delta_T < T_{\text{target}}, \\ T_i - \Delta_T, & T_i - \Delta_T > T_{\text{target}}, \end{cases} \\ \tau_{i+1} &= \begin{cases} \tau_{\text{target}}, & |\tau_i - \tau_{\text{target}}| < \Delta_\tau, \\ \tau_i + \Delta_\tau, & \tau_i + \Delta_\tau < \tau_{\text{target}}, \\ \tau_i - \Delta_\tau, & \tau_i - \Delta_\tau > \tau_{\text{target}}. \end{cases} \end{aligned} \quad (2)$$

The use of restrictions on the growth rate of the current pulses field structure parameters of neuro-electrostimulation avoids the appearance of patient's painful sensations during treatment procedure. The application for control of the field structure of current pulses and its characteristics of a microcontroller makes it possible to implement a large number of programs for neuro-electrostimulation.

Aforementioned computational procedures are required to implement the functions of the second block: when specifying the structure of the spatially distributed field of current pulses and the characteristics of this field, as well as the formation of various commands. These tasks can be handled on the basis of a computer or any specialized computation units that, in essence, will perform the functions of a specialized interface of the neuro-electrostimulator. To ensure the compactness and mobility of this specialized interface, we have chosen a smartphone.

The second unit of the neuro-electrostimulator is implemented in the form of an original cross-platform application for mobile devices based on Android, iOS, and Windows Phone. The application is structurally implemented in the form of two activities: the search for the first block and the control of the stimulation process. To organize the operation of the telemetric communication channel, the Bluetooth low energy API is used. In this case, a virtual control panel for the medical process is formed, which allows the physician to monitor the battery charge level in real time, the telemetry communication channel serviceability, the current pulse field structure, their biotropic parameters, and the position of the stimulation targets. Figure 5 shows a picture of specialized neuro-electrostimulator interface display in the virtual control panel mode of the treatment process.

Thus, the implementation of a neuro-electrostimulator in the form of two blocks will allow for the performance of the functions of restorative medicine:

- (1) For a patient to form a spatially distributed field of current pulses for the organization of a multifactorial neuro-electrostimulation of the neck nervous

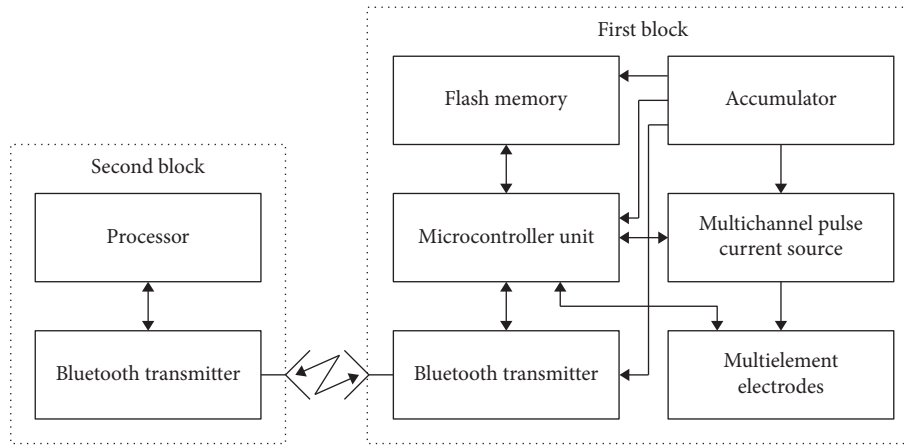


FIGURE 1: Diagram of the neuro-electrostimulator.



FIGURE 2: The first block of the neuro-electrostimulator

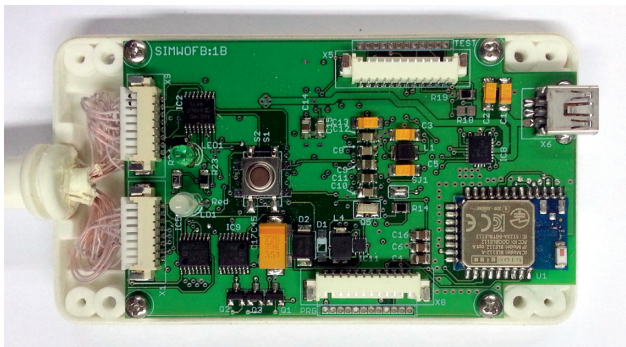


FIGURE 3: Printed circuit board of the first block.

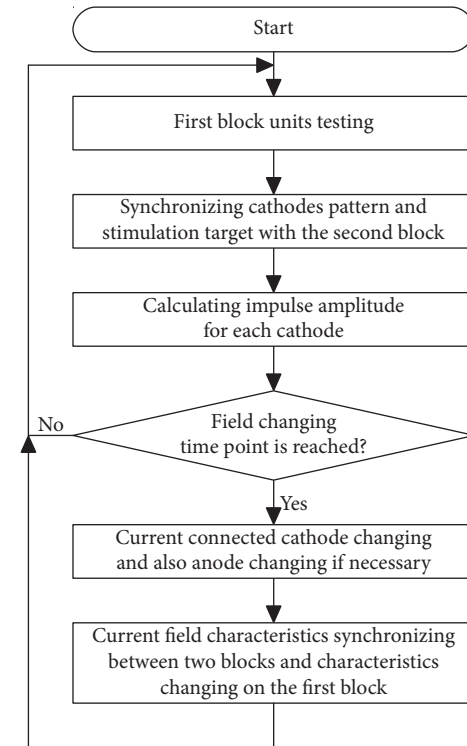


FIGURE 4: The first block functioning algorithm.

(2) For the physician to obtain wide opportunities for virtual management and control of the medical process, including the following:

- (a) In real time to monitor the operation of the first block of the neuro-electrostimulator, including monitoring the level of charge of its battery
- (b) Change the structure of the spatially distributed field of current pulses in the neck region, their parameters (amplitude, frequency, and duration)
- (c) Choose targets for the impact in the projection of the *sympathetic trunk*, *carotid plexus*, *cervical*

spinal plexus, *vagus nerve*, *nervus accessories* and branches of the *glossopharyngeal nerve* by selecting respectable functioning anodes

- (d) Change the number of partial cathodes participating in the formation of the field
- (e) Control the formation of a spatially distributed field of current pulses in the neck region in several patients (up to 10) via a telemetry channel using a single smartphone
- (f) Use the potential of telemedicine technology through the organization of remote monitoring of the rehabilitation process by highly qualified medical personnel

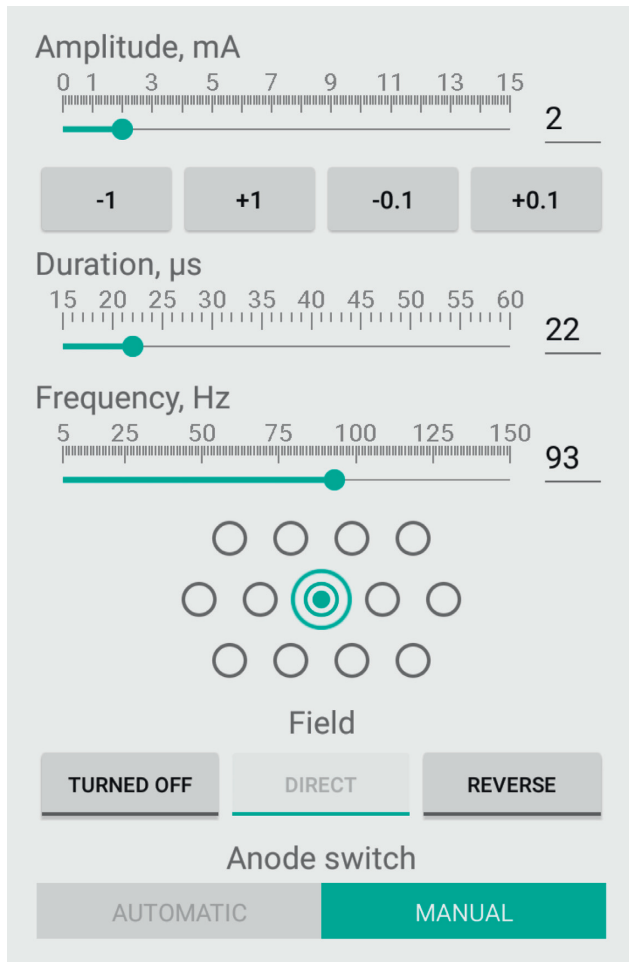


FIGURE 5: Specialized neuro-electrostimulator interface display.

- (g) Ensure the collection and processing of data on changes in the functional status of patients with an option to monitor the treatment process
- (h) Access to specialized databases of neuro-rehabilitation, storing personalized information about the course of the treatment process

The structure of such specialized database, named the neuro-electrostimulation service, implemented as a database model in the notation “Entity-Relationship” is shown in Figure 6.

Core elements of the database are entities (tables):

- (i) «Physician», having lines physician_id; Name, Surname
- (ii) «Patient», having lines patient_id; Name, Surname; Age; Sex; physician_id
- (iii) «Procedure», having lines procedure_id, physician_id, patient_id, Date, procedure type (examination, neuro-electrostimulation procedure, functional load), device_id
- (iv) «Device», having lines device_id, physician_id, patient_id, Stimulation features

- (v) «Data», having lines data_id, physician_id, patient_id, procedure_id, Data type (arterial pressure, electrocardiography signal, biochemical tests, psychological tests), Content

The proposed structure provides quick access to information on the treatment process available for a particular patient, allows storing and systematizing registered data, and making decisions for management of treatment based on this data. The use of such a database allows the formation of complex search queries that can be used for further analysis and processing.

4. Discussion: Prospects of Artificial Intelligence Application for Neurorehabilitation Management

At present, high hopes are placed on the use of artificial intelligence and machine learning for use in the diagnosis and control of the therapeutic process [17–19]. Thus, in our early work on the clinical example of arterial hypertension, it was shown that the application of quadratic discriminant analysis and methods for selecting diagnostic features of heart rate variability signals allows not only to perform express diagnostics of arterial hypertension, but also to evaluate the effectiveness of the therapeutic process with the use of neuro-electrostimulation [20]. Thus, it is possible to create an information decision support system for a physician.

The use of artificial intelligence in determining the paradigm of rehabilitation personally for each patient is made possible by taking into account the opportunities of telemedicine. As noted earlier, the specialized interface of the neuro-electrostimulator control is implemented as an application for a smartphone. A smartphone can interact with a global computer network. This allows not only to transmit the information generated in the neuro-electrostimulator, but also to obtain information from the specialized databases, to support the decision-making of the physician in treatment. Thus, the aforementioned information decision support system can be integrated with the neuro-electrostimulation service for the purpose of information exchange. The interaction of the neuro-electrostimulation service with the information decision support system provides express diagnostics of the cardiovascular system regulation disorders.

Creation of the neuro-electrostimulation service allows to close the contour of the physician and patient interaction and implements the functions of the biotechnical system for neurorehabilitation. The structural diagram of the biotechnical system is shown in Figure 7.

Such biotechnical system implements a number of principles of the patient-oriented approach in health care, such as personalized medicine and active involvement of the patient in the medical process. The presence of the neuro-electrostimulation service solves the problem of storing registered diagnostic data centrally on the server of the medical institution, simultaneous work with several patients

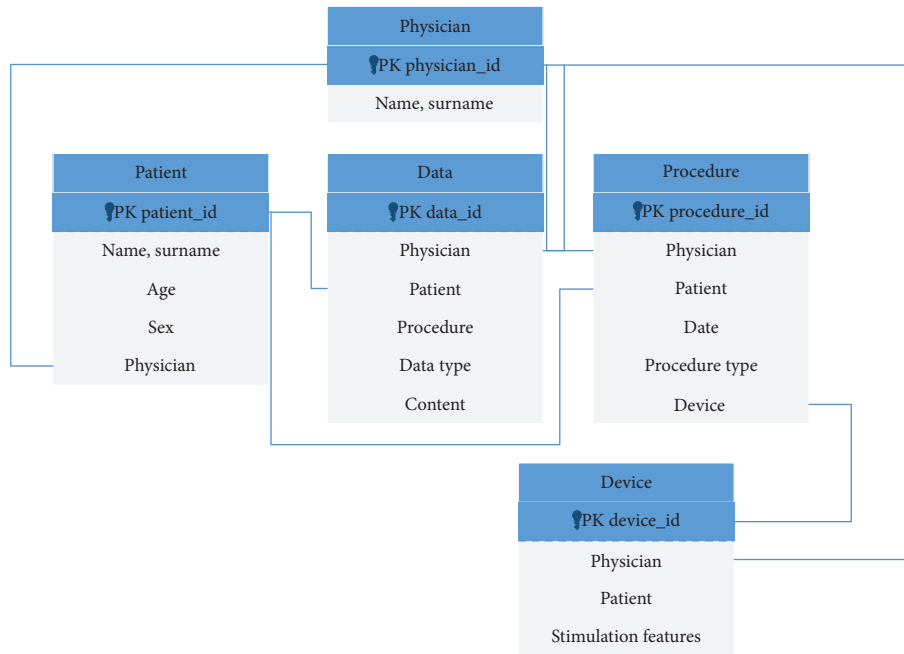


FIGURE 6: The neuro-electrostimulation service.

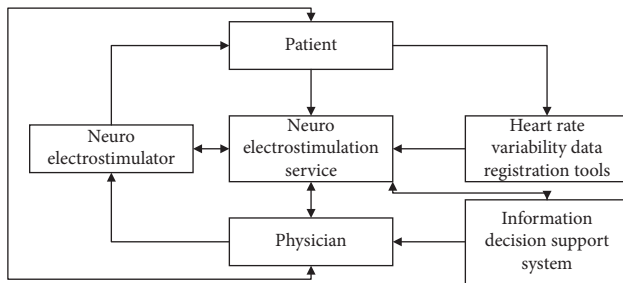


FIGURE 7: Biotechnical system for neurorehabilitation.

and effective use of the resources of the medical institution, and protection of patients' personal data from unauthorized access. The result of the interaction between the neuro-electrostimulation service and the information decision support system is the processing and automated analysis of the patient's data by means of machine learning to obtain express assessments for the diagnosis of arterial hypertension and the effectiveness of the therapeutic process, tracking the dynamics of changes in patient data, and information support by means of artificial intelligence for a physician.

5. Conclusion

As suggested in the article, principles of organization, circuit, and engineering solutions allowed to create mobile and compact hardware-information system for neuro-rehabilitation. Application of artificial intelligence and machine learning opens possibilities for treatment process management in accordance with the personalized medicine principles.

At present, the neuro-electrostimulation device has undergone clinical approbation in the treatment of depressive anxiety disorders, children with attention deficit disorder, and rehabilitation of patients after traumatic brain injuries. Clinical studies have shown that in comparison with known methods, a higher effectiveness of treatment is achieved through involvement in the regulatory process in addition to the autonomic nervous system and brain structures responsible for cognitive, motor, visual, auditory, vestibular, and other brain functions. These results are presented in more detail in the specialized publications of our physician co-authors [21, 22].

Data Availability

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

Conflicts of Interest

The authors declare that they have intellectual property rights in the field of use reported in this article.

Acknowledgments

This study was supported by the Act 211 of the Government of the Russian Federation (contract no. 02.A03.21.0006) and was funded by RFBR (project no. 18-29-02052).

References

- [1] P. A. Lapchak and J. H. Zhang, *Neuroprotective Therapy for Stroke and Ischemic Disease*, Springer, Berlin, Germany, 2017.
- [2] Global Health Estimates 2016, *Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016*, World Health Organization, Geneva, 2018.

- [3] E. J. Benjamin, M. J. Blaha, S. E. Chiuve et al., "American heart association statistics committee and stroke statistics sub-committee," *Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association*, vol. 135, no. 10, pp. e146–e603, 2017.
- [4] Q. Ashton Acton, *Advances in Central Nervous System Research and Treatment: 2013 Edition*, ScholarlyEditions, Atlante, Georgia, 2013.
- [5] V. S. Kublanov, "A hardware-software system for diagnosis and correction of autonomic dysfunctions," *Biomedical Engineering*, vol. 42, no. 4, pp. 206–212, 2008.
- [6] J. C. Wildenberg, M. E. Tyler, Y. P. Danilov, K. A. Kaczmarek, and M. E. Meyerand, "Electrical tongue stimulation normalizes activity within the motion-sensitive brain network in balance-impaired subjects as revealed by group independent component analysis," *Brain Connectivity*, vol. 1, no. 3, pp. 255–265, 2011.
- [7] D. Adair, M. Bikson, D. Q. Truong, L. Ho, and H. Borges, "Cognition and electrical stimulation of cranial nerves," *Brain Stimulation*, vol. 10, no. 2, p. 477, 2017.
- [8] A. K. Srivastava and C. S. Cox Jr, *Pre-Clinical and Clinical Methods in Brain Trauma Research*, Springer Science+Business Media, LLC, part of Springer Nature, Houston, TX, USA, 2018.
- [9] E. Krames, P. H. Peckham, and A. R. Rezai, *Neuromodulation: Comprehensive Textbook of Principles, Technologies, and Therapies*, Academic Press, Cambridge, MA, USA, 2018.
- [10] R. S. Orlov and A. D. Nozdrachev, *Normal Physiology*, A. Textbook (in Russian). M., Geotar-Media, Moscow, Russia, 2010.
- [11] S. S. Michailov, A. V. Chukbar, and A. G. Tsybul'kin, *Human Anatomy*, (in Russian). M., Geotar-Media, Moscow, Russia, 2013.
- [12] V. S. Kublanov, V. I. Shmirev, A. S. Shershever, and J. E. Kazakov, "About innovative possibilities of device "SIMPATOCOR-01" in management of functional disorders of vegetative and central nervous system in neurology, kremljovskaya medicine," *Clinichesky Vestnik*, vol. 4, pp. 60–64, 2010.
- [13] R. Nonis, K. D'Ostilio, S. Sava, J. Schoenen, and D. Magis, "Non-invasive vagus nerve stimulation with gammaCore (R) in healthy subjects: is there electrophysiological evidence for activation of vagal afferents?," *Headache*, vol. 56, p. 56, 2016.
- [14] A. Straube, J. Ellrich, O. Eren, B. Blum, and R. Ruscheweyh, "Treatment of chronic migraine with transcutaneous stimulation of the auricular branch of the vagal nerve (auricular t-VNS): a randomized, monocentric clinical trial," *Journal of Headache and Pain*, vol. 16, no. 1, p. 63, 2015.
- [15] M. I. Johnson, *Transcutaneous electrical nerve stimulation (TENS)*. eLS, John Wiley, Hoboken, NJ, USA, 2012.
- [16] V. S. Kublanov, M. V. Babich, and T. S. Petrenko, "New principles for the organization of neurorehabilitation," *Biomedical Engineering*, vol. 52, no. 1, pp. 9–13, 2018.
- [17] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 26–47, 2017.
- [18] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [19] M. Espinilla, J. Medina, A.-L. García-Fernández, S. Campaña, and J. Londoño, "Fuzzy intelligent system for patients with preeclampsia in wearable devices," *Mobile Information Systems*, vol. 2017, Article ID 7838464, 10 pages, 2017.
- [20] A. Y. Dolganov, V. S. Kublanov, D. Belo, and H. Gamboa, "Comparison of machine learning methods for the arterial hypertension diagnostics," *Applied Bionics and Biomechanics*, vol. 2017, Article ID 5985479, 13 pages, 2017.
- [21] V. S. Kublanov, K. Y. Retyunskii, and T. S. Petrenko, "A new method for the treatment of korsakoffs (amnesic) psychosis: neurostimulation correction of the sympathetic nervous system," *Neuroscience and Behavioral Physiology*, vol. 46, no. 7, pp. 748–753, 2016.
- [22] T. Petrenko, V. Kublanov, and K. Retyunskiy, "The role of neuroplasticity in the treatment of cognitive impairments by means multifactor neuro-electrostimulation of the segmental level of the autonomic nervous system," *European Psychiatry*, vol. 41, p. S770, 2017.

Review Article

WearableDL: Wearable Internet-of-Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future

Aras R. Dargazany ^{1,2}, Paolo Stegagno ², and Kunal Mankodiya ¹

¹Wearable Biosensing Lab, Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, Rhode Island, USA

²Intelligent Control & Robotics Lab, Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, Rhode Island, USA

Correspondence should be addressed to Aras R. Dargazany; arasdar@uri.edu

Received 23 July 2018; Accepted 2 October 2018; Published 14 November 2018

Guest Editor: Giuseppe De Pietro

Copyright © 2018 Aras R. Dargazany et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work introduces Wearable deep learning (WearableDL) that is a unifying conceptual architecture inspired by the human nervous system, offering the convergence of deep learning (DL), Internet-of-things (IoT), and wearable technologies (WT) as follows: (1) the brain, the core of the central nervous system, represents deep learning for cloud computing and big data processing. (2) The spinal cord (a part of CNS connected to the brain) represents Internet-of-things for fog computing and big data flow/transfer. (3) Peripheral sensory and motor nerves (components of the peripheral nervous system (PNS)) represent wearable technologies as edge devices for big data collection. In recent times, wearable IoT devices have enabled the streaming of big data from smart wearables (e.g., smartphones, smartwatches, smart clothings, and personalized gadgets) to the cloud servers. Now, the ultimate challenges are (1) how to analyze the collected wearable big data without any background information and also without any labels representing the underlying activity; and (2) how to recognize the spatial/temporal patterns in this unstructured big data for helping end-users in decision making process, e.g., medical diagnosis, rehabilitation efficiency, and/or sports performance. Deep learning (DL) has recently gained popularity due to its ability to (1) scale to the big data size (scalability); (2) learn the feature engineering by itself (no manual feature extraction or hand-crafted features) in an end-to-end fashion; and (3) offer accuracy or precision in learning raw unlabeled/labeled (unsupervised/supervised) data. In order to understand the current state-of-the-art, we systematically reviewed over 100 similar and recently published scientific works on the development of DL approaches for wearable and person-centered technologies. The review supports and strengthens the proposed bioinspired architecture of WearableDL. This article eventually develops an outlook and provides insightful suggestions for WearableDL and its application in the field of big data analytics.

1. WearableDL: Conceptual Architecture

Wearable DL is a concept derived from a holistic comparison between the evolving big data system and the human nervous system (NS) in terms of architecture and functionalities. Although the human NS is a biological mechanism, it essentially inspires the convergence, collaboration, and coordination of three key elements such as wearable tech (WT), Internet of things (IoT), and deep learning (DL) in the development of big data system for actionable outcomes and informed decision making.

The article views the big data system with respect to its close resemblance with the human nervous system (NS). The NS is responsible for coordinating the actions such as the transmissions of signals to and from the human body, identification, perception, decision making, and information storage [1]. Similarly, the big data system (or model) is evolving and conversing various domains such as wearable sensors, edge computing, fog computing, cloud computing, and deep learning (DL) to achieve equivalent functions such as signal communication, perception, decision making, analytics, and storage. As the complexity of the big data

system rises, it becomes important to understand the architectural and functional components of the NS. This could guide us to develop more improved and sophisticated version of a big data system.

1.1. A Brief Overview of the Human Nervous System. The NS is composed of two subsystems:

- (1) Central nervous system (CNS) consists of the brain and spinal cord
- (2) Peripheral nervous system (PNS) consists of nerves with sensory and motor fibers

1.1.1. PNS. The end elements of PNS are sensory and motor fibers which are connected to the parts and organs of the body. The sensory fibers sense various sensations including pressure, temperature, and pain on the body and sends them to the nerves leading to the spinal cord (a part of CNS). The motor fibers receive the commands from the CNS to actuate and activate the muscles and organs. The bundle of fibers which collectively forms nerves connected to the spinal cord relay the information back and forth between CNS and PNS.

1.1.2. CNS. The spinal cord is a part of CNS which serves two purposes:

- (1) It acts as a bidirectional relay for the signals to flow between the body and the brain. This function supports the NS to make centralized decisions.
- (2) The spinal cord also coordinates the reflexes in which the decisions are made in real-time to avoid delays in critical conditions. A simple example of the reflex is removing the hands from a hot object.

The ultimate top layer of CNS is the human brain made of approximately 100 billion neurons [1]. Each neuron connects to one or more other neurons. The brain receives the signals from the spinal cord and other sensor organs such as eyes, nose, tongue, and ears. The brain processes the incoming signals and makes decisions. It generates commands that pass through the spinal cord to the PNS. The commands activate the muscles or organs of the body. Apart from the processing and decision making, the brain also stores the information that is used in a short or long-term decision making process.

1.2. PNS vs Wearable Tech/Wearable Edge Devices for Big Data Collection and Application (Actions). WT is comparable to the sensory and motor fibers of PNS because of the following:

- (i) Fibers are the carriers of the information similar to WiFi backbone in WT
- (ii) WT is located onto the periphery of IoT architecture, interacting with the environment for sensing and actuation

For example, modern smartwatches come with built-in sensors such as heart rate, motion, ambient light, and also actuators including touch screen, audio speakers, and tactile (or haptic) feedback. Edge devices, such as smartphones, act similar to the nerves (as part of PNS). The smartphone receives the data from the connected smartwatch sensors and also commands the smartwatches to alert the wearers through the actuations on haptic, visual, or audio feedback. This helps us collect the sensor data and send them to the upper layer such as edge devices.

1.3. CNS-Spinal Cord vs IoT and Fog Gateways for the Big Data Flow and Local Intelligence. IoT and fog gateways are equivalent to the spinal cord in CNS (Figure 1) as follows:

- (1) Big data transfer (flow) between PNS (sensory and motory nerves) and the brain (the central processing and intelligence unit)
- (2) Local intelligence for locally responding to some stimuli such as extreme heat and pain

As described earlier, the spinal cord plays an important role in reacting and responding to some specific stimuli such as feeling pain and reacting to the pain, e.g., caused by extreme heat. It is also responsible to deliver the motor response and reactions from our brain to the PNS (our motor actuators and muscles) for any dynamic (kinematic) movement (motion) internally and externally, i.e., inside our body or outside. IoT intelligence, as a local intelligence, is functioning similar to the spinal cord. For example, the smartwatch sensor data can be processed onto fog gateways which are located in homes or hospitals away from the centralized cloud servers. In this case, the sensor data are processed on the gateway for the local decision support in time-critical applications, e.g., the sensor data streams could help the detection of a fall event in an elderly person living alone at home. In this way, the fog or IoT gateway provides reflex-type services to alert appropriate individuals such as medical personnel or caretakers to respond to the event immediately. This reduces the potential delays in time-sensitive events.

1.4. CNS-Brain vs DL and Cloud Computing for Big Data Analytics. The cloud computing servers are equivalent to the physical architecture of the brain, and the DL-based big data analytics resembles the function of the brain. The human brain is a centralized processor to receive the incoming stimulus from the spinal cord (connected to PNS) or other sensor organs. Upon receiving, it perceives and makes decisions on how and when to respond to the stimuli. It also stores the information. Similarly, the cloud computing servers receive the big data from WT via fog computing. Upon receiving, it uses high-performance computers to apply DL methods (explained in the next section) that help in decision making. Very similar to the brain, the cloud computers derive when and how to respond to the incoming queries. It often stores the sensor data to learn the patterns

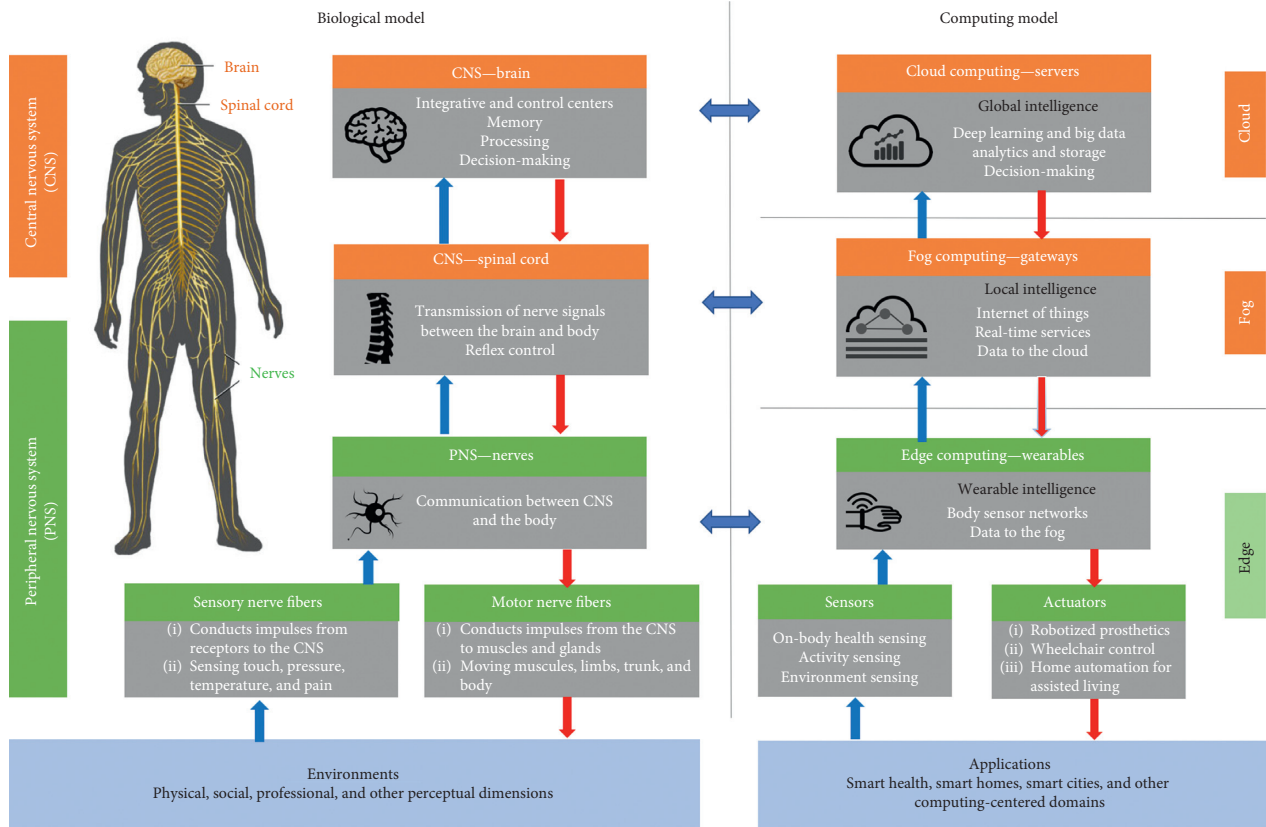


FIGURE 1: The WearableDL conceptual architecture: the human nervous system as the main biological model and inspiration (right) vs the human-made computing model as the actual architecture (left).

and create historical database to enable informed decision making in the future.

1.5. Outline and Contributions. In this article, we endeavor to describe the benefits and challenges associated with the use of DL in the wearable big data. We have conducted a thorough survey of more than 100 literatures related to DL and its applications in wearable IoT. The survey allowed us to create a holistic picture combining wearable sensors, IoT, DL, and big data. This work's key contributions are structured as follows:

- (i) Section 2 provides an overview of wearable IoT including the concept, its different categories of wearable IoT devices, and its future direction in a nutshell.
- (ii) Section 3 provides a research roadmap for DL thorough understanding of its past, its present, and its future. Here, we focus on how understanding the human brain, specifically neocortex, links to the development of the artificial intelligence (AI) and how that is mainly divided into three areas: ML, DL, and Cortical Learning (CL) which are covered in this section.
- (iii) Section 4 emphasizes on the recent similar work applications of WearableDL in big data analytics.

Over 100 recently published literatures were reviewed and included in this section to correlate with the paradigms of WearableDL and its applications.

- (iv) Section 5 projects WearableDL future research and application direction in association with the wearable big data.

2. Wearable Internet-of-Things

In 1965, an observation, later regarded as Moore's law, estimated that the number of transistors on integrated circuits (IC) doubles every two years [2]. Moore's law prediction played an important role in the semiconductor industry and motivated the evolution of miniaturized yet high performance computing (HPC) chips which revolutionized the modern world. This evolution caused an explosion in the production of electronic devices and therefore brought a limitless expansion in the use and applicability of the computing chips that today drive smart wearable devices, smartphones, personal computers, smart homes, and smart cities along with WiFi, Internet, and other communication devices. As a result of the aforementioned evolution, explosion, and expansion, the wearable devices are booming in the market, and therefore, we witness the growth of personalized big data that hold a significant value to the end-users including citizens, communities, hospitals, and

governments to improve health or performance, reduce medical cost, and increase efficiency [3].

2.1. Wearable Devices Categories. Overall, the wearable devices can be categorized into three main classes (Figure 2):

- (1) **Implantable devices:** these devices are implanted inside the body for a long period of time, e.g., cardiac pacemaker and deep brain stimulator are implanted for 5–10 years to provide current to specific organs.
- (2) **Wearable contact devices:** this is the largest category among the three types. These devices are targeted to stay on the body unobtrusively to collect various parameters including heart rate, physical activity, body temperature, muscular activity, blood/tissue oxygenation, and other physiological parameters. The most common devices in this category are smartwatches, smart clothing, smart footwear, fitness trackers, HR chest belts, and ECG Holter monitors.
- (3) **Wearable ambient monitors:** these devices are made to sense outside environment instead of the body's physiological state. Google glasses (smart glasses) are a simple example of this category, in which, a wearable camera allows to record the surrounding scenes [4].

2.2. Wearable IoT: Convergence of Wearable Devices and Internet-of-Things (IoT). The convergence and deployment of wearable devices, Internet-of-things (IoT), and cloud computing together allow us to record, monitor, and store a wide range of the big data from individuals such as personalized health and wellness data, body vital parameters, physical activity, and behaviors, which are all critical data indicating the quality and the trend of daily life [5]. In the past, wearable devices were a stand-alone system. However, bringing wearable devices into the framework of IoT makes it possible to stream the data from an individual to a centralized location such as cloud servers. The continuous accumulation of the wearable data becomes a massive big data [6] that, in general, are a set of sequential time-series signals and logs containing biometrical, behavioral, physiological, and biological information depending on the nature of wearable devices categorized above. One of the key objectives of collecting the wearable big data is to support remote or on-site decision making by detecting symptoms, events, and anomalies, or by producing contextual awareness [7].

2.3. Wearable Data Categories. Wearable biosensing devices can collect a large variety of physiological data continuously, all-day long and in any-place health, mental, and activity status monitoring. These multiparameter physiological sensing systems provide us with reliable and crucial measurements for supporting online decision making by detecting the symptoms and producing contextual awareness [8]. A wide range of wearable data in biomedical and health is provided by this overview [9].

2.4. Emerging Unobtrusive Wearable Devices. Wearable sensors can be either woven or integrated into clothing, accessories, and the living environment, such that individuals' or patients' data can be collected in their daily life. According to an overview [10], four emerging unobtrusive wearable technologies (WT) which are essential for collecting the individuals' health data are the following:

- (1) Unobtrusive sensing methods
- (2) Smart textile technology
- (3) Flexible-stretchable-printable electronics
- (4) Sensor fusion

2.5. Data Reliability. Data reliability strongly depends on the type of collected data and specifically on the category of the collected wearable data in general. In the wearable DL scenario, it is not the role of the wearable devices to assess data reliability. A presifting of the data, particularly in case of structured data, can be implemented directly on the device by embedding data sifting policies dictated by a prior interaction with medical specialist, physicians, and studies. However, data reliability should be assessed at IoT and DL level as discussed later.

3. Artificial Intelligence

Artificial Intelligence (AI) is ultimately the ability to reconstruct the human biological intelligence for modern machines. AI domain is currently divided into three active learning-based areas of research: machine learning (ML) [11, 12], deep learning (DL) [13], and cortical learning (CL) [14] (Figure 3).

3.1. Cortical Learning. Cortical learning (CL) is inspired from our cortical structure (i.e., based on studying the neocortex) and coined by Hawkins et al. [14, 15] from <https://numenta.com/Numenta>. The cortical area is the largest part of the brain in humans and monkeys compared to other species and the main source of our intelligence [14, 16]. The CL algorithm/approach (CLA), inspired by the architecture of the human cortex [14, 17], is applied to an approach called hierarchical temporal memory (HTM) [16, 18–20]. Cortex learns the spatial and temporal patterns in the sequential data, e.g., for visual perception, spoken language comprehension, manipulating objects, and navigation in a complex 3D world [17].

3.2. Machine Learning. Machine learning is the mother subject for deep learning and many other statistical or probabilistic analysis approaches but not necessarily related to CL which is neuroscience-based endeavor for AI (computational neuroscience or systematic neuroscience). ML is mostly referring to shallow ML approaches which are not scalable to the data size. This set of shallow artificial learning algorithms [11, 12] helps machine directly learn from the data, model the data, and generate machine intelligence. ML is highly founded on mathematics, e.g., linear algebra,

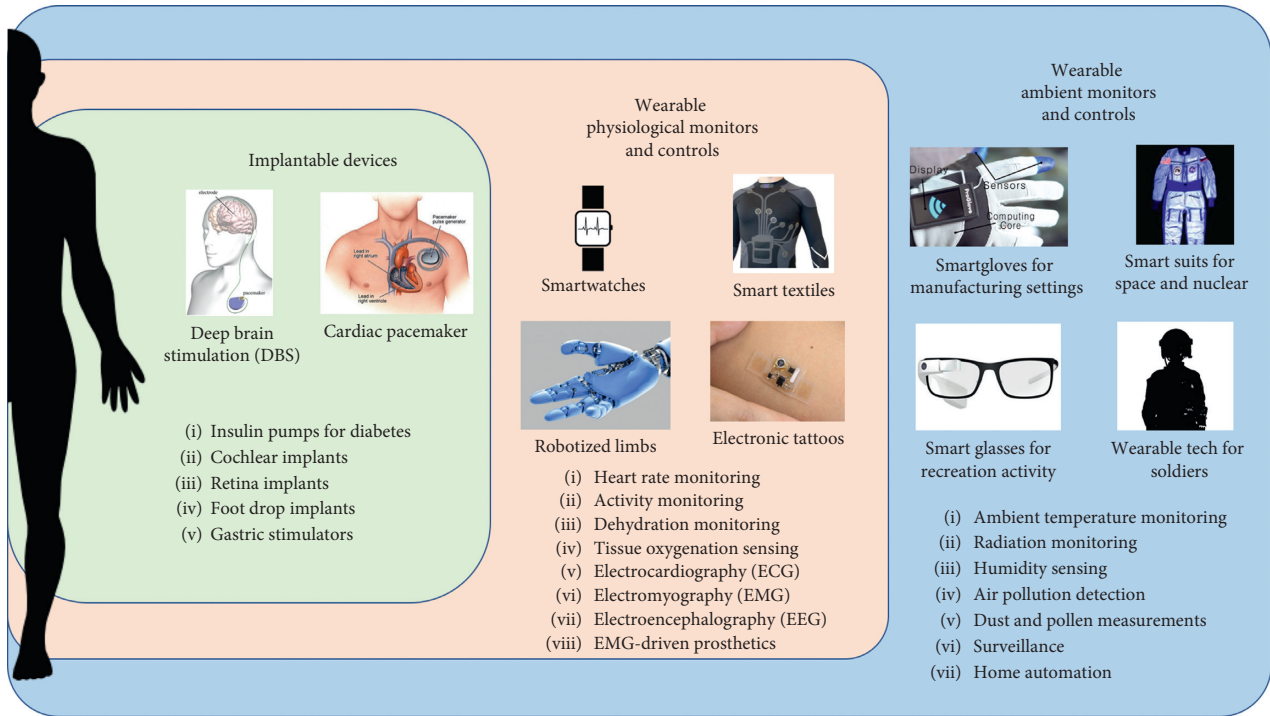


FIGURE 2: Different types of wearable IoT-based devices for wearable big data collection.

calculus, statistics, probability, and stochastic optimization approaches such as evolutionary algorithms (EA) and Monte Carlo search. Some of the ML limitations are as follows:

- (1) It is very broad and often mathematically proved but not biologically inspired. This is a problem since biologically inspired algorithms often are proved to be extremely powerful and robust such as genetic algorithms. On top of that, AI is targeting biological intelligence at the first place and ultimately aims to replicate/reconstruct our biological intelligence, human intelligence.
- (2) It is often shallow and not scalable to the data size, i.e., as the data size or dimensionality grows exponentially (big data problem), the traditional ML approaches (e.g., SVM) can not scale up the data size. This causes a problem so-called under-fitting which means there are not enough parameters in the learning approach for approximating the best fitting function.
- (3) It is also hard to apply it to high-dimensional data directly. That is why we have to apply dimension reduction to the data first by either manually do the feature extraction or engineering (hand-crafted features) and then apply the ML approach to the data features with the reduce dimensionality.
- (4) ML approach accuracy and robustness for noisy data is almost not comparable to DL approaches since ML approaches are learning from few examples or small training data compared to DL approaches which are capable of learning from massive dataset (big data).

3.3. Deep Learning. DL approaches differ from shallow ML algorithms in terms of scalability, i.e., depth (number of hidden layers) and width (number of cells or units or neurons in each layer). DL (or deep ML) is a scalable ML approach capable of scaling to the data size in terms of high number of data samples or data dimensionality. DL is applied to artificial neural networks (ANN or NN) and that is why it is also known as deep neural networks (DNN) [13, 21]. DL is the ability to learn the deep architectures of NN using backpropagation (BP) [22, 23]. Error backpropagation [24] is the dominant training approach for NN which was proposed in 1986 for training multilayer perceptrons (MLP) which is backpropagation of the resulting error between the predicted output and the given labels into the network for fine-tuning the weights in order to minimize the loss/cost function in a nonconvex surface. DL is loosely inspired by the visual cortex [25–27]. It is mimicking our brain [27] in terms of learning and recognizing the spatial and temporal patterns (or spatiotemporal) in the data. DNN are basically deep hierarchical layers of perceptrons [28], as artificial neurons, for representation and regression learning [29, 30].

4. Deep Learning

The research question of "How can the massive wearable big data be analyzed to produce actionable outcomes?" is difficult to answer when the wearable big data is heterogeneous, unlabeled, and unstructured. This means the wearable big data seeks unsupervised learning methods that can not only analyze the data but also identify helpful patterns leading to

	Machine learning (ML)	Deep learning (DL)	Cortical learning (CL)
Math	Yes	Yes	—
Shallow	Yes	No	No
Scalable	Yes	Yes	Yes
Multimodal	Almost	Yes	Yes
Biological	No	Yes	Yes
Foundation	Math	Math + bio	Bio
	Support vector machine (SVM)	Convolutional neural net (CNN)	Hierarchical temporal memory (HTM)
Example			

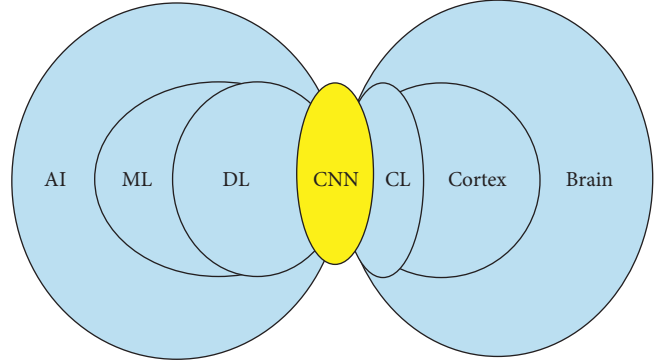


FIGURE 3: Comparing AI and brain along with machine learning vs deep learning vs cortical learning.

informed decision making. In recent years, deep learning (DL) has been established as a new area of machine learning research which aims to advance artificial intelligence [13]. A plethora of studies provide evidence that DL has achieved state-of-the-art results in various fields related to computational intelligence and big data including computer vision and image processing [31], speech processing [32], natural language processing (NLP), and machine translation [33, 34]. Similarly, these DL advancements bring a new promise to analyze the unsupervised wearable big data in order to recognize the spatial and temporal patterns related to health, wellness, medical condition, sports performance, and safety (Table 1).

Deep learning (DL) is exponentially gaining interest in research and development (R&D) community at academia and industry as they are also being heavily invested on by giant software and hardware companies such as Google, Nvidia, and Intel [35–37].

4.1. Deep Learning History: Receptive Fields of Neurons Inspired from Cat’s Visual Cortex. Simple-cells and complex-cells [38] were found in the receptive fields of single neurons in the cat’s visual cortex. The discovery of receptive fields of neurons in the cat’s visual cortex [38] contributed enormously to NN, AI, computer vision, and neuroscience community (demonstrated visually with timeline in Figure 4).

4.2. Deep Learning History: The First Conceptual Architecture—Cognitron and Neocognitron. The discovery of simple and complex cells was followed by the introduction of cognitron and neocognitron (by Fukushima et al. in 1975 [39, 42–44]). Neocognitron (as the first proposed deep NN architecture) was the inspiration behind the introduction of the convolutional neural networks (CNNs) by LeCun in 1989 [45] (as shown in the past part of the timeline in Figure 4). Cognitron and neocognitron (Fukushima et al. [39, 42, 43]) were introduced as self-organizing multilayered neural networks. The proposed cognitron and neocognitron architectures, by Fukushima [42, 43], is composed of the simple-cells and complex-cells inside the CNN architecture.

4.3. Deep Learning History: Neural Networks. Schmidhuber’s survey [46] thoroughly reviews the history of DL in NN since the birth of ANN along with different types of learning approaches applied to DNN architectures such as unsupervised learning (UL), supervised learning (SL), and reinforcement learning (RL). It also discusses evolutionary algorithms (EA) and optimization approaches (e.g., genetic algorithms) along with the learning algorithms for minimizing mean squared error (MSE) and sum of squared error (SSE).

4.4. Deep Learning Math Foundation: Artificial Neural Networks Universal Approximation Theorem. ANN is a universal function approximator based on the universal approximation theorem [23, 47–49]. This theorem proves that ANN, even with a single hidden layer of finite size, can approximate any continuous functions [47]. This approximation theorem [47] is applicable to high-dimensional as well as low-dimensional function approximations [23]. For example, 2-dimensional-CNN (2D-CNN) or 3-dimensional-CNN (3D-CNN) for image and point-cloud classification (high-dimensional 2D or 3D data) was compared to feed-forward neural network (FFNN) for a low-dimensional time-series signal classification. In this case, CNN contains much more parameters for high-dimensional function approximation compared to FFNN which is a low-dimensional function and containing much less parameters.

4.5. Deep Learning Applications: Recent Breakthroughs and State-of-the-Art Results. DL has achieved the state-of-the-art results in many fields such as computer vision, speech processing, and machine translation as the following:

- (1) A breakthrough in 2012 for computer vision using DL: deep convolutional neural nets (CNN), LeNet by LeCun in 1989 [45], proved to be enormously efficient in an end-to-end image classification and analysis [31].
- (2) A breakthrough in 2012 for speech processing using DL: another important breakthrough, almost in the same year as Krizhevsky [31], was applying DL to TIMIT, massive dataset for speech recognition

TABLE 1: The AI domain and DL review table.

Learning-based AI approach	Data-based learning approach		
	Unsupervised learning (UL) for unlabeled data	Supervised learning (SL) for labeled data	Reinforcement learning (RL) for rewarded labeled data (labeled data with cost function)
Cortical learning (CL): neuroscience of brain cortex (cortical areas)	Sequence learning: hierarchical temporal memory (HTM) and cortical learning algorithm (CLA)	—	—
Machine learning (ML): shallow ML	Dimension reduction: principle component analysis (PCA) and independent component analysis (ICA), clustering: expectation maximization (EM), K-means, K-nearest neighbors (KNN), approximate nearest neighbor (ANN), and fast library for approximate nearest neighbor (FLANN)	Linear discriminant analysis (LDA), random forest, search trees (Monte Carlo search), artificial neural network (ANN) or multilayer perceptron (MLP), and support vector machine (SVM)	Q-learning, policy-learning, and inverse RL (IRL)
Deep learning (DL): deep ML	Deep unsupervised learning (DUL): restricted Boltzmann machine (RBM), deep belief network (DBN), deep Boltzmann machine (DBM), autoencoder (AE), variational autoencoder (VAE), generative adversarial network (GAN), and sequence learning	Deep supervised learning (DSL): Feed-forward neural network (FFNN), deep neural network (DNN), spike neural network (SNN), sequence-to-sequence learning, recurrent neural network (vanilla RNN), long short-term memory (LSTM), convolutional LSTM (ConvLSTM), and gated recurrent unit (GRU)	Deep reinforcement learning (DRL): Deep Q-Network (DQN), AlphaGo, and inverse DRL (inverse RL & GAN)

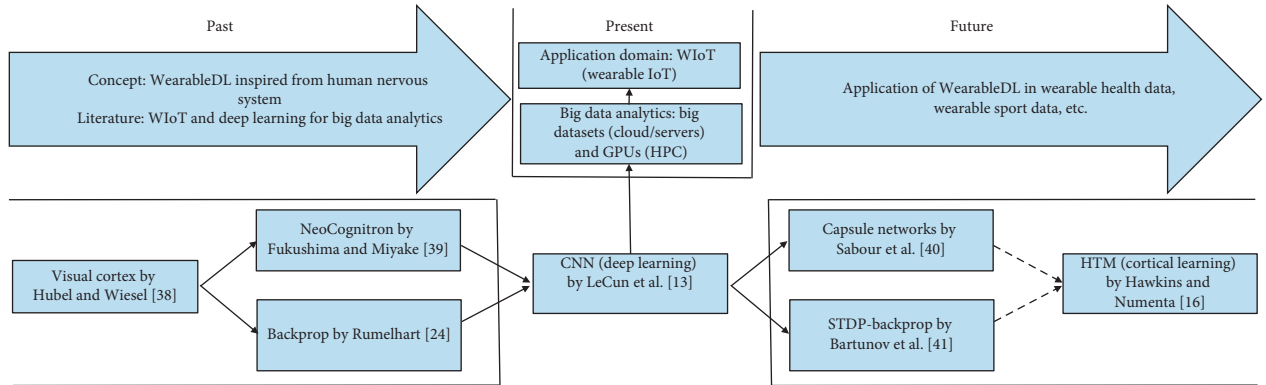


FIGURE 4: The simplified research roadmap for DL: (past) how it was inspired by visual cortex research, (present) how it is related to wearable IoT and big data analytics, and (future) how it is connected to cortical learning. The related works mentioned/included in this figure are the following: visual cortex by Hubel and Wiesel [38]; neocognitron by Fukushima and Miyake [39]; Backprop by Rumelhart et al. [24]; CNN (deep learning) by LeCun et al. [13]; Capsule networks by Sabour et al. [40]; STDP-backprop by Bartunov et al. [41]; HTM (cortical learning) by Hawkins et al. [16].

[32]. Microsoft immediately started adopting and applying this approach to its own AI assistant, Cortana for Windows 10 [32].

- (3) Google brain project—The first large-scale DL project in 2012: this project, as a large scale distributed deep networks led by Dean et al. [50], applied deep belief network (DBN) to massive data from Youtube (videos of cats) using 16,000 computers in distributed parallel configuration. This

large-scale implementation of DBN, on distributed parallel computing platforms, successfully recognized cats in videos after watching millions of cat videos on Youtube without any supervision or teaching signals (unsupervised setting).

- (4) Bridging the gap between human-level translation and machine translation in NLP by Google neural machine translation: Google neural machine translation (GNMT) [33, 34] is as an end-to-end DL model

for automated translation which has outperformed the conventional phrase-based translation systems by far. The proposed GNMT [33, 34] system requires big computational power (big compute) and massive datasets (big data) for both training and translation inference for building big model (big net).

4.6. Deep Learning Architectures. Some of the important and famous DNN architectures are the following:

- (1) Feed-forward neural network (FFNN): this is the simplest NN, also known as multilayer perceptron (MLP), with feed-forward connections. FFNN is also referred to as fully connected network (FCN) inside CNN architectures.
- (2) Convolutional neural network (CNN): CNN is loosely inspired by the cat's visual cortex [38]. It was initially proposed as Cognitron [42] and Neocognitron [39]. CNN architecture was initially applied to digit recognition and trained using BP by LeCun in 1989 [45]. This CNN architecture is also referred to LeNet [45].
- (3) Deep belief network (DBN), Deep Boltzmann Machines, and Restricted Boltzmann Machines (RBM): DBN was initially proposed and trained using backpropagation by Hinton et al. [51, 52] as a deep unsupervised learning (DUL) approach using greedy pre-trained stacked up layers of RBM.
- (4) Autoencoder (AE): AE is another DUL approach for dimensionality reduction [53] and data compression. Variational AE (VAE) [54] is another recent version of AE which improved the AE precision in generating images and data as a generative model using Bayesian distribution.
- (5) Spike neural network (SNN): This type of NN is mimicking the Spike stimulation of brain inside the ANN, i.e., loosely inspired from how Spikes are activating neurons in our biological NN (brain).
- (6) Deep Q-networks (DQN): DQN [55, 56] is the first deep reinforcement learning (DRL) approach proposed by Google DeepMind. This DL approach achieved human-level control in playing variety of Atari games.

4.7. Deep Learning Dominant Training Approach: Backpropagation. DL is mainly related to the algorithms for learning big and deep NN architectures. BP is dominantly the learning algorithm used in DL [29, 57] which is the main power behind the scalability of DL architectures such as CNN and DBN. DL is coined mainly by LeCun et al. in 2015 [13]. Goodfellow et al. [21] published a book providing a thorough explanation of DL theory and approaches.

4.8. Deep Learning Categories and Subdomains. The DL approaches, regardless of their application domains, are mainly categorized into three dominant groups (the same as

ML): deep unsupervised learning (DUL), deep supervised learning (DSL), and deep reinforcement learning (DRL). There are also some subcategories (subdomains) which are currently the active area of research in DL as well such as transfer learning (TL), semisupervised learning, learning-by-demonstration, and imitation learning.

4.9. Deep Learning Biological and Neurological Inspiration Related to Backpropagation. BP (proposed by Rumelhart et al. [24, 29, 57]) in DL is supported neurologically by random synaptic in Lillicrap et al. [26]. Lillicrap et al. [26] argues that BP is functioning similar to an error feedback neuron for error optimization (minimization). Yasmin and DiCarlo [25] also provide another strong biological foundation for BP and CNN architecture in DL. They demonstrate visually how the goal-oriented convolutional hierarchical layers are inspired from sensory cortex.

4.10. Deep Supervised Learning. DSL is divided into three categories [46]: FFNN & CNN, recurrent neural network (RNN), and the hybrid one (combination of both: convolutional long short-term memory (LSTM) and convolutional RNN).

4.10.1. Feed-Forward Neural Network and Convolutional Neural Network. FFNN was also traditionally known as MLP. FFNN, so-called FCN, is often the last two years inside CNN architecture in DSL. CNN was at first applied to optical character recognition (OCR), specifically for digit recognition, and trained using BP by LeCun et al. in 1989 [45]. This CNN, named LeNet after LeCun et al. [45], was brought back to the attention in 2012 after reducing the classification-error rate almost in half in Imagenet contest by Alexnet, named after Krizhevsky et al. [31].

4.10.2. Recurrent Neural Network and Long Short-Term Memory. LSTM was proposed by Hochreiter and Schmidhuber in 1997 [58]. LSTM was successfully trained using BP through time (BPTT) unlike vanilla RNN which was heavily suffering from the problem of vanishing gradient and exploding gradients [59]. LSTM-based RNN is often applied to sequential learning for temporal pattern recognition.

4.10.3. Hybrid DL: Convolutional LSTM (ConvLSTM). Xingjian et al. [60] proposed convolutional LSTM (ConvLSTM) as a hybrid approach which is a combination or an integrated version of CNN [13] and RNN (LSTM [58]). In this regard, Zhang et al. [61] show reliable results using hybrid approach for speech recognition. Residual bi-directional ConvLSTM [61] is a very deep network including bidirectional LSTM and CNN with residual connections for an end-to-end speech recognition which is an efficient and powerful deep hybrid model for acoustic speech recognition.

4.11. Deep Unsupervised Learning. DUL focuses on unlabeled big data which are abundantly available nowadays on the web. Bengio et al. [62] review DUL approaches and provide new perspectives on them. Yeung et al. [63] propose an approach for learning the big unlabeled data existing on web (i.e., also referred to as *wild*). Rupprecht et al. in 2016 [64] also propose a DUL framework for the unlabeled big data as new methodology of learning multiple hypothesis. Mirza et al. [65] provide a DUL architecture for the generalization of the aligned features, specifically to perform TL across multiple task domains.

4.11.1. Deep Belief Network, Restricted Boltzmann Machine, and Google Brain Project. DBN is, in fact, the stacked up layers of pretrained restricted Boltzmann machine (RBM). In 2012, Google brain project (as a "large scale distributed deep networks") led by Dean et al. [50], was applying DBN using massive data from Youtube videos on cats and 16,000 computers in distributed parallel configuration. This large-scale implementation of DBN, on distributed parallel computing platforms, successfully recognized cats in videos after watching millions of cat videos on Youtube without any supervision or teaching signals, i.e., entirely accomplished in unsupervised setting (DUL). When DBN is trained on a massive dataset (big data) as DUL, it can learn how to probabilistically reconstruct its inputs. DBN layers (layers of representation) can act as feature detectors (extractors) on inputs [32, 50, 51, 66]. After this learning step, a DBN can be further trained in a supervised way to perform classification [32, 51] for TL.

4.11.2. Generative Adversarial Networks (GANs). Goodfellow et al. [67] proposed GAN as two networks which are competing against each other. One of these networks is *generator* and another one is *discriminator*. The generator tries to produce fake input data similar to the real one to fool the discriminator. This adversarial training performed on GANs is entirely based on game theory.

4.11.3. Autoencoders. AE is an unsupervised DL architecture for DUL applied to denoising, dimensionality reduction [53], data compression, and image or data generation (generative models). VAE [54] is an improved AE as a generative model using Bayesian distribution. It can also be trained and transferred to a DSL architecture (TL) for classification and regression purposes [68].

4.12. Deep Learning (DL) and Reinforcement Learning (RL) Started Getting Published in Nature: Quick Review of the Recent Years Progress. This part briefly walks you through how DL and RL were combined/accomplished with an incredible speed since 2015 only and only from Nature publication perspective:

- (1) In 2015, deep learning (DL) models found their way into Nature publications by producing incredible results in AI [13].
- (2) At the same year, neuroscientists found a very interesting relationship between goal-driven DL models and our sensory cortex in the brain [25]. This was a huge leap toward biological inspiration of deep reinforcement learning models.
- (3) In that year, one DL-based AI agent created super-human-level performance in many Atari games [55]. This approach, so-called Deep Q-Networks (DQN), demonstrated least human-level control (performance) in playing many Atari games [56]. This was the birth of deep reinforcement learning (DRL) because of the combination of RL (initially proposed by Sutton in 1984 [69]) and DL [13].
- (4) In 2016 (one year later), another DL-based AI agent, AlphaGo, dominated the Go game by only watching the previously human-played Go games [70]. AlphaGo (Silver et al. from Google DeepMind in 2016 [70]) made a considerable impact on DRL community by dominating the game of Go (a Chinese ancient chess-like game) using two deep co-operative networks [70]: deep policy network and deep value network (DQN [55]). The policy network was basically recommending the next possible moves (actions) and the value network (i.e., Q-network) evaluate the moves intuitively based on the previous experiences. Eventually Q-network (value-network) picks the most valuable move based on the selected move with maximum rewarded value. The co-operative networks in AlphaGo [70] are cooperating with each other on contrary with GAN.
- (5) Recently, AlphaGo Zero [71] started learning the Go game from scratch only-and-only by playing in a try-and-error fashion and even beats the previous AlphaGo eventually.
- (6) Finally, an important implementation of grid-like cells in mice [72] can loosely demonstrate how the navigation is performed using these grid-like cells and how they are represented in artificial agents. These Grid cells were discovered in 2005 by Wills et al. [73] and a team of scientists in Norway [74]. They were awarded the 2014 Nobel Prize for their discoveries of cells that constitute a local/global positioning system in the brain.

DRL has opened a new frontier in AI so-called artificial general intelligence (AGI) which is exponentially growing and succeeding in demonstrating human-level and even super-human level performance not even in playing Atari games but in robotics [75] and other domains performing complex task such as imitation learning or learning by demonstration:

- (1) Combining inverse RL and GANs: since GAN is a generative model to maximize the reward function

to fool a discriminator network, it is related to RL in terms of learning how to maximize the reward function. In RL, learning the reward function for an observed action is coined as inverse reinforcement learning (IRL). Finn et al. [76] show that IRL [77–79] is equivalent to GAN [67] by highlighting the mathematical connection between them.

- (2) Generative adversarial imitation learning (GAIL): it is the combination of GANs and imitation learning [80]. This model is also introduced earlier in 2016 by Baram et al. [81] as model-based adversarial imitation learning. These models generally aim to model human behaviors and motives using IRL [82] which is dominantly targeting lack of reward of function for variety of complex task or the difficulty of defining a reward function for these tasks rather than learning a reward function for them.

Another recent development in DRL is applying RNNs, specifically LSTM [58], for learning the temporal dependencies since the RL tasks are all sequential, known as generative RNNs. In this regard, Schmidhuber team is one for the main fore-frontiers by introducing the world models [83]. A beautiful combination of GAIL and generative RNNs are proposed very recently in Zhu et al. [84] in order to apply these RNN-based GAIL models for diverse visuomotor skills, specifically in robotics manipulation across both simulation and real domains. In this direction, generative query networks (GQN) [85] have shown very promising results in terms of an agent predicting how the environment would look like taken a specific action. GQN is another great combination of RNN, GAN, and imitation learning.

5. WearableDL: Literature Review

Feature extraction is the key in understanding and modeling the repetitive patterns of the collected physiological and behavioral data. Traditionally hand-crafted features were extracted based on the expert knowledge, which were labor-intensive and time-consuming, for classification or regression purposes. Moreover, the manual feature extraction process does not scale well when the wearable data is growing rapidly in size temporally (number of samples in time) or spatially (number of dimensions). That is why our article aims to explore DL approaches since they are capable of scaling to the data size. In this section, we review the literature related to DL approaches for analyzing different types of wearable data as demonstrated and mapped briefly in this Table 2.

5.1. Embedding DL in Mobile, Wearable, and IoT Devices. Lane et al. [86] presents a study on embedded DL in wearables, smartphones, and IoT devices in order to build the knowledge of the performance characteristics, resource requirements, and the execution bottlenecks for DL models. Regarding DL for mobile, wearable, and embedded sensory applications, DL requires a significant amount of device (and

processor) resources. The limited availability of memory, computation, and energy on mobile and embedded platforms is a serious problem for powerful DL approaches.

5.1.1. SparseSep: Large-Scale-Embedded DL in Smartphones. SparseSep [87] leverages the sparsification of fully connected layers and separation of convolutional kernels to reduce the resource requirements of DL algorithms. SparseSep [87] allows large-scale DNN (with fully connected layers and with convolutional layers) to run and execute efficiently on mobile and embedded hardware with minimal impact on inference accuracy.

5.1.2. DeepX and Demo: Embedded DL Execution Accelerator. DeepX [89] is a software accelerator for efficient embedded DL execution. DeepX significantly lowers the wearable resources (e.g., memory, computation, and energy) required by DL which is a severe bottleneck to mobile (smartphone) adoption. DeepX [89] is an embedded efficiently executable large-scale DL model on mobile (smartphone) processors versus the existing cloud-based offloading. Demo [92] and DeepX [89–91] are good case studies for adapted and embedded low-powered DL software for mobile devices and smartphones, specialized for wearable and behavioral big data analytics.

5.1.3. Embedded DL for Wearable Multimodal Sensor Data Fusion and Integration. Radu et al. [6] used smartphone and smartwatch for human or user activity recognition (HAR). Data integration and fusion, using DL from smartphone and smartwatch, is the focus of this work [6]. DL, specifically RBM, is proposed in [6] for integration (or fusion) of sensor data from multiple sensors (different modalities). Bhattacharya and Lane [7] performed a smartwatch-centric HAR using DL, specifically RBM. Behavior and context recognition tasks related to smartwatches (such as transportation mode, physical activities, and indoor/outdoor detection) using DL (RBM) is performed and focused in [7]. Although DL-based (RBM) human activity recognition outperforms other alternatives, DL resource consumption is unacceptably high for constrained WT devices like smartwatches. Therefore, a complementary study is conducted in Bhattacharya and Lane [7] related to the overhead of DL (RBM models) on smartwatches.

5.1.4. DeepEye: Embedded DL in Wearables with Built-in Camera for Wearable Image Analytics. Wearables with built-in camera provide us with the opportunities to record our daily activities from different perspectives and angles. This is potentially useful in terms of a low vision over our daily lives. DeepEye [93] is a match-box sized wearable camera capable of running multiple cloud-scale-embedded DL models in the device for almost real-time image analysis without offloading them to the cloud. DeepEye [93] is powered by a commodity wearable processor to address the bottleneck of executing multiple DL models (CNN) on wearable limited resources with specifically limited runtime

TABLE 2: The reviewed literature table.

Wearable data	DSL: FFNN, MLP, DNN, CNN	DL approach	
		DUL: DBN, RBM, AE, VAE, GAN, sequence learning, DRL: DQN, AlphaGo, Deep IRL	DSL: RNN, LSTM, GRU, ConvLSTM, sequence-to-sequence learning
Embedding DL in mobile, wearable, and IoT devices	Lane et al. [86], SparseSep [87, 88], DeepX [89–91], Demo [92], DeepEye [93]	DeepEar [94]	—
Embedded DL for wearable multimodal sensor data fusion and integration	—	Radu et al. [6], Bhattaharya and Lane [7]	—
Embedded DL in wearables with built-in cameras for wearable image analysis	Chen et al. [95], DeepEye [93]	—	—
Embedding DL in iOS mobile devices	Chen et al. [95]	—	—
Embedded DL in smartphones for audio signal analytics	—	DeepEar [94]	—
Embedded DL in mobile sensing framework	Harari et al. [96]	Survey [97], Lane and Georgiev [7, 98], Radu et al. [6]	DeepSense [99], DeepSpy [100]
Mobile crowdsensing framework	Harari et al. [96]	Survey [97], Lane and Georgiev [7, 98], Radu et al. [6]	DeepSense [99], DeepSpy [100]
Time-series data	—	Survey [101]	Gamboa [102]
Mobile big data (MBD)	DeepSpace [103]	Alsheikh et al. [104, 105]	—
Mobile wireless sensor network (WSN) data	—	Marjovi et al. [106]	—
EEG data	Stober et al. [107–110]	Wulsin et al. [111], Narejo et al. [112], Stober et al. [107, 113]	Ma et al. [114]
Physiological data	—	Wang and Shang [115]	—
Big data	Najafabadi et al. [116, 117]	—	—
Image and signal data	Xie et al. [118]	Xie et al. [118]	—
Multimodal sequential data	VINet [119, 120]	—	VINet [119, 120]
Mobile gait analytics	Hannink et al. [121–123]	—	—
Embedded DL for inertial data analytics	Survey [124], Ravi et al. [125–127]	Survey [124]	Survey [124]
Embedding DL in low-power devices for health care	Survey [124], Ravi et al. [125–127]	Survey [124]	Survey [124]
Electronic health-care records data	Survey [128]	DeepPatient [129], Miotto et al. [130, 131], Survey [128]	Gram [132], Choi et al. [133–135], DoctorAI [136], Survey [128]
Electronic medical records data	Deepr [137]	Nemati [138]	DeepCare [139]
ECG data	Shashikumar [140]	—	—
Cybersecurity data	—	—	DeepSpy [100]
Smartglass and smartglove data	Advani [4]	—	—
Wearable 3D point cloud data	Poggi et al. [141, 142], Ji et al. [143]	—	—
Multimodal physiological data	—	Du et al. [144]	Alhanai and Ghassemi [145]

memory. Chen et al. [95] propose an embedded deep CNN into iOS smartphones by maximizing data reusability for approaching the high bandwidth burden in DL, specifically the convolution layers of CNN. The effective data reuse makes it possible to parallelize all the computing threads without data loading latency. Chen et al. [95] enhance the capability of DL on local iOS mobile (smartphone) devices.

5.1.5. DeepEar: Embedded DL in Smartphones for Audio Signal Analytics. Regarding mobile audio sensing and analysis, DL has radically changed related audio modeling domains like speech recognition [146]. DeepEar [94] is

a framework for mobile audio sensing using DL, which is trained in an unsupervised setting using a large-scale unlabeled dataset (big audio data) from 168 place visits. With 2.3 M parameters, DeepEar [94] is more robust to background noise compared to conventional approaches in the wearables, specifically in smartphones (mobile devices).

5.2. Embedded DL in Mobile Sensing Framework. Lane et al. [97] is a survey on mobile sensing architecture composed of sensing, learning, and distribution. This survey [97] reviews the existing mobile phone sensing algorithms, applications, and systems related to the architectural framework for

mobile phone sensing research. Harari et al. [96] discusses the potentials and limits of smartphones in collecting wearable biometric and physiological data for behavioral science since smartphones help us enormously collect continuous behavioral data in our daily lives without attracting any attention. The collected continuous behavioral data includes social interactions, daily activities (physical activity), and mobility patterns. Harari et al. [96] look at the practical guidelines for facilitating the use of smartphones as a behavioral observation tool in psychological science. Lane and Georgiev [98] provide a low-power embedded DL using a smartphone System-on-Chip (SoC). This work highlights the critical need for further exploration of DL in mobile sensing towards robust and efficient wearable sensor data inference. DeepSense [99] is a DL framework to address the noisy mobile sensor data and feature engineering problems in mobile sensing. DeepSense [99] integrates CNN and RNN to extract temporal and spatial patterns in the mobile sensor data dynamics for car tracking, HAR, and user identification.

5.3. DL for Time-Series Data Analytics. In many real-world applications (e.g., speech recognition or sleep stage classification), data are collected over the course of time. This time-series data contains temporal patterns related to different classes of behaviors (behavior prediction). Hand-crafted features are expensive to extract since they require the expert knowledge of the field. That is why DUL offers powerful feature learning for time-series data analysis and forecast (prediction). Since wearable data are often collected as time-series signal data, DL plays an important role for learning and recognizing (inference) the temporal pattern in this data. In this aspect, LSTM [58] is dominating other DL approaches. A review of the recent developments, in DUL for time-series data, is given by Långkvist et al. [101] and Gombao [102]. Although DL has shown promising performance in modeling the static data (e.g. computer vision and image classification [31]), applying them to time-series data has not yet been well-studied and explored (understudied). Långkvist et al. [101] and Gombao [102] provide current challenges, projects, and works that either applied DL to time-series data analysis or modified the DL to account for the current challenges in time-series data.

5.4. DL for Mobile Big Data Analytics. The availability smartphones and IoT gadgets led to the recent mobile big data (MBD) era. Collecting MBD is profitable if there is learning methods for analytics to recognize the hidden spatial and temporal patterns from the collected MBD. Alsheikh et al. [104] propose DL in MBD analytics as a scalable learning framework over Apache Spark. Mobile crowdsensing is an efficient MBD collection approach combining the crowd intelligence, smartphones, wearables, and IoT devices (gadgets). Regarding MBD analytics, Alsheikh et al. [105] focuses on the accuracy and privacy aspects of mobile and people-centric crowdsensing as a true MBD collection approach by service providers. DeepSpace,

Ouyang et al. [103], is a DL approach for MBD analytics applied to predicting human trajectory by understanding their mobility patterns. DeepSpace [103] is composed of two models: course and fine prediction models.

5.5. DL for Mobile Wireless Sensor Network Data Analytics. Marjovi et al. [106] explains how to collect data using mobile wireless sensor network (WSN) on public transportation vehicles and analyzing them using DL (AE) for temporal pattern recognition.

5.6. DL for EEG Data Analytics. Stober et al. [107–110, 113] are applying DL approaches for classifying and recognition of EEG recordings for rhythm perception. It specifically applied stacked AE and CNN on the collected EEG data to distinguish the rhythms on a group and individual participants. Given the EEG data, Stober et al. [107–110, 113] use DL for detection and classification of EEG signal in terms types and genres. Wulsin et al. [111] also model EEG waveform data (brain time-series signal) for anomaly measurement, detection, and recognition (classification) using DL approaches, specifically DBN. Narejo et al. [112] classify EEG data (brain time-series signal) for eye states using DUL, specifically DBN and AE. DL for compressed sensing, in brain-computer interface (BCI), is demonstrated in Ma et al. [114] for extracting the motion-onset visual evoked potential (mVEP) BCI features. Ma et al. [114] combine DL with compressed sensing to analyze discriminative mVEP features to improve the mVEP BCI performance. Ma et al. [114] demonstrate DL effectiveness for extracting the mVEP feature for compressed sensing in BCI systems.

5.7. DL for Physiological Data Analytics. Wang and Shang [115] modeled physiological data (time-series biometric signals) using DL, specifically DBN. DBN, as a DUL approach, can automatically extract features from raw physiological data of multiple channels. Using the pretrained DBN, Wang and Shang [115] built multiple classifiers to predict the levels of arousal, valance, and liking based on the learned features. Based on the experimental results, DBN is applied to raw physiological data effectively learns relevant features, emotional patterns, and predict emotions.

5.8. DL for Big Data Analytics. Big data analytics and DL are two highly focused areas in the data science. Big data is the result of collecting massive amounts of data with useful information in different domains such as national intelligence, cybersecurity, fraud detection, marketing, and medical informatics [147]. DL can extract high-level abstractions as data representation layers through a hierarchical learning process. A key benefit of DL is the analysis through learning the massive amounts of unsupervised data. This key benefit makes DL an extremely valuable tool for big data analytics since the available raw data are largely unlabeled, unannotated, and uncategorized. Najafabadi et al. [116, 117] explore how DL is utilized for big data analytics by

extracting complex patterns from massive volumes of data, semantic indexing, data tagging, fast information retrieval, and simplifying discriminative tasks. Najafabadi et al. [116, 117] also investigate DL in terms of analyzing the streaming data, high-dimensional data, scalability of models, and distributed computing.

5.9. DL for Mobile Gait Analytics. Hannink et al. [121–123] estimate mobile stride length in human gait using DL, specifically deep CNN. Spatial gait pattern recognition and mobile gait analysis are performed in [121–123] to address motor impairment in neurological disease. Deep CNN is used for stride length estimation to map stride-specific inertial sensor data to the resulting stride length.

5.10. Embedded DL for Inertial Data Analytics. In Ravi et al.'s studies [124–127], DL is applied to inertial sensor data analysis for real-time human activity recognition & classification.

5.11. DL for Electronic Healthcare Records Data Analytics. dos Santos et al. [128] discuss DL applications in health-care management and diagnostics as most of the studies suggest DL for clinical diagnosis due to its accurate pattern recognition of disease in electronic medical records (EMR). Based on Dos Santos and Carvalho [128], DL assists in medical decisions, the accuracy of the diagnosis, and medical treatment recommendations. DL for clinical data analysis is discussed in Miotto et al. [129–131]. DeepPatient [129] is an application of DL for massive patient electronic health-care records (EHR) data analytics and prediction. Miotto et al. [129, 131] clearly demonstrate the transition from ML approaches [130] to DL due to the fact that DL overperformed ML on patients' massive EHR datasets. Choi et al. [132–134, 136] review DL approaches and applications for EHR for population health research.

5.12. DL for Electronic Medical Records Data Analytics. An electronic medical record (EMR) is a digital paper chart containing the patient's medical history. Personalized predictive medicine requires modeling of patient illness and care processes long-term temporal patterns.

5.12.1. DeepCare: Personalized Medicine Recommender System. DeepCare [139] analyze and recognize the patients' EMRs long-term temporal patterns. Health-care observations, recorded in EMRs, are episodic and irregular in time. EMRs are collected via health-care observations, patient's disease, and personal care history. DeepCare [139] reads EMRs, predicts future medical outcomes, and recommends proper medications. DeepCare models patient health state trajectories with explicit memory of illness. Built on LSTM [58], DeepCare introduces time parameterizations to handle irregular timing by moderating the forgetting and consolidation of illness memory.

5.12.2. DeepR: Deep Record for EMR Data Analytics. Nguyen et al. [137] propose DeepR (deep record) for analyzing the massive EMRs in medicine. DeepR [137] is a predictive system for analyzing EMRs and detecting predictive regular clinical motifs from irregular episodic records. DeepR is an end-to-end DL system to extract features from EMRs and predicts automatically any future risk and transforms a record into a sequence of discrete elements separated by coded time gaps and hospital transfers.

5.12.3. Deep Reinforcement Learning for Clinical EMR Data Analysis in Medication Dosing. Nemati et al. [138] optimizes medication dosing from suboptimal clinical examples using the DRL approach. A clinician-in-the-loop sequential decision-making framework [138] is proposed for an individualized dosing policy of each patient's evolving clinical phenotype using the publicly available MIMIC II intensive care unit database with a DRL that learns an optimal heparin dosing policy from sample dosing trails and their associated outcomes in large EMRs. The proposed DRL system [138] demonstrates that a sequential modeling approach, learned from retrospective data, could potentially be used at the bedside to derive individualized patient dosing policies.

5.13. DL for ECG Data Analytics. Wearables have enormous potential to provide low-risk and low-cost long-term monitoring of electrocardiography (ECG), but these signals highly suffer from significant movement-related noise. Shashikumar et al. [140] present DL-based atrial fibrillation (AF) detection in a sequence of short windows with significant movement artifact. Pulsatile photoplethysmographic (PPG) data and triaxial accelerometry were captured using a multichannel wrist-worn device. A single-channel electrocardiogram (ECG) was recorded (for rhythm verification only) simultaneously. A DL approach was developed on these data to classify AF from wrist-worn PPG signals. A continuous wavelet transform was applied to the PPG data, and CNN was trained on the derived spectrograms to detect AF.

5.14. DL for Cybersecurity Data Analytics. DeepSpying [100] is a mobile-sensing framework for data collection and DL for data analytics in information security (i.e., cybersecurity) domain to protect individual privacy. DeepSpying [100] pioneers WT-based data collection and DL-based data analysis for patient's information security and privacy protection.

5.15. DL for Smartglass and Smartglove Data Analytics. Advani et al. [4] build a multitask AI visual-assistance system for assisting visually impaired people in grocery shopping using smart glass, smart glove, and shopping carts for providing auditory and tactile feedback. This AI system [4] is part of the visual cortex on Silicon project aimed at developing interfaces, algorithms, and hardware platforms

to assist the visually impaired with a focus on grocery shopping.

5.16. DL for Wearable 3D Point Cloud Data Processing and Analytics. Poggi et al. [141, 142] recognize the crosswalk (i.e., crosswalk recognition) on the route using DL for point cloud processing (i.e., 3D data learning) with a suitable wearable mobility aid for the visually impaired people. Poggi and Mattoccia [142] present a wearable mobility aid for the visually impaired individuals using embedded 3D vision and DL-based approach. Poggi et al. [141] relies on an RGBD camera and FPGA embedded in a wearable eyeglass for effective point cloud data processing with a compact and lightweight embedded computer. The computer also provides feedback to the user using a haptic interface as well as audio messages. Poggi et al. [141] does crosswalk recognition for several visually impaired users as a crucial requirement in an effective design of a mobility aid. Poggi et al. [141] propose a system to detect and categorize crosswalks by leveraging on point-cloud processing and DL techniques. Ji et al. [143] processes 3D data using CNN for HAR. They develop a novel 3D CNN model for action recognition of both the spatial and the temporal patterns using 3D convolutions for capturing the motion information encoded in multiple adjacent frames. They also apply the developed models to HAR in the real-world environment of airport surveillance videos.

5.17. DL for Multimodal Physiological Data Analytics. Du et al. [144, 145] discuss the effects of DL in mortality prediction. In these works [144, 145], a combination of auditory, text, and physiological signals are utilized to predict the mood (happy or sad) of 31 narrations from subjects engaged in personal story-telling. They extracted 386 audio and 222 physiological features (using the Samsung wearable simband) from the data. A subset of 4 audio, 1 text, and 5 physiologic features was identified using sequential forward selection (SFS) for inclusion in DNN. These features included subject movement, cardiovascular activity, energy in speech, probability of voicing, and linguistic sentiment (i.e., negative or positive).

6. WearableDL: Future Insights

We have presented a biologically inspired architecture "WearableDL" for the wearable big data analytics that resembles the human NS. We also reviewed briefly the current frontiers in AI, specifically DL approaches and architecture. We carefully selected more than 100 recently published research articles related to WearableDL architecture with focus on DL, IoT, and WT (Section 4). Although WearableDL meets with obstacles and challenges, we believe that it could be practically and potentially useful especially when the wearable data are massive (volume), heterogeneous (variety), and sampled at different frequency (velocity). In this section, we intend to provide our future view of "WearableDL" challenges and its potential application to wearable big data analytics.

6.1. Health Insurance Decision Making. DL brings a great promise and could increase the value of the wearable big data by making them actionable, e.g., health insurance companies thrive on the data to minimize the cost. Therefore, it becomes extremely important that they learn more about their customers and their lifestyles. They are also interested in knowing the information such as how often their customers perform physical activity such as walking, jogging, or other exercises. The health insurance industry wants to track if their customers have smoking or drinking habits. Due to the promise of accuracy, precision, and efficiency, the application of DL on personalized wearable data can play a major role to estimate the insurance policy cost and also to give rebates if their customers cultivate healthy habits [148]. The trend of a data-driven health insurance policy has already been considered in many countries including North America, Europe, and Asia.

6.2. High Performance in Sports and Athletics. Another area that will be impacted by DL is the billion-dollar sports industry. The performance of athletes is not only a moment of pride for their country or state team but also an economical model and therefore the athletes strive to outperform. Today, they use WT in their training to improve their performance inch by inch [149]. Such precision in their performance also demands WT to offer fine-grain quality in the measurement of body's kinematic motion such as agility and balance and physiological parameters such as heart rate, oxygenation, and muscular strength. Various DL methods can be applied to analyze highly sampled wearable big data and extract the actionable information to improve sports performance. DL could also help detect sports injuries during the game or in the training, so effective decisions are made in time.

6.3. Supporting Elderly Population. Aging population across the globe is a well-known phenomenon. By 2030, 20% or more population will be 65+ years of age [150]. This indicates that we will need to seek technological solutions to support senior citizens who are more prone to disorders, severe health conditions, and injuries due to decaying physical and mental capabilities. In the last decade, WT have specifically been targeted to provide health-care services and comfortable assisted living. However, it is not enough to just collect the data from WT. It is equally important to make the WT personalized to the specific condition experienced by an elderly individual. DL could fill this gap by learning the daily patterns in the wearable big data and by offering the decision makers the relation between the historic and current data. In this way, DL could lead the prediction of underlying health conditions which are often not detected by WT alone.

6.4. Challenges. Although DL comes with several promises for the wearable big data, it also needs to overcome a number of barriers and obstacles for its wide spread adoption.

6.4.1. Unlabeled Wearable Big Data. This is a very common important problem when it comes to analyzing the wearable

big data since this data are often collected in a complete unlabeled or unannotated fashion. That is why UL is becoming an important scope for applying DL to the collected big data. As reviewed and talked about, this scope is often known as DUL and it is still an active area of research, specifically when it comes to wearable big data which are time-series and sequential. Sequence learning is one of the attractive ways approaching this problem using LSTM, RNN, and ConvLSTM.

6.4.2. Computational Bottlenecks, Demand, and Complexity.

Currently, deep models face the burden of computational demand to achieve exceptional performance on large-size datasets. These models are currently aimed to run on cloud servers. However, fog computers which require lightweight algorithms will demand new type of DL models that learn from small datasets. As also mentioned in Section 4 and Table 2, embedding DL into mobile, wearable, and IoT devices has two important bottlenecks: memory bandwidth for matrices and computational power for matrix multiplication operation in parallel or distributed setting.

6.4.3. Data Reliability. In many situations, data collected by wearable devices can be affected by noise and error due to nonideal collection setting, particularly for structured and complex data. In this regard, the wearable devices can be designed to perform a presifting and prefiltering of the data. Therefore, DL can be applied to identify and isolate the corrupted data in the decision-making process. DL can generalize the data in an extraordinary way and that is how it can isolate the corrupted/noisy data and identify the distinctive, repetitive, and robust spatiotemporal pattern in such data.

Abbreviations

AGI:	Artificial general intelligence
AI:	Artificial intelligence
ANN:	Artificial neural net
Backprop:	Backpropagation
BCI:	Brain-computer interface
BP:	Backpropagation
BPTT:	Backpropagation through time
CL:	Cortical learning
CLA:	Cortical learning algorithm
ConvLSTM:	Convolutional LSTM
CNN:	Convolutional neural net or ConvNet
CNS:	Central nervous system
DBN:	Deep Boltzmann machine
DBS:	Deep brain stimulation
DL:	Deep learning
DRL:	Deep reinforcement learning
DSL:	Deep supervised learning
DNN:	Deep neural network
DQN:	Deep Q-network
DUL:	Deep unsupervised learning
EA:	Evolutionary algorithm
ECG:	Electrocardiography
EEG:	Electroencephalography

EM:	Expectation maximization
EMG:	Electromyography
EMR:	Electronic medical record
FCN:	Fully connected network
FFNN:	Feed-forward neural network
FLANN:	Fast library for approximate nearest neighbor
FPGA:	Field programmable gate arrays
GAN:	Generative adversarial nets
GNMT:	Google neural machine translation
GPU:	Graphical processing unit
GQN:	Generative query network
GRU:	Gated recurrent units
HR:	Heart rate
HPC:	High-performance computing
HTM:	Hierarchical temporal memory
HAR:	Human activity recognition
IoT:	Internet-of-Things
IRL:	Inverse reinforcement learning
IC:	Integrated circuits
ICA:	Independent component analysis
KNN:	K-nearest neighbors
LDA:	Linear discriminant analysis
LSTM:	Long short-term memory
ML:	Machine learning
MLP:	Multilayer perceptron
MBD:	Mobile big data
MSE:	Mean-squared error
NS:	Nervous system
NN:	Neural net
NLP:	Natural language processing
PNS:	Peripheral nervous system
PCA:	Principle component analysis
RBM:	Restricted Boltzmann machine
RL:	Reinforcement learning
RNN:	Recurrent neural nets
STDP:	Spike-time-dependent plasticity
SNN:	Spike neural net
SL:	Supervised learning
SSE:	Sum of squared error
SVM:	Support vector machine
TL:	Transfer learning
UL:	Unsupervised learning
VAE:	Variational autoencoder
WIoT:	Wearable Internet-of-things
WT:	Wearable tech.
WearableDL:	Wearable deep learning.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under grant numbers 1652538 and 1565962. Research reported in this publication was also supported by the National Institute of Mental Health/National Institutes of Health (NIH) under award no. 1R01MH108641-01A1.

References

- [1] E. R. Kandel, J. Schwartz, and T. M. Jessell, *Principles of Neuroscience*, McGraw-Hill Education, New York, NY, USA, 4th edition, 2000.
- [2] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998.
- [3] Key investments opportunities for data-driven healthcare," http://www.topbots.com/5-key-investment-opportunities-data-driven-healthcare-ai/?utm_medium=article&utm_source=facebook&utm_campaign=aihealthcare5&utm_content=groups.
- [4] S. Advani, P. Zientara, N. Shukla et al., "A multitask grocery assist system for the visually impaired: smart glasses, gloves, and shopping carts provide auditory and tactile feedback," *IEEE Consumer Electronics Magazine*, vol. 6, no. 1, pp. 73–81, 2017.
- [5] S. Hiremath, G. Yang, and K. Mankodiya, "Wearable internet of things: concept, architectural components and promises for person-centered healthcare," in *Proceedings of 2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*, pp. 304–307, Milan, Italy, November 2014.
- [6] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar, "Towards multimodal deep learning for activity recognition on mobile devices," in *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 185–188, Heidelberg, Germany, 2016.
- [7] S. Bhattacharya and N. D. Lane, "From smart to deep: robust activity recognition on smartwatches using deep learning," in *Proceedings of 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pp. 1–6, Sydney, Australia, March 2016.
- [8] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–12, 2010.
- [9] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G.-Z. Yang, "Big data for health," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1193–1208, 2015.
- [10] Y.-L. Zheng, X.-R. Ding, C. C. Y. Poon et al., "Unobtrusive sensing and wearable devices for health informatics," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1538–1554, 2014.
- [11] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.
- [12] J. L. Solé, "Book review: pattern recognition and machine learning, Christopher M. Bishop, information science and statistics, Springer 2006, 738 pages," *SORT-Statistics and Operations Research Transactions*, vol. 31, no. 2, 2007.
- [13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [14] J. Hawkins and S. Ahmad, "Why neurons have thousands of synapses, a theory of sequence memory in neocortex," *Frontiers in Neural Circuits*, vol. 10, 2016.
- [15] S. Ahmad and J. Hawkins, "How do neurons operate on sparse distributed representations? a mathematical theory of sparsity, neurons and active dendrites," arXiv preprint arXiv:1601.00720, 2016.
- [16] J. Hawkins, S. Ahmad, and D. Dubinsky, "Hierarchical temporal memory including HTM cortical learning algorithms," Technical report, Numenta, Inc., Redwood City, CA, USA, 2011, <http://www.numenta.com/htm-overview/education/HTM/CorticalLearningAlgorithms.pdf>.
- [17] J. Hawkins, S. Ahmad, and Y. Cui, "Why does the neocortex have layers and columns, a theory of learning the 3d structure of the world," bioRxiv, p. 162263, 2017.
- [18] S. Ahmad and J. Hawkins, "Properties of sparse distributed representations and their application to hierarchical temporal memory," arXiv preprint arXiv:1503.07469, 2015.
- [19] S. Billaudelle and S. Ahmad, "Porting HTM models to the heidelberg neuromorphic computing platform," arXiv preprint arXiv:1505.02142, 2015.
- [20] Y. Cui, S. Ahmad, and J. Hawkins, "The HTM spatial pooler: a neocortical algorithm for online sparse distributed coding," bioRxiv, p. 085035, 2016.
- [21] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," 2016.
- [22] Y. Bengio, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [23] P. Andras, "High-dimensional function approximation with neural networks for large volumes of data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 2, pp. 500–508, 2017.
- [24] 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.
- [25] D. L. Yamins and J. J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex," *Nature Neuroscience*, vol. 19, no. 3, pp. 356–365, 2016.
- [26] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, "Random synaptic feedback weights support error back-propagation for deep learning," *Nature Communications*, vol. 7, p. 13276, 2016.
- [27] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, p. 607, 1996.
- [28] Y. Freund and R. E. Schapire, "Large margin classification using the perceptron algorithm," *Machine Learning*, vol. 37, no. 3, pp. 277–296, 1999.
- [29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error-propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, no. 6088, pp. 318–362, MIT Press, Cambridge, MA, USA, 1986.
- [30] R. Collobert, "Deep learning for efficient discriminative parsing," in *Proceedings of AISTATS*, vol. 15, pp. 224–232, Ft. Lauderdale, FL, USA, April 2011.
- [31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, NV, USA, December 2012.
- [32] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [33] Y. Wu, M. Schuster, Z. Chen et al., "Google's neural machine translation system: bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.
- [34] M. Johnson, M. Schuster, Q. V. Le et al., "Google's multilingual neural machine translation system: enabling

- zero-shot translation,” arXiv preprint arXiv:1611.04558, 2016.
- [35] “How google detects cancer using deep learning,” <http://www.techtimes.com/articles/200311/20170306/here-s-how-google-s-ai-helps-detect-cancer-via-deep-learning.htm>.
 - [36] “Dl advances in medicine,” <http://www.nvidia.com/object/deep-learning-in-medicine.html>.
 - [37] “Dl analytics for healthcare: UCSF, intel join forces to develop deep learning analytics for health care,” <https://www.ucsf.edu/news/2017/01/405536/ucsf-intel-join-forces-develop-deep-learning-analytics-health-care>.
 - [38] D. H. Hubel and T. N. Wiesel, “Receptive fields of single neurones in the cat’s striate cortex,” *Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.
 - [39] K. Fukushima and S. Miyake, “Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition,” in *Competition and Cooperation in Neural Nets*, pp. 267–285, Springer, Berlin, Germany, 1982.
 - [40] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in *Proceedings of Advances in Neural Information Processing Systems*, pp. 3856–3866, Long Beach, CA, USA, December 2017.
 - [41] S. Bartunov, A. Santoro, B. A. Richards, G. E. Hinton, and T. Lillicrap, “Assessing the scalability of biologically-motivated deep learning algorithms and architectures,” arXiv preprint arXiv:1807.04587, 2018.
 - [42] K. Fukushima, “Cognitron: a self-organizing multilayered neural network,” *Biological Cybernetics*, vol. 20, no. 3-4, pp. 121–136, 1975.
 - [43] K. Fukushima, “Neocognitron: a hierarchical neural network capable of visual pattern recognition,” *Neural Networks*, vol. 1, no. 2, pp. 119–130, 1988.
 - [44] K. Fukushima, “Artificial vision by multi-layered neural networks: neocognitron and its advances,” *Neural Networks*, vol. 37, pp. 103–119, 2013.
 - [45] Y. LeCun, B. Boser, J. S. Denker et al., “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
 - [46] J. Schmidhuber, “Deep learning in neural networks: an overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
 - [47] B. C. Csáji, “Approximation with artificial neural networks,” vol. 24, p. 48, M.Sc. thesis, Faculty of Sciences, Eötvös Loránd University, Hungary, M.Sc. thesis, Faculty of Sciences, Eötvös Loránd University, 2001.
 - [48] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of Control, Signals, and Systems (MCSS)*, vol. 2, no. 4, pp. 303–314, 1989.
 - [49] K. Hornik, “Approximation capabilities of multilayer feed-forward networks,” *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991.
 - [50] J. Dean, G. Corrado, R. Monga et al., “Large scale distributed deep networks,” in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1223–1231, Lake Tahoe, NV, USA, December 2012.
 - [51] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
 - [52] G. E. Hinton, “Deep belief networks,” *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.
 - [53] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
 - [54] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.
 - [55] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
 - [56] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Playing Atari with deep reinforcement learning,” arXiv preprint arXiv:1312.5602, 2013.
 - [57] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” Technical Report, California Univ San Diego La Jolla Inst, Oakland, CA, USA, 1985.
 - [58] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
 - [59] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
 - [60] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, “Convolutional LSTM network: a machine learning approach for precipitation nowcasting,” in *Proceedings of Advances in Neural Information Processing Systems*, pp. 802–810, Montreal, Canada, December 2015.
 - [61] Y. Zhang, W. Chan, and N. Jaitly, “Very deep convolutional networks for end-to-end speech recognition,” arXiv preprint arXiv:1610.03022, 2016.
 - [62] Y. Bengio, A. C. Courville, and P. Vincent, “Unsupervised feature learning and deep learning: a review and new perspectives,” *CoRR*, abs/1206.5538, vol. 1, 2012.
 - [63] S. Yeung, V. Ramanathan, O. Russakovsky, L. Shen, G. Mori, and L. Fei-Fei, “Learning to learn from noisy web videos,” arXiv preprint arXiv:1706.02884, 2017.
 - [64] C. Rupprecht, I. Laina, M. Baust, F. Tombari, G. D. Hager, and N. Navab, “Learning in an uncertain world: representing ambiguity through multiple hypotheses,” arXiv preprint arXiv:1612.00197, 2016.
 - [65] M. Mirza, A. Courville, and Y. Bengio, “Generalizable features from unsupervised learning,” arXiv preprint arXiv:1612.03809, 2016.
 - [66] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle et al., “Greedy layer-wise training of deep networks,” in *Proceedings of Advances in Neural Information Processing Systems*, vol. 19, p. 153, Vancouver, British Columbia, Canada, December 2007.
 - [67] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., “Generative adversarial nets,” in *Proceedings of Advances in Neural Information Processing Systems*, pp. 2672–2680, Montreal, Canada, December 2014.
 - [68] C. Finn, X. Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel, “Deep spatial autoencoders for visuomotor learning,” in *Proceedings of International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, May 2016.
 - [69] R. S. Sutton, “Temporal credit assignment in reinforcement learning,” Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA, USA, 1984.
 - [70] D. Silver, A. Huang, C. J. Maddison et al., “Mastering the game of go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
 - [71] D. Silver, J. Schrittwieser, K. Simonyan et al., “Mastering the game of go without human knowledge,” *Nature*, vol. 550, no. 7676, p. 354, 2017.
 - [72] A. Banino, C. Barry, B. Uria et al., “Vector-based navigation using grid-like representations in artificial agents,” *Nature*, vol. 557, no. 7705, p. 429, 2018.


- [73] T. J. Wills, F. Cacucci, N. Burgess, and J. O'Keefe, "Development of the hippocampal cognitive map in preweanling rats," *Science*, vol. 328, no. 5985, pp. 1573–1576, 2010.
- [74] M. Fyhn, T. Hafting, M. P. Witter, E. I. Moser, and M.-B. Moser, "Grid cells in mice," *Hippocampus*, vol. 18, no. 12, pp. 1230–1238, 2008.
- [75] D. Kalashnikov, A. Irpan, P. Pastor et al., "Qt-opt: scalable deep reinforcement learning for vision-based robotic manipulation," arXiv preprint arXiv:1806.10293, 2018.
- [76] C. Finn, P. Christiano, P. Abbeel, and S. Levine, "A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models," arXiv preprint arXiv:1611.03852, 2016.
- [77] A. Y. Ng and S. J. Russell, "Algorithms for inverse reinforcement learning," in *Proceedings of ICML*, pp. 663–670, Stanford, CA, USA, June–July 2000.
- [78] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of Twenty-First International Conference on Machine Learning*, Banff, Alberta, Canada, July 2004.
- [79] P. Pieter and A. Y. Ng, "Inverse reinforcement learning," in *Encyclopedia of Machine Learning*, pp. 554–558, Springer, Berlin, Germany, 2011.
- [80] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 4565–4573, Barcelona, Spain, December 2016.
- [81] N. Baram, O. Anschel, and S. Mannor, "Model-based adversarial imitation learning," arXiv preprint arXiv:1612.02179, 2016.
- [82] B. Wang, T. Sun, and S. X. Zheng, "Beyond winning and losing: modeling human motivations and behaviors using inverse reinforcement learning," arXiv preprint arXiv:1807.00366, 2018.
- [83] D. Ha and J. Schmidhuber, "Recurrent world models facilitate policy evolution," arXiv preprint arXiv:1809.01999, 2018.
- [84] Y. Zhu, Z. Wang, J. Merel et al., "Reinforcement and imitation learning for diverse visuomotor skills," arXiv preprint arXiv:1802.09564, 2018.
- [85] S. A. Eslami, D. J. Rezende, F. Besse et al., "Neural scene representation and rendering," *Science*, vol. 360, no. 6394, pp. 1204–1210, 2018.
- [86] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, and F. Kawsar, "An early resource characterization of deep learning on wearables, smartphones and internet-of-things devices," in *Proceedings of 2015 International Workshop on Internet of Things towards Applications*, pp. 7–12, Seoul, South Korea, 2015.
- [87] N. Lane and S. Bhattacharya, "Sparsifying deep learning layers for constrained resource inference on wearables," in *Proceedings of 14th ACM Conference on Embedded Network Sensor Systems*, pp. 176–189, Stanford, CA, USA, November 2016.
- [88] S. Bhattacharya and N. D. Lane, "Sparsification and separation of deep learning layers for constrained resource inference on wearables," in *Proceedings of ACM Conference on Embedded Networked Sensor Systems (SenSys) 2016*, Stanford, CA, USA, November 2016.
- [89] N. D. Lane, S. Bhattacharya, P. Georgiev et al., "A software accelerator for low-power deep learning inference on mobile devices," in *Proceedings of 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 1–12, Vienna, Austria, April 2016.
- [90] N. D. Lane, S. Bhattacharya, P. Georgiev et al., "Demonstration abstract: accelerating embedded deep learning using DeepX," in *Proceedings of 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 1–2, Vienna, Austria, 2016.
- [91] N. D. Lane, S. Bhattacharya, P. Georgiev et al., "Accelerating embedded deep learning using DeepX: demonstration abstract," in *Proceedings of 15th International Conference on Information Processing in Sensor Networks*, p. 61, Vienna, Austria, April 2016.
- [92] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, and F. Kawsar, "Demo: accelerated deep learning inference for embedded and wearable devices using DeepX," in *Proceedings of 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*, , Singapore, June 2016.
- [93] A. Mathur, N. D. Lane, S. Bhattacharya, A. Boran, C. Forlivesi, and F. Kawsar, *Deepeye: Resource Efficient Local Execution of Multiple Deep Vision Models Using Wearable Commodity Hardware*, ACM, New York, NY, USA, 2017.
- [94] N. D. Lane, P. Georgiev, and L. Qendro, "Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning," in *Proceedings of 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 283–294, Osaka, Japan, September 2015.
- [95] C.-F. Chen, G. G. Lee, V. Sritapan, and C.-Y. Lin, "Deep convolutional neural network on iOS mobile devices," in *Proceedings of 2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 130–135, Dallas, TX, USA, 2016.
- [96] G. M. Harari, N. D. Lane, R. Wang, B. S. Crosier, A. T. Campbell, and S. D. Gosling, "Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges," *Perspectives on Psychological Science*, vol. 11, no. 6, pp. 838–854, 2016.
- [97] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.
- [98] N. D. Lane and P. Georgiev, "Can deep learning revolutionize mobile sensing?," in *Proceedings of 16th International Workshop on Mobile Computing Systems and Applications*, pp. 117–122, Santa Fe, NM, USA, February 2015.
- [99] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: a unified deep learning framework for time-series mobile sensing data processing," arXiv preprint arXiv:1611.01942, 2016.
- [100] T. Beltramelli and S. Risi, "Deep-spying: spying using smartwatch and deep learning," arXiv preprint arXiv:1512.05616, 2015.
- [101] M. Långkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11–24, 2014.
- [102] J. C. B. Gamboa, "Deep learning for time-series analysis," arXiv preprint arXiv:1701.01887, 2017.
- [103] X. Ouyang, C. Zhang, P. Zhou, and H. Jiang, "Deepspace: an online deep learning framework for mobile big data to understand human mobility patterns," arXiv preprint arXiv:1610.07009, 2016.
- [104] M. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and Apache spark," *IEEE Network*, vol. 30, no. 3, pp. 22–29, 2016.

- [105] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, and Z. Han, "The accuracy-privacy tradeoff of mobile crowd-sensing," arXiv preprint arXiv:1702.04565, 2017.
- [106] A. Marjovi, A. Arfire, and A. Martinoli, "Extending urban air quality maps beyond the coverage of a mobile sensor network: data sources, methods, and performance evaluation," in *Proceedings of International Conference on Embedded Wireless Systems and Networks*, Madrid, Spain, February 2017.
- [107] S. Stober, D. J. Cameron, and J. A. Grahn, "Classifying EEG recordings of rhythm perception," in *Proceedings of ISMIR*, pp. 649–654, Taipei, Taiwan, 2014.
- [108] S. Stober, D. J. Cameron, and J. A. Grahn, "Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1449–1457, Montreal, Canada, December 2014.
- [109] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," arXiv preprint arXiv:1511.04306, 2015.
- [110] A. Sternin, S. Stober, J. Grahn, and A. Owen, "Tempo estimation from the EEG signal during perception and imagination of music," in *Proceedings of International Workshop on Brain-Computer Music Interfacing/International Symposium on Computer Music Multidisciplinary Research (BCMI/CMMR)*, Plymouth, UK, June 2015.
- [111] D. Wulsin, J. Gupta, R. Mani, J. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of Neural Engineering*, vol. 8, no. 3, p. 036015, 2011.
- [112] S. Narejo, E. Pasero, and F. Kulsoom, "EEG based eye state classification using deep belief network and stacked autoencoder," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 6, pp. 3131–3141, 2016.
- [113] S. Stober, "Learning discriminative features from electroencephalography recordings by encoding similarity constraints," in *Proceedings of Bernstein Conference*, London, UK, September 2016.
- [114] T. Ma, H. Li, H. Yang et al., "The extraction of motion-onset VEP BCI features based on deep learning and compressed sensing," *Journal of Neuroscience Methods*, vol. 275, pp. 80–92, 2017.
- [115] D. Wang and Y. Shang, "Modeling physiological data with deep belief networks," *International Journal of Information and Education Technology (IJIET)*, vol. 3, no. 5, p. 505, 2013.
- [116] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, p. 1, 2015.
- [117] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning techniques in big data analytics," in *Big Data Technologies and Applications*, pp. 133–156, Springer International Publishing, Berlin, Germany, 2016.
- [118] D. Xie, L. Zhang, and L. Bai, "Deep learning in visual computing and signal processing," *Applied Computational Intelligence and Soft Computing*, vol. 2017, Article ID 1320780, 13 pages, 2017.
- [119] H. Wen, "Vinet: visual-inertial odometry as a sequence-to-sequence learning problem," in *Proceedings of AAAI*, Phoenix, AZ, USA, February 2016.
- [120] R. Clark, S. Wang, H. Wen, A. Markham, and N. Trigoni, "Vinet: visual-inertial odometry as a sequence-to-sequence learning problem," arXiv preprint arXiv:1701.08376, 2017.
- [121] J. Hannink, T. Kautz, C. F. Pasluosta et al., "Stride length estimation with deep learning," arXiv preprint arXiv:1609.03321, 2016.
- [122] J. Hannink, T. Kautz, C. Pasluosta, K.-G. Gassmann, J. Klucken, and B. Eskofier, "Sensor-based gait parameter extraction with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 85–93, 2017.
- [123] J. Hannink, T. Kautz, C. Pasluosta et al., "Mobile stride length estimation with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 354–362, 2018.
- [124] D. Ravi, C. Wong, F. Deligianni et al., "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.
- [125] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: a resource efficient implementation on low-power devices," in *Proceedings of 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 71–76, San Francisco, CA, USA, June 2016.
- [126] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "A deep learning approach to on-node sensor data analytics for mobile or wearable devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 56–64, 2017.
- [127] D. Ravi, C. Wong, F. Deligianni et al., "Special section on deep learning for biomedical and health informatics," *Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.
- [128] A. B. V. dos Santos and D. R. Carvalho, "Deep learning for healthcare management and diagnosis," *Iberoamerican Journal of Applied Computing*, vol. 5, no. 2, 2016.
- [129] 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, 2016.
- [130] R. Miotto and C. Weng, "Unsupervised mining of frequent tags for clinical eligibility text indexing," *Journal of Biomedical Informatics*, vol. 46, no. 6, pp. 1145–1151, 2013.
- [131] R. Miotto, L. Li, and J. T. Dudley, "Deep learning to predict patient future diseases from the electronic health records," in *Proceedings of European Conference on Information Retrieval*, pp. 768–774, Padua, Italy, March 2016.
- [132] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "Gram: graph-based attention model for healthcare representation learning," arXiv preprint arXiv:1611.07012, 2016.
- [133] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 361–370, 2017.
- [134] E. Choi, M. T. Bahadori, E. Searles et al., "Multi-layer representation learning for medical concepts," in *Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1495–1504, San Francisco, CA, USA, August 2016.
- [135] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Medical concept representation learning from electronic health records and its application on heart failure prediction," arXiv preprint arXiv:1602.03686, 2016.

- [136] E. Choi, M. T. Bahadori, and J. Sun, "Doctor AI: predicting clinical events via recurrent neural networks," arXiv preprint arXiv:1511.05942, 2015.
- [137] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deepir: a convolutional net for medical records," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 22–30, 2016.
- [138] S. Nemati, M. M. Ghassemi, and G. D. Clifford, "Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach," in *Proceedings of 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pp. 2978–2981, Orlando, FL, USA, August 2016.
- [139] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Deepcare: a deep dynamic memory model for predictive medicine," in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 30–41, Auckland, New Zealand, April 2016.
- [140] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proceedings of 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 141–144, Orlando, FL, USA, February 2017.
- [141] M. Poggi, L. Nanni, and S. Mattoccia, "Crosswalk recognition through point-cloud processing and deep-learning suited to a wearable mobility aid for the visually impaired," in *Proceedings of International Conference on Image Analysis and Processing*, pp. 282–289, Genoa, Italy, September 2015.
- [142] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning," in *Proceedings of 2016 IEEE Symposium on Computers and Communication (ISCC)*, pp. 208–213, Messina, Italy, June 2016.
- [143] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
- [144] H. Du, M. M. Ghassemi, and M. Feng, "The effects of deep network topology on mortality prediction," in *Proceedings of 2016 IEEE 38th Annual International Conference of Engineering in Medicine and Biology Society (EMBC)*, pp. 2602–2605, Orlando, FL, USA, 2016.
- [145] T. Alhanai and M. M. Ghassemi, *Predicting Latent Narrative Mood Using Audio and Physiologic Data*, MIT, Cambridge, MA, USA, 2017.
- [146] A.-r. Mohamed, G. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Transactions on Audio, Speech, and Language Processing*, no. 99, p. 1, 2010.
- [147] B. Furht and F. Villanustre, *Big Data Technologies and Applications*, Springer, Berlin, Germany, 2016.
- [148] N. Sultan, "Reflective thoughts on the potential and challenges of wearable technology for healthcare provision and medical education," *International Journal of Information Management*, vol. 35, no. 5, pp. 521–526, 2015.
- [149] S. L. Halson, J. M. Peake, and J. P. Sullivan, "Wearable technology for athletes: information overload and pseudo-science?," *International Journal of Sports Physiology and Performance*, vol. 11, no. 6, pp. 705–706, 2016.
- [150] J. M. Ortman, V. A. Velkoff, H. Hogan et al., *An Aging Nation: the Older Population in the United States*, United States Census Bureau, Economics and Statistics Administration, US Department of Commerce, Suitland, MD, USA, 2014.

Research Article

Artificial Intelligence to Prevent Mobile Heart Failure Patients Decompensation in Real Time: Monitoring-Based Predictive Model

Nekane Larburu ^{1,2}, Arkaitz Artetxe,^{1,2} Vanessa Escobar,³ Ainara Lozano,³ and Jon Kerexeta¹

¹*Vicomtech, Paseo Mikeletegi 57, 20009 Donostia/San Sebastian, Spain*

²*Biodonostia Health Research Institute, P. Doctor Begiristain s/n, 20014 San Sebastian, Spain*

³*Hospital Universitario de Basurto (Osakidetza Health Care System), Avda. Montevideo 18, 48013 Bilbao, Spain*

Correspondence should be addressed to Nekane Larburu; nlarburu@vicomtech.org

Received 29 June 2018; Accepted 30 September 2018; Published 5 November 2018

Guest Editor: Giovanna Sannino

Copyright © 2018 Nekane Larburu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rapid advances in ICT and collection of large amount of mobile health data are giving room to new ways of treating patients. Studies suggest that telemonitoring systems and predictive models for clinical support and patient empowerment may improve several pathologies, such as heart failure, which admissions rate is high. In the current medical practice, clinicians make use of simple rules that generate large number of false alerts. In order to reduce the false alerts, in this study, the predictive models to prevent decompensations that may lead into admissions are presented. They are based on mobile clinical data of 242 heart failure (HF) patients collected for a period of 44 months in the public health service of Basque Country (Osakidetza). The best predictive model obtained is a combination of alerts based on monitoring data and a questionnaire with a Naive Bayes classifier using Bernoulli distribution. This predictive model performs with an AUC = 67% and reduces the false alerts per patient per year from 28.64 to 7.8. This way, the system predicts the risk of admission of ambulatory patients with higher reliability than current alerts.

1. Introduction

Since these early days, the advances on ICT have given a huge opportunity to telemedicine applications and new e-Health services [1]. Along with this phenomenon are the large quantities of mobile data that are being collected and processed these days. The growth in these two areas are leading in advanced health-care systems that not only provide continuous support to clinicians or informal care givers (e.g., family members), but also to patients. In this context, telemedicine systems that monitor ambulatory patients and guide them in their daily routine are emerging. Nevertheless, often all the potential of the mobile-health data used to support clinical professionals and patients is not sufficiently exploited. Other times, the exploited clinical data, in the form of, for example, predictive models to identify patients at high risk, are not applied in a real setting to support clinicians and patients.

Studies suggest that artificial intelligence by means of predictive models and telemonitoring systems for clinical support and patient empowerment may improve several pathologies [2], such as heart failure.

Heart failure (HF) is a clinical syndrome caused by a structural and/or functional cardiac abnormality. HF patients suffer decompensations, which is defined by Mangini et al. [3] as a clinical syndrome in which a structural or functional change in the heart leads to its inability to eject and/or accommodate blood within physiological pressure levels, thus causing a functional limitation and requiring immediate therapeutic intervention [3]. Hence, decompensations may lead in hospital admissions, which in this study are defined as emergency admissions and hospital admissions, and home interventions. As Ponikowski et al. presented in [4], the prevalence of HF depends on the definition applied, but it is approximately 1-2% of the adults

in developed countries, rising to more than 10% among people >70 years of age. Hence, due to the aging population, an increase in the number of HF patients is expected in the future. Therefore, predicting the risk of a patient to suffer a decompensation may prevent admissions and readmissions, improving both patient care and hospital management, which has a high impact on costs and clinical professionals time. The first step to predict the risk of decompensation is to telemonitor ambulatory patients. Next, we need reliable systems to assess the risk. Most telemedicine systems apply alerts or rule-based systems to detect potential complications of ambulatory patients [5–8]. But these usually contain large number of false alerts, and hence, these systems are not trustworthy (Table 1).

Our hypothesis is that with the usage of artificial intelligence (AI) by means of, for instance, predictive models, it is possible to detect decompensations of ambulatory patients and reduce false alerts. In this context, this research extends the study for readmissions detection [9] and presents predictive models of a telemedicine system for heart failure patients, called INCAR. INCAR has been developed to (i) be generally applicable in HF patients, (ii) improve the clinical practice by developing an accurate system that detects the risk of decompensation and suggest actions to prevent them on time, (iii) allow professionals to maintain an efficient and personalized support and follow-up of patient, (iv) give patients support when required and guide them in risk situations, informing clinicians accordingly, and (v) reduce HF patients admission and readmission rate, which have a high economic impact.

This paper focuses on the development of predictive models to detect decompensations, and it is structured as follows: First, the *Related Work* section summarizes the state of the art on telemedicine systems for heart failure and the role of predictive models on telemedicine systems. *Materials* section presents the database used in this study and the characteristics of the dataset. *Methods* presents the applied methods to assess the risk of an ambulatory HF patient to suffer a decompensation that may lead into admission. In *Results*, the outcomes obtained for each of the developed predictive models is presented. Finally, *Discussion* presents the results and limitations of the study, and *Conclusion* gives a summary of the contributions and future work.

2. Related Work

2.1. Telemedicine Systems for HF Patients. Being HF a disease with high prevalence and high readmission rate, the usage of telemedicine systems in this area is common [7]. Chaudhry et al. [2] telemonitored patients by means of telephone-based interactive voice-response system and concluded that the simple phone-based telemonitorization does not improve the outcomes (i.e., readmission, death). Nevertheless, most of current telemonitoring systems do not simply implement telephone-based monitorization, but also the transmission of mobile health data, such as bodyweight, heart rate, and blood pressure [7]. Besides, more advanced noninvasive systems transfer electrocardiograph (ECG) tracings, oxygen saturation, and physical activity (e.g., pedometer)

data. Apart from noninvasive telemedicine systems, invasive systems enable the transfer of variables measured invasively, such as transthoracic impedance and pulmonary and left atrial pressures. But literature studies do not present significantly better results when implementing invasive measurements into their telemedicine systems in terms of HF decompensation prevention. Nonetheless, some benefits have been presented when applying impedance instead of weight for detecting HF patients early decompensation, as presented by Abraham et al. [5] and Gyllensten et al. [6].

2.2. Alerts for HF Patients. Most studies implement “simple” alerts to prevent decompensations based on these data. One of the implemented techniques is *Rule of Thumb (RoT)* based on simple rules (i.e., when a measurement goes beyond or below a given threshold or when they are based on simple difference between the current value of an attribute and a previous measurement that occurs a predefined number of days in the past) [5–8]. Other studies, such as Zhang et al. [7], Gyllensten et al. [6], and Ledwidge et al. [8], make also use of more sophisticated techniques, such as the *Moving Average (MA)* or similar techniques that calculate the variations applied to usually weight. The *Cumulative Sum (CUMSUM)*, applied by Adamson et al. [10], is typically used for detecting changes and implies that when a continuous variation of a measurement is produced over time, that tendency will result in an alert. Additionally, Gilliam et al. [11], apply the *multivariate method*, which consists on the usage of several data elements that are incorporated into a multivariate logistic regression model to form the probability of an event occurring. From the studied papers, we could conclude that each type of alert may work best depending on the applied attribute. For instance, techniques related with MA work best when applied to weight. On the other hand, CUMSUM is one of the best methods when applied to transthoracic or intrathoracic impedance.

Table 1 presents the results of different studies that determine whether a monitored HF patient will have a decompensation, usually implementing alerts. Due to the large number of days that do not end in an admission, even when the computed specificity values are high, the number of false positives could remain too high for the clinical practice, so it is not an optimal testing value in this scope. Taken into account this limitation, based on the literature studies, we could consider the number of false alerts per patient per year (FA/pt-y) as de facto standard to determine the number of false positives. However, as shown in Table 1, some of the studies present the specificity value for determining how well the no admissions are detected using own techniques to compute it.

2.3. Predictive Models on Telemedicine Systems. As shown above, most telemedicine systems apply alerts or rule-based systems to detect potential complications of ambulatory patients. This is not only present in the context of HF, but also in diabetes, atrial fibrillation, and other clinical domains [12, 13]. Hence, there is a lack of the usage of collected data that could lead in more accurate solutions by means of, for instance, predictive models.

TABLE 1: Summary of decompensation detection studies.

Study	Data type	Dataset	Method	Results
Zhang et al. [7]	Weight	135 patients; 1964 days monitoring	RoT	Se = 58.3%, Sp = 54.1%
			MACD	Se = 20.4%, Sp = 89.4% (AUC = 0.55%)
Gyllensten et al. [6]	Weight	91 patients; 10 months	RoT	Se = 20%, Sp = 90%
			MACD	Se = 33%, Sp = 91%
			CUMSUM	Se = 13%, Sp = 91%
	Noninvasive transthoracic bioimpedance	91 patients; 10 months	RoT	Se = 13%, Sp = 91%
			MACD	Se = 13%, Sp = 91%
			CUMSUM	Se = 13%, Sp = 91%
Adamson et al. [10]	Blood pressure	274 patients	CUMSUM	Se = 83.1%, FA = 4.1/pt-y
Abraham et al. [5]	Intrathoracic impedance	156 patients; 537 \pm 312 days	RoT	Se = 76.4%; FA = 1.9/pt-y
	Weight	156 patients; 537 \pm 312 days	RoT	Se = 21%; FA = 4.3/pt-y
Ledwidge et al. [8]	Weight	87 patients; 23.9 \pm 12 weeks	RoT	Se = 21%; Sp = 86%
			HeartPhone algorithm (based on MA)	Se = 82%; Sp = 68%
Gilliam et al. [11]	Multivariate	201 patients		Se = 41%; FA = 2/pt-y

Several studies in the context of HF develop predictive models to determine whether a patient will be readmitted within 30 days after discharge [14–20]. These predicting models make use of baseline information of patients, such as age, sex, or left ventricular injection fraction, but not daily (or weekly) tele-monitored patient mobile data, such as weight, heart rate, or blood pressure, which could be crucial for detecting and preventing an ambulatory patient admission. In several telemedicine studies applied in diverse pathologies, such as chronic obstructive pulmonary disease [21, 22] and preeclampsia [23], predictive models have been successfully applied. However, in the context of HF, limited studies apply predictive models. Lafta et al. [24] is one of these studies that using several telemonitored attributes (i.e., heart rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, and oxygen saturation) applied basic time series prediction algorithm, regression-based time series prediction algorithm, and hybrid time series prediction algorithm. The obtained results showed that up to 75% and 98% of accuracy values could be obtained across different patients under three algorithms, but still the accuracy value is not objective enough to determine how well the system performs.

The presented study goes beyond the state of the art and applies classifiers based on alerts applied in current medical practice and state-of-the-art studies. Additionally, this study makes use of baseline information and ambulatory tele-monitored information to build an integral telemedicine system that applies predictive models with double goal: assess ambulatory patients' admission risk to provide both patients and clinicians the appropriate guidance to prevent potential decompensations that may lead to hospital admissions.

3. Materials

3.1. Database. The public hospital OSI Bilbao-Basurto (Osakidetza), located in Basque Country (Spain), has been gathering HF patients' information from June 2014 until

February 2018 (44 months) to closely monitor HF patients. For the present study, the dataset contained a cohort of 242 HF patients. Clinicians have collected baseline data (i.e., information collected by a clinician when the patient is diagnosed, Table 2), ambulatory patient monitored data (i.e., information collect from three to seven times per week, Tables 3 and 4), and patients admissions information (i.e., emergency admissions, hospital admissions, and home care interventions that are associated to HF associated with a patient decompensation).

Besides vital signs, a questionnaire is also included into the telemonitoring system to ask patients about their condition, with potential impact on decompensation prediction (Table 4).

3.2. Characteristics of Ambulatory Patients Dataset. In the whole study, 242 patients have been enrolled from June 2014 until February 2018. Of these 242 patients, one patient has been excluded as it is a cirrhotic man who often has interventions of evacuational paracentesis due to a liver pathology not related to HF. There is an average follow-up of 13.5 ± 9.11 months. In this time period, there have been 254 decompensations of which 202 are considered as predictable, since 52 decompensations do not have previous tele-monitoring information (i.e., less than 3 times in the last week before the decompensation).

4. Methods

Following the methodology applied for the generation of the predictive models is presented: (i) training and testing dataset construction, (ii) application of alerts implemented in current clinical setting, (iii) selection of the alerts for the study, (iv) generation of the dataset to apply the machine learning classifiers, and (v) the application and comparison of different classifiers.

TABLE 2: Baseline characteristics of the study population.

Characteristics	Description	Median \pm SD (percentage)
Age	The age of the patient (years)	78 \pm 10.9
Height	The height of the patient (mm)	162.37 \pm 10.34
Sex	The sex of the patient (men/women)	57% men
Smoker	If the patient smokes, did smoke, and now do not or never has smoked	15.35% do smoke, 22% did smoke (not now)
LVEF	Left ventricular ejection fraction (%)	42.4 \pm 15.21
First diag	Years since first diagnosis	5.8 \pm 7.04
Implanted device	If implanted device (pacemaker, implanted cardioverter defibrillator, and cardiac resynchronisation therapy)	22.7%
Need oxygen	If the patient needs oxygen	4.7%
Barthel	Barthel scale	82.98 \pm 15.23
Gijón [25]	Sociofamily assessment scale in the elderly that allows the detection of risk situations or social problems.	7.47 \pm 2.29
<i>Laboratory</i>		
Urea	Urea (mg/dl)	75.12 \pm 37.8
Creatinine	Creatinine (mg/dl)	1.3 \pm 0.54
Sodium	Sodium (mEq/L)	140.12 \pm 4.14
Potassium	Potassium (g/dl)	4.28 \pm 0.74
Haemoglobin	Haemoglobin (g/dl)	13 \pm 9.6
<i>Comorbidities</i>		
Rhythm	If sinus rhythm, AF or atrial flutter	Sinus: 37.1%
Atrial fibrillation	If the patient has atrial fibrillation (AF)	57.4%
Pacemaker	If the patient has a pacemaker	14.5%

TABLE 3: Ambulatory patients monitored characteristics of the study population.

Characteristics	Description
SBP	Systolic blood pressure (mmHg)
DBP	Diastolic blood pressure (mmHg)
O2Sat	Oxygen saturation (%)
HR	Heart rate (bpm)
Weight	Body weight (kg)

4.1. Splitting Training and Testing Datasets. To build and test a predictive model, the clinical data are divided in training and testing datasets. The training dataset is used to develop the model, and once it is finished, the resulting model is tested with the testing dataset. This way, the overfitting is prevented, and it is possible to check whether the created model will generalize well. The whole dataset is from telemonitored patients starting from June of 2014 until February 2018. The training dataset contains 132 predictable decompensations (i.e., with at least 3 monitorizations in the last week before a decompensations) out of 174 patients, with an average follow-up per patient of 13.47 ± 7.47 months. The testing set contains 70 predictable decompensations out of 162 patients, with an average follow-up per patient of 5.41 ± 3.48 months.

4.2. Applied Alerts for Ambulatory Patients Admission. The alerts implemented in current medical practice are used as a filtering method to obtain the instances for training and building the classifiers. This way, we discard the days when

there is no sign of destabilization of any attribute, leading into a more balanced dataset. Therefore, this section presents the different types of alerts that are implemented in medical practice and their performance to select the optimal ones to be applied in our study.

4.2.1. Generic Alerts. The following tables describe the alerts that are being implemented in OSI Bilbao-Basurto Hospital and their sensitivity (Se) and false alerts per patient per year (FA/pt-y) when applied to the training dataset. They are differentiated into “yellow” and “red” alerts, being these last ones more restrictive and, therefore, more critical.

Simple Rules. Table 5 presents the rules based on each parameter individually. The alerts’ thresholds presented in Table 5 are the generic ones. But based on personalized clinical cases, clinicians modified some patients’ alerts thresholds. For example, if a patient’s O2Sat values are always lower than 90, but the patient is stable, the O2Sat alerts are adapted. This study uses the adapted alerts.

Weight Tendency. Besides simple rules, OSI Bilbao-Basurto Hospital also checks the tendency of weight values in order to trigger an alarm (Table 6). This weight change “red” (“yellow”) alert performs with a Se value of 0.52 (0.64) and a FA/pt-y of 9.55 (16.38).

Questionnaire. Additionally, OSI Bilbao-Basurto Hospital clinicians make use of the questionnaire (Table 4) and apply the following alert based on the answers from the questionnaire: if three or more answers are the wrong ones, the questionnaire alert would trigger. This alert achieves

TABLE 4: Ambulatory patients questionnaire.

n	Tag	Question	Possible answer
1	Well-being	Comparing with the previous 3 days, I feel:	B/W/S*
2	Medication	Is the medication affecting me well?	Yes/No
3	New medication	During the previous 3 days, did I take any medication without my clinicians' prescription?	Yes/No
4	Diet and exercise	Am I following the diet and exercise recommendations provided by my clinician and nurse?	Yes/No
5	Ankle	In the last 3 days, my ankles are:	B/W/S*
6	Walks	Can I go walking like previous days?	Yes/No
7	Shortness of breath	Do I have fatigue or shortness of breath when I lay down in the bed?	Yes/No
8	Mucus	Do I notice that I started coughing of with phlegm?	Yes/No

* B/W/S = better/worse/same.

TABLE 5: Simple rules implemented by Osakidetza.

Parameter to study	Threshold number	Type of alert	Se	FA/pt-y
SBP	<95	Yellow	0.28	11.4
	>150			
	<85	Red	0.08	1.4
	>180			
DBP	<60	Yellow	0.23	9.1
	>100			
	<50	Red	0.04	0.9
	>110			
HR	<55	Yellow	0.30	11.2
	>90			
	<50	Red	0.08	1.4
	>110			
O2Sat	<94	Yellow	0.15	3
	<90	Red	0.39	13.5

TABLE 6: Weight alerts implemented by Osakidetza.

Parameter to study	Time period	Minimum (kg)	Maximum (kg)	Type of alert
Weight change	5 days	1	2	Yellow
	3 days	1	25	Red
	5 days	2	25	Red

a Se of 0.31 and FA/pt-y of 9.55. To determine which are the questions that perform best, Table 7 presents the Se and FA/pt-y for each of them based on each possible answer.

The answers of "Worse" in the questions of n1 and n5 (Table 4) result in very good predictors of the de-compensations considering Se and FA/pt-y values. Questions n3, n4, n6, and n7 also have predictive power, but not as good as n1 and n5. The other questions cannot be considered as alerts, because of their low/null prediction capacity (Table 7).

4.2.2. Implemented Alerts Based on Moving Average. As presented in the *Related Work* section, weight-associated alerts have been improved, and hence, tendency rules for weight have been substituted for a more advanced method,

based on moving average. Moving Average Convergence Divergence (MACD) algorithm calculates the difference between the average value taken from two windows and generates an alert when this difference exceeds a prespecified threshold. Following the same moving average (MA) concept, a similar method is implemented which consists on the following (Figure 1):

- (i) *a*: immediate previous days (starting from the checking day and continuing backwards) over which the average is calculated
- (ii) *b*: previous days (starting from at least the latest day from *a* and continuing backwards) over which the average should be calculated
- (iii) *d*: distance between the last day of *a* and first day from *b*
- (iv) Difference threshold (THR): size of difference between *a* and *b* average that should generate an alert

In Figure 2, different scores for each possible variable's value for the MA alert are presented. The tested and illustrated results are from all possible combinations of the following variable's values: $a = (2, 3, 7)$, $b = (3, 4, 7, 14)$, $d = (0, 1, 3, 7)$, and $THR = (0.2, 0.5, 0.75, 1, 1.5, 1.8, 2, 3)$.

After representing all the results of the MA algorithm and applying the Youden index [26], the optimal value of these combinations is the one obtained with $a = 2$, $b = 3$, $d = 0$, and $THR = 0.75$ (green dot in Figure 2). This alert achieves Se value of 0.56 and FA/pt-y of 11.06 in the training set, similar to the results of the already alert-implemented weight alert. But based on the literature [6–8], this latest one is best.

4.3. Selection of Alerts for Instances Generation. To obtain the right dataset of instances, the best combination of alerts is sought. Once the alerts are selected, when at least one of these alerts is triggered, the patient data of that day are used to build the dataset for machine learning model building (see *Built Dataset for Machine Learning Classifiers*). In Table 8, the results of the combinations of different alerts are presented.

R1 refers to the sum of MA weight alert and the two best alerts from the questionnaire related to *ankle* (n5) and

TABLE 7: Questionnaire questions' performance.

n	Tag	Answer	Se	FA/pt-y
1	Well-being	Same	0.88	120.7
		Better	0.42	90.8
		Worse	0.37	2.7
2	Medication	Yes	1	210
		No	0.05	3.8
3	New medication	Yes	0.15	5.3
		No	1	209
4	Diet and exercise	Yes	1	203
		No	0.22	11.18
5	Ankle	Same	0.86	114
		Better	0.44	96
		Worse	0.35	2.9
6	Walks	Yes	0.99	196
		No	0.37	18
7	Shortness of breath	Yes	0.41	19.93
		No	0.96	194
8	Mucus	Yes	0.44	60.5
		No	0.84	153.5

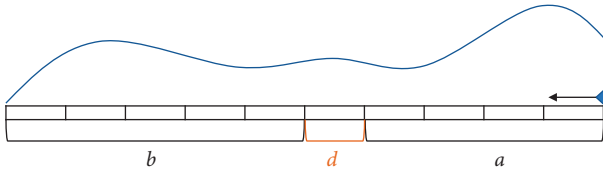


FIGURE 1: Representation of the applied MA algorithm.

well-being (n1) (Tables 4 and 7). If some of these alerts are triggered, R1 is also triggered. R2 refers to the R1 plus the yellow alerts of SBP, DBP, O2Sat, and HR (Table 5). Finally, R3 refers to R2 plus the questions n3, n4, n6, and n7 from Table 4. Since R2 (Table 8) detects almost all decompensations (95%), though FA/pt-y is quite high (FA/pt-y = 51.12), this is the one used to generate the instances for the machine-learning classifiers.

4.4. Built Dataset for Machine-Learning Classifiers. Next, the attributes that are considered for each of the instances and that are applied in the classifiers are presented. Note that the applied attributes come from (i) the telemonitoring dataset, (ii) the baseline dataset, and (iii) the readmission dataset.

(i) *Telemonitoring dataset:*

- (a) The value of SBP, DBP, HR, O2Sat attributes, and, in the case of the weight, the values of the MA algorithm
- (b) The number of consecutive alerts for each type of alert:
 - (1) Yellow alerts: the number of yellow alerts that have been triggered in the previous consecutive days related to SBP, DBP, HR, and O2Sat (4 attributes)
 - (2) Red alerts: the number of red alerts that have been triggered in the previous consecutive

days related to SBP, DBP, HR, and O2Sat (4 attributes)

- (3) MA: the number of alerts that have been triggered in the previous consecutive days for the MA algorithm (1 attribute)

(c) Questionnaires: the answers of the 8 questions of the questionnaire, shown in Table 7 (8 attributes)

(ii) *Baseline dataset:* the baseline information of the patient shown in Table 2 (24 attributes)

(iii) *Readmissions dataset:* whether in the moment of the instance is about a readmission, i.e., if the last 30 days, the patient has discharged because HF (1 attribute)

4.5. Applied Machine-Learning Classifiers. In this section, we briefly describe the main classification algorithms that were used during the experiments carried out in this work. Since classifier definitions are well known in the literature, we will provide just a summary overview about them.

4.5.1. Naïve Bayes. Naive Bayes methods follow the “naive” assumption that the components of the feature vectors are statistically independent, so that the posterior probability of the class can be approximated as

$$p(y | x) = \frac{p(y) \prod_i^n (p(x_i | y))}{p(x)}, \quad (1)$$

where $p(x_i)$ is the likelihood of the i -th feature, and $p(y)$ the a priori probability of the class. The Gaussian Naive Bayes assumes that the likelihood follows a Gaussian distribution, where the mean and standard deviation of each feature are estimated from the sample. On the other hand, the Bernoulli Naive Bayes assumes Bernoulli's distribution in the parameters, and hence, it estimates the probability of $p(x_i | y)$ following this last distribution.

4.5.2. Decision Tree. Decision Trees (DTs) [27, 28] are built by recursive partitioning of the data space using a quantitative criterion (e.g., mutual information, gain-ratio, gini index), maybe followed by a pruning process to reduce overfitting. Tree leaves correspond to the probabilistic assignment of data samples to classes. One of the most popular implementations of the algorithm is C4.5 [27], which is an extension of the previous ID3 [29] algorithm. At each node, the algorithm selects the feature that best splits the samples according to the normalized information gain.

4.5.3. Random Forest. Random forest [30] is an ensemble classifier consisting of multiple decision trees trained using randomly selected feature subspaces. This method builds multiple decision trees at the training phase. Often, a pruning process is applied to reduce both tree complexity and training data overfitting. In order to predict the class of a new instance, it is put down to each of these trees. Each tree

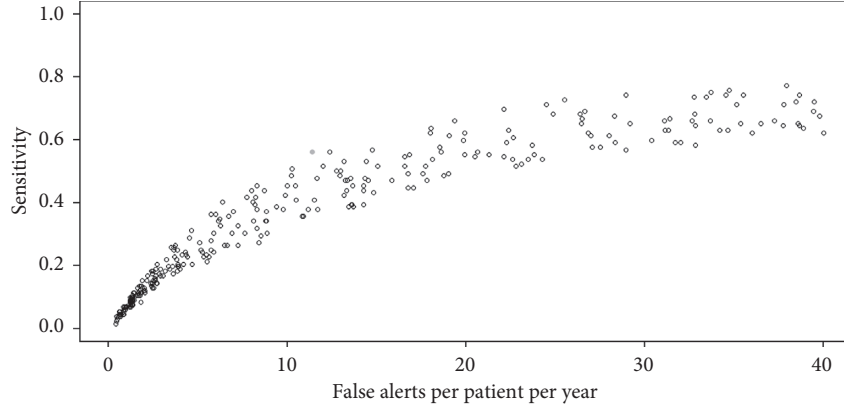


FIGURE 2: Representation of several MA to determine the Youden index.

TABLE 8: Inclusion Criterion performance.

Rules	Description of the rules	Se	FAPy
R1	Weight + ankle + well-being	0.79	15.33
R2	R1 + yellow	0.95	51.12
R3	R2 + questionnaire alerts	1	84.5

gives a prediction (votes) and the class having most votes over all the trees of the forest will be selected (majority voting). The algorithm uses the bagging method [31], where each tree is trained using a random subset (with replacement) of the original dataset. In addition, each split uses a random subset of features.

4.5.4. Support Vector Machine. Support vector machines (SVMs) [32, 33] look for the set of support vectors that allow to build the optimal discriminating surface in the sense of providing the greatest margin between classes. In this way, the decision function can be expressed in terms of the support vectors only:

$$f(x) = \text{sign}\left(\sum \alpha_i y_i K(s_i, x) + w_0\right), \quad (2)$$

where $K(x_i, x_j) \equiv \phi(x_i)T\phi(x_j)$ is a kernel function, α_i is a weight constant derived from the SVM process, and the s_i is the support vector [33]. Nonlinear kernel functions filling some conditions allow to map a nonlinearly separable discrimination problem into a linearly separable equivalent problem in higher dimensional space.

4.5.5. Neural Network. Multilayer Perceptron (MLP) is a neural network that consists of at least three layers of nodes, namely: (i) an input layer, (ii) one or more hidden layers, and (iii) an output layer. The input layer consists of a set of neurons that represents input features. The hidden layer transforms the outputs of the input layer by means of nonlinear activation functions. The output layer collects the values of the hidden layer and builds the output value. The model is trained using backpropagation, and it can classify data that is not linearly separable.

4.5.6. Class Balancing. In this work, like in many other supervised classification problems, imbalanced class distribution leads to important performance evaluation issues and problems to achieve desired results. The underlying problem with imbalanced datasets is that classification algorithms are often biased towards the majority class and hence, there is a higher misclassification rate of the minority class instances. Although there are several methods that can be used to tackle the class imbalance problem, we have followed an over-sampling approach. Random oversampling is the simplest oversampling method, which consists of randomly replicating minority class samples. Despite its simplicity, this method leads easily to overfitting, since it generates exact copies of existing instances [34]. In order to deal with such problems, we have used a more sophisticated technique, namely, synthetic minority oversampling technique (SMOTE). This method over samples the minority class by creating synthetic instances based on its nearest neighbours [35].

Depending on the percentage of synthetic samples that want to be generated (in respect to the original minority class instances), some, or all, minority samples are selected. Having specified beforehand the number of nearest neighbours k , for each sample, the k nearest neighbours are found using the Euclidean distance. Once the nearest samples are selected, a random value between 0 and 1 is generated and multiplied to the distance of each feature between the actual instance and the neighbour. In other words, the vector of coefficients of a random convex linear combination is generated and applied to the k nearest neighbours to create a new sample.

5. Results

This section presents the results obtained after the development of the machine learning classifiers presented in *Applied Machine-Learning Classifiers* and the final results of the selected classifier in the testing dataset.

5.1. Validation Method. Although there are many ways to assess the generalization ability of a ML model, such as cross-validation, time series can be problematic for such validation techniques, as there is an ordered timeline factor

to be considered. Henceforth, we use cross-validation on a rolling basis [36], as it is explained in Figure 3.

The training set is separated in the five sets shown in Figure 3. The number written inside the blocks is the number of decompensations corresponding to that period, which is the reason why the dates (on top) are chosen. The splits are not exactly equitable, since all the predecessors of a decompensation must fit within the same block. In Step 1, the classifier is trained in the first block (55 decompensations) and tested in the next block (17 decompensations) getting the score for Step 1. Following, in Step 2, the classifier is trained in Step 1 and tested in the new one (19 decompensations), getting the score for Step 2. Repeating the same with Step 3 and Step 4, we get four scores. It is supposed that the first step is the more unstable, as there are less data to train the classifiers, but, while the training set increases, it is believed that the results will become stable, and the score will converge to its real testing value.

The score value used to test the classifiers is the area under the ROC curve (AUC) [37], a measure that evaluates the balance between sensitivity and specificity and that gets an accurate estimation even in moderately imbalanced datasets, which is our case. The AUC value is used to check how well the classifiers perform and consequently select the best one. To test the global predictive model, we use Se and FA/pt-y which are the ones used in the literature.

5.2. Classifiers Comparison. In this section, the results of the classifiers explained in *Applied Machine-Learning Classifiers* are presented applied for the training dataset. Additionally, the rolling cross-validation method, presented above, is applied to avoid the overfitting. This way, the classifier(s) with best outcomes and generalizable (and therefore, stable) can be selected for the predictive model.

In Figure 4, the AUC values of each classifier are illustrated for each of the steps defined in the rolling cross-validation method. The points are the mean of the AUCs achieved in each case, with its standard deviation drawn with whiskers. High standard deviation value indicates that the classifier is less generalizable, while low standard deviation hints a stable classifier.

It is expected that the AUCs values converge as the number of steps grow, although with the available dataset, there are a trend of significative improvement in the second step and a worsening trend in the third one. However, Figure 4 clearly shows that the best classifiers are Naïve Bayes (NB) with Bernoulli method and the random forest (RF). NB classifier has lower AUC value than RF, but the standard derivation is almost negligible, and the trend through the steps is more stable. Hence, it is expected that its performance will not vary significantly over time with new data. RF gets the best scores, but is unstable, and it has high standard derivations. Henceforth, NB with Bernoulli method and RF classifiers are selected to validate the models.

Decision tree and SGD classifiers give the lowest results. The other three classifiers (NB with Gaussian distribution, SVM, and MLP classifiers) perform better, but not as good as the selected two.

5.3. Final Results

5.3.1. Alerts Performance. Since the alerts are used to generate the instances for the machine-learning classifiers (see *Selection of Alerts for Instances Generation*), first, the performance of these in the testing dataset is presented (Table 9).

Comparing these results (Table 9) with the obtained in the training set (Table 5), the weight-associated alerts get worse result. In the case of the questionnaire alerts (Table 10), there is a general worsening comparing with the training set (Table 7). Hence, it is possible to get worse results than the expected when testing the predictive model in the testing set.

5.3.2. Validation Results. In the current medical practice, their alerts all together obtain the following results: Se = 0.76 and FA/pt-y = 28.64 with the red alerts plus the questionnaire alert, and Se = 1 and FA/pt-y = 88.41 with the yellow alerts plus the questionnaire alert.

After applying the R2 alerts in the testing dataset (see *Selection of Alerts for Instances Generation* section), the selected machine learning classifiers achieved the following AUC values: NB with Bernoulli, AUC = 0.67 and RF, AUC = 0.62. As it was expected, NB with Bernoulli maintains the AUC value in accordance with the results obtained in the training dataset, and in the case of RF, due to the classifier instability, the score deteriorates (Figure 4).

Once the classifier is selected and trained, the results are given depending on the probability of the patient to suffer a decompensation. For that, the probability given by the classifier (0 if none of the alerts of the inclusion criterion is triggered) is split in terciles. Each tercile is associated with a colour: if the probability is less than 0.33, “green” group; if the probability is between 0.33 and 0.66, “yellow” group; and if it is upper than 0.66, “red” group. This way, the clinicians can base their decisions on the risk group. Setting the probabilities of the classifiers to the risk groups, the results achieved are the next (Table 11).

As presented in Table 11, the RF classifier results in a poor predictive model. However, NB reaches acceptable scores comparing with literature studies that have similar attributes (see section *Alerts for HF Patients*). Comparing the obtained results in the “red” group to the current medical practice, though NB (in the red group) gets 38% less of Se (0.76 \rightarrow 0.47) value, it achieves 72% less of FA/pt-y (28.64 \rightarrow 7.8) value. Henceforth, this predictive model improves the results of the actual alerts method and it is more reliable.

6. Discussion

Current medical practice may use sensitive alerts, that although they detect most of the decompensations due to their high sensitivity, they also have too many false alerts. Therefore, the main goal of this study is the reduction of these false alerts. This study has shown an improvement from current alerts system implemented in the hospital. The system reduces the number of false alerts notably, from 28.64 FA/pt-y of the current medical practice to even 7.8 FA/pt-y

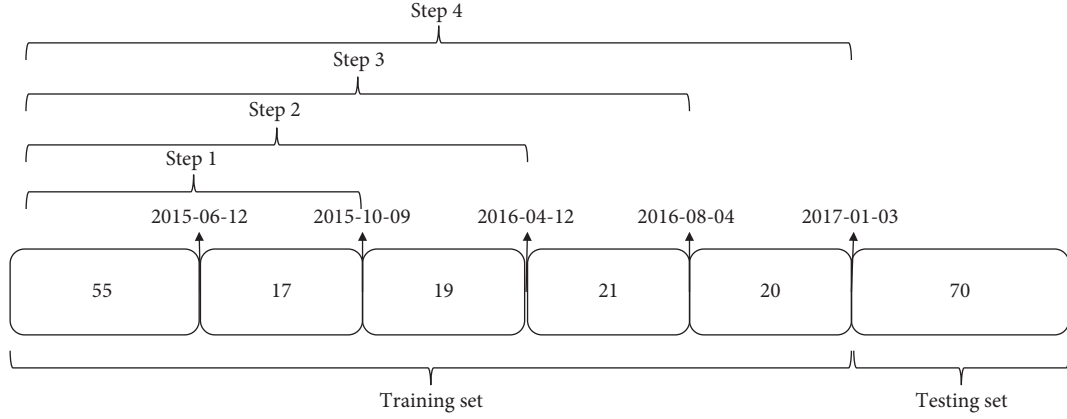


FIGURE 3: Cross-validation on a rolling basis applied in the study.

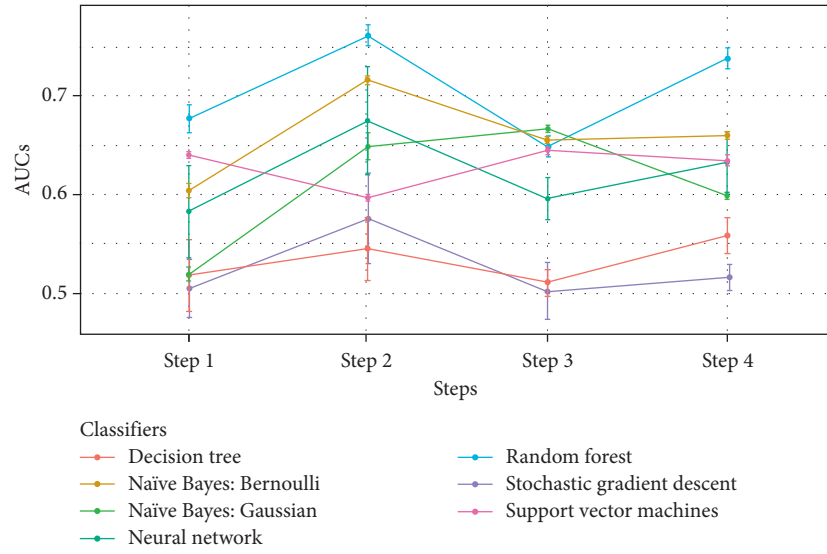
FIGURE 4: AUC values of the classifiers (colours) depending on the steps (axis x).

TABLE 9: Alerts' performance in the testing set.

Alert tag	Colour alert	Se	FA/pt-y
Weight	—	0.4	13.36
SBP	Yellow	0.49	18.9
	Red	0.1	2.44
DBP	Yellow	0.34	15.4
	Red	0.07	1.3
HR	Yellow	0.37	19.7
	Red	0.06	3.1
O2Sat	Yellow	0.5	27.27
	Red	0.2	4.2

TABLE 10: Questionnaire alerts' performance in the testing set.

Question tag	Answer	Se	FA/pt-y
Well-being	Worse	0.25	2.9
New Medication	Yes	0.13	5.2
Diet and exercise	No	0.16	8.75
Ankle	Worse	0.13	2.86
Walks	No	0.4	24.7
Shortness of breath	Yes	0.43	21.9

TABLE 11: Results of the predictive models.

Group	Random forest		Naïve Bayes	
	Se	FA/pt-y	Se	FA/pt-y
Green	1	79	0.75	59.4
Yellow	0.08	1.29	0.41	13.2
Red	0	0.11	0.47	7.8

for the “red” group, which is denoted as the most restrictive group. This last result is achieved with the predictive model built by applying NB with Bernoulli to the combination of telemonitoring alerts and questionnaire alerts (R2). However, as expected, the application of machine learning techniques entails a decrement on sensitivity values. The result obtained in this study for the “red” group is $Se = 0.47$, while the alerts used in the current medical practice applied to the same testing dataset achieve $Se = 0.76$. Despite this Se worsening, it is notorious that the FA/pt-y has much higher decrement, with which we conclude that this new predictive model improves the current medical practice. Moreover, when comparing the obtained results with the state of the

art, the Se values are similar or better to these studies that do not consider transthoracic impedance (Table 1). Especially considering that in the SoA, most of the studies reduce the real FA/pt-y concatenating the neighbour alerts, since they assume that once an alert has been triggered, the clinician will take action, and hence, next consecutive alerts will not be triggered.

The current study also presents some limitations. Firstly, as presented in *Characteristics of Ambulatory Patients Dataset*, there are patients that did not monitor regularly. As a consequence, from 254 decompensations during this telemonitored period, only 202 of them had 3 or more measurements during the last week previous to the admission, and hence, could be used in our study. The rest did not have even 3 measurements, and hence, they were not predictable.

Secondly, as the clinical data used in the study are from Caucasian patients, the model may perform differently in different settings, such as in non-Caucasian population. Finally, we must stress that heart failure is a very complex disease with multiple factors, and its predictiveness is complex. Nevertheless, larger amount of data and the registration of all type of decompensations is key to improve the current model.

7. Conclusion and Future Work

This article presents the methodology to develop predictive models for HF decompensations prediction based on ambulatory patients' telemonitored data, extending the study for readmissions detection [9].

The results on these studies have been successfully implemented in a telemedicine system, called INCAR. This way INCAR provides the patient with the confidence of being monitored and guided with an advanced technology and clinical professionals' supervision.

Currently, new devices that monitor physical activity and sleeping quality are incorporated in the telemonitoring program in order to determine whether these features could have an impact in the results and improve the outcome. To finish, we will study the possibility of including in the telemonitoring plan a new device that monitors transthoracic impedance, and explore raising deep learning techniques, which have demonstrated their good performance and may improve the presented results.

Data Availability

The dataset used for this study contains personal information, and therefore, following the Research Data policy of Hindawi, it is not available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been funded by Basque Government by means of the RIS3 Program, under reference No 2017222015.

References

- [1] O. Hamdi, M. A. Chalouf, D. Ouattara, and F. Krief, "eHealth: survey on research projects, comparative study of tele-monitoring architectures and main issues," *Journal of Network and Computer Applications*, vol. 46, pp. 100–112, 2014.
- [2] S. I. Chaudhry, J. A. Mattera, J. P. Curtis et al., "Tele-monitoring in patients with heart failure," *New England Journal of Medicine*, vol. 363, no. 24, pp. 2301–2309, 2010.
- [3] S. Mangini, P. V. Pires, F. G. M. Braga, and F. Bacal, "Decompensated heart failure," *Einstein*, vol. 11, no. 3, pp. 383–391, 2013.
- [4] P. Ponikowski, A. A. Voors, S. D. Anker et al., "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC," *European Heart Journal*, vol. 37, no. 27, pp. 2129–2200, 2016.
- [5] W. T. Abraham, S. Compton, G. Haas et al., "Intrathoracic impedance vs daily weight monitoring for predicting worsening heart failure events: results of the Fluid Accumulation Status Trial (FAST)," *Congestive Heart Failure*, vol. 17, no. 2, pp. 51–55, 2011.
- [6] I. C. Gyllenstein, A. G. Bonomi, K. M. Goode et al., "Early indication of decompensated heart failure in patients on home-telemonitoring: a comparison of prediction algorithms based on daily weight and noninvasive transthoracic bio-impedance," *JMIR Medical Informatics*, vol. 4, no. 1, p. e3, 2016.
- [7] J. Zhang, K. M. Goode, P. E. Cuddihy, J. G. F. Cleland, and TEN-HMS Investigators, "Predicting hospitalization due to worsening heart failure using daily weight measurement: analysis of the Trans-European Network-Home-Care Management System (TEN-HMS) study," *European Journal of Heart Failure*, vol. 11, no. 4, pp. 420–427, 2009.
- [8] M. T. Ledwidge, R. O'Hanlon, L. Lalor et al., "Can individualized weight monitoring using the HeartPhone algorithm improve sensitivity for clinical deterioration of heart failure?," *European Journal of Heart Failure*, vol. 15, no. 4, pp. 447–455, 2013.
- [9] J. Kerexeta, A. Artetxe, V. Escolar, A. Lozano, and N. Larburu, "Predicting 30-day readmission in heart failure using machine learning techniques," in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, Funchal, Portugal, January 2018.
- [10] P. B. Adamson, M. R. Zile, Y. K. Cho et al., "Hemodynamic factors associated with acute decompensated heart failure: part 2—use in automated detection," *Journal of Cardiac Failure*, vol. 17, no. 5, pp. 366–373, 2011.
- [11] F. R. Gilliam III, G. A. Ewald, and R. J. Sweeney, "Feasibility of automated heart failure decompensation detection using remote patient monitoring: results from the decompensation detection study," *Innovations in Cardiac Rhythm Management*, vol. 3, pp. 735–745, 2012.
- [12] G. García-Sáez, M. Rigla, I. Martínez-Sarriegui et al., "Patient-oriented computerized clinical guidelines for mobile decision support in gestational diabetes," *Journal of Diabetes Science and Technology*, vol. 8, no. 2, pp. 238–246, 2014.
- [13] M. Peleg, Y. Shahar, S. Quaglini et al., "Assessment of a personalized and distributed patient guidance system," *International Journal of Medical Informatics*, vol. 101, pp. 108–130, 2017.
- [14] B. J. Mortazavi, N. S. Downing, E. M. Bucholz et al., "Analysis of machine learning techniques for heart failure

- readmissions,” *Circulation: Cardiovascular Quality and Outcomes*, vol. 9, no. 6, pp. 629–640, 2016.
- [15] K. Zolfaghar, “Predicting risk-of-readmission for congestive heart failure patients: a multi-layer approach,” 2013, <http://arxiv.org/abs/1306.2094>.
 - [16] B. Zheng, J. Zhang, S. W. Yoon, S. S. Lam, M. Khasawneh, and S. Poranki, “Predictive modeling of hospital readmissions using metaheuristics and data mining,” *Expert Systems with Applications*, vol. 42, no. 20, pp. 7110–7120, 2015.
 - [17] N. Meadem, N. Verbiest, K. Zolfaghar et al., “Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients,” in *Proceedings of International Conference on Knowledge Discovery and Data Mining (KDD), Data Mining and Healthcare (DMH)*, vol. 150, Chicago, IL, USA, August 2013.
 - [18] H. M. Krumholz, Y.-T. Chen, Y. Wang, V. Vaccarino, M. J. Radford, and R. I. Horwitz, “Predictors of readmission among elderly survivors of admission with heart failure,” *American Heart Journal*, vol. 139, no. 1, pp. 72–77, 2000.
 - [19] R. Amarasingham, B. J. Moore, Y. P. Tabak et al., “An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data,” *Medical Care*, vol. 48, no. 11, pp. 981–988, 2010.
 - [20] S. Sudhakar, W. Zhang, Y.-F. Kuo, M. Alghrouz, A. Barbajelata, and G. Sharma, “Validation of the readmission risk score in heart failure patients at a tertiary hospital,” *Journal of Cardiac Failure*, vol. 21, no. 11, pp. 885–891, 2015.
 - [21] M. van der Heijden, B. Lijnse, P. J. F. Lucas, Y. F. Heijdra, and T. R. J. Schermer, “Managing COPD exacerbations with telemedicine,” in *Artificial Intelligence in Medicine*, pp. 169–178, Westview Press, Boulder, CO, USA, 2011.
 - [22] M. S. Mohktar, *A Decision Support System for the Home Management of Patients with Chronic Obstructive Pulmonary Disease (COPD) Using Telehealth*, Graduate School of Biomedical Engineering, University of New South Wales, Kensington, NSW, Australia, 2012.
 - [23] M. Velikova, P. J. F. Lucas, and M. Spaanderman, “A predictive bayesian network model for home management of preeclampsia,” in *Artificial Intelligence in Medicine*, pp. 179–183, Westview Press, Boulder, CO, USA, 2011.
 - [24] R. Lafta, J. Zhang, X. Tao et al., “An intelligent recommender system based on predictive analysis in telehealthcare environment,” *Web Intelligence*, vol. 14, no. 4, pp. 325–336, 2017.
 - [25] M. T. Alarcón and J. I. González-Montalvo, “La escala sociofamiliar de Gijón, instrumento útil en el hospital general,” *Revista Española de Geriatria y Gerontología*, vol. 33, no. 1, pp. 178–179, 1998.
 - [26] W. J. Youden, “Index for rating diagnostic tests,” *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
 - [27] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, Burlington, MA, USA, 1993.
 - [28] L. Breiman, R. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees 1984*, Wadsworth and Brooks, Monterey, CA, USA, 1984.
 - [29] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
 - [30] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [31] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
 - [32] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
 - [33] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
 - [34] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, “An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics,” *Information Sciences*, vol. 250, pp. 113–141, 2013.
 - [35] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
 - [36] C. Bergmeir and J. M. Benítez, “On the use of cross-validation for time series predictor evaluation,” *Information Sciences*, vol. 191, pp. 192–213, 2012.
 - [37] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.