

## Review Article

# Validation Techniques for Sensor Data in Mobile Health Applications

Ivan Miguel Pires,<sup>1,2,3</sup> Nuno M. Garcia,<sup>1,3,4</sup> Nuno Pombo,<sup>1,3</sup>  
Francisco Flórez-Revuelta,<sup>5</sup> and Natalia Díaz Rodríguez<sup>6</sup>

<sup>1</sup>*Instituto de Telecomunicações, Universidade of Beira Interior, Covilhã, Portugal*

<sup>2</sup>*Altranportugal, Lisbon, Portugal*

<sup>3</sup>*Assisted Living Computing and Telecommunications Laboratory (ALLab), Computer Science Department, Universidade of Beira Interior, Covilhã, Portugal*

<sup>4</sup>*ECATI, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal*

<sup>5</sup>*Department of Computer Technology, Universidad de Alicante, Alicante, Spain*

<sup>6</sup>*Department of Computer Science and Artificial Intelligence, CITIC-UGR, University of Granada, Granada, Spain*

Correspondence should be addressed to Ivan Miguel Pires; [impieres@it.ubi.pt](mailto:impieres@it.ubi.pt)

Received 27 February 2016; Revised 25 August 2016; Accepted 4 September 2016

Academic Editor: Francesco Dell'Olio

Copyright © 2016 Ivan Miguel Pires 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.

Mobile applications have become a must in every user's smart device, and many of these applications make use of the device sensors' to achieve its goal. Nevertheless, it remains fairly unknown to the user to which extent the data the applications use can be relied upon and, therefore, to which extent the output of a given application is trustworthy or not. To help developers and researchers and to provide a common ground of data validation algorithms and techniques, this paper presents a review of the most commonly used data validation algorithms, along with its usage scenarios, and proposes a classification for these algorithms. This paper also discusses the process of achieving statistical significance and trust for the desired output.

## 1. Introduction

There has been an increase of the number of mobile applications that make use of sensors to achieve a plethora of goals. Many of these applications are designed and developed by amateur programmers, and that in itself is good as it confirms an increase in the overall set of skills of the developer community. Nevertheless, and even when the applications are developed by professionals or by companies, there are not many applications that publicize or disclose how the sensors' data is processed. This is a problem, in particular when these applications are meant to be used in a scenario where they can influence their users' lives, as for example, when the data is expected to be used to identify Activities of Daily Living (ADLs) or, to an extreme, when the applications are used in medical scenarios.

Due to the nature of the mobile device itself, multi-processing, with limited computational power and limited

battery life, the data that is collected from the sensors is often unusable in its primary form, requiring further processing to allow it to be representative of the event or object that it is supposed to measure. The recording of sensor data and the sequent processing of this data need to include validation subtasks that guarantee that the data are suitable to be fed into the higher-level algorithms.

Moreover, the use of the sensors' data to feed higher-level algorithms needs to guarantee a minimum degree of error, with this error being the difference between the output of these applications, built on limited computational mobile platforms, and the output of a golden standard. To achieve a minimum degree of error, statistical methods need to be applied to ensure that the output of the mobile application is to maximum extent similar to the output given by the relevant golden standard, if and when this is possible.

To mitigate this problem, this paper presents and discusses the most used data validation algorithms and

techniques and their usage in a mobile application that relies on the sensors' data to give meaningful output to its user. The algorithms are listed and their use is discussed. The discussion of the statistical process to ensure maximum reliability of the results is also presented.

The remainder of this paper is organized as follows: this paragraph concludes Section 1, where a short introduction to the problem and a proposal to achieve its mitigation are disclosed; Section 2 presents the most commonly found data validation methods, along with a critical comparison of its usage scenarios; Section 3 deepens the analysis presenting a classification of the data validation methods; Section 4 discusses the applicability of these methods, including the discussion of the degree of trust the data can be expected to provide; finally, Section 5 presents relevant conclusions.

## 2. Data Validation Methods

Sensor data validation is an important process executed during the data acquisition and data processing modules of the multisensor mobile system. This process consists of the validation of the external conditions of the data and the validity of the data for specific purpose, in order to obtain accurate and reliable results. The sequence of this validation may be applied not only in data acquisition but also in data processing since increase, as these increase the degree of confidence of the systems, with the confidence in the output being of great importance, especially for systems involved in medical diagnosis, but also for the identification of ADLs or sports monitoring.

In addition, data validation methods must be used during the different phases of the conception of a new system, such as design, development, tests, and validation. Therefore, the data validation methods with verified reliability during the conception should be also used to validate the data automatically during the execution time.

One of the causes for the presence of incorrect values during the data acquisition process may be existence of environmental noise. Even when the data is correctly collected, the data may still be incorrect because of noise. Therefore, very often the data captured or processed has to be cleaned, treated, or imputed to obtain better and reliable results. Following the existence of missing values at random instants of time, the causes may be the mechanical problems or power failures of sensors. At this case, data correction methods should be applied, including data imputation and data cleaning. The data validation process may be simplified as presented in Figure 1.

The selection of the best technique for sensor data validation also depends on the type of data collected, the purpose of its application, and the computational platform where the algorithm will be run. Data validation techniques are commonly composed by statistical methods. Due to the characteristics of mobile devices, data validation techniques can be executed locally in the mobile device or at the server-side, depending on the amount of data to validate simultaneously, the frequency of the validation tasks, and the computational, communication, and storage resources needed for the validation. The characteristics of the sensors

are also important for the selection of the best techniques, which may be separated in three large groups, which are sensor performance characteristics, pervasive metrics, and environmental characteristics [1].

While data validation is important for improving the reliability of a system, it also depends on other factors, such as power instability, temperature changes, out-of-range data, internal and external noises, and synchronization problems that occur when multiple sensors are integrated into a system [2]. However, the reconstruction of the data and correction for the correct measurement is also important, and several research studies have proposed systems, methods, models, and frameworks to improve the data validation and reconstruction [3, 4].

Sensor data validation methods can be separated in three large groups, such as faulty data detection methods, data correction methods, and other assisting techniques or tools [5].

Firstly, faulty data detection methods may be either simple test based methods or physical or mathematical model based methods, and they are classified in valid data and invalid or missingness data [6, 7]. For the detection of faulty data, the authors in [7] presented an order of methods that should be applied to obtain better results, which are as follows: zero value detection, flat line detection, minimum and maximum values detection, minimum and maximum thresholds based on last values, statistical tests that follow certain distributions, multivariate statistical tests, artificial neural networks (ANNs) [8], one-class support vector machine (SVM) [9], and classification and physical models. On the one hand, simple test based methods include different techniques, such as physical range check, local realistic range detection, detection of gaps in the data, constant value detection, the signals' gradient test, the tolerance band method, and the material redundancy detection [7, 10, 11]. On the other hand, physical or mathematical model based methods include extreme value check using statistics, drift detection by exponentially weighted moving averages, the spatial consistency method, the analytical redundancy method, gross error detection, the multivariate statistical test using Principal Component Analysis (PCA), and data mining methods [7, 12, 13].

Secondly, data correction methods can be carried out by interpolation, smoothing, data mining, and data reconciliation [10, 12, 14]. For the application of the interpolation, the authors of [11] proposed the use of the value measured from the last measurement or the use of the trend from previous sets of measurements. The smoothing methods, for example, moving average and median, may be used to filter out the random noise and convert the data into a smooth curve that is relatively unbiased by outliers [10]. The application of data mining techniques allows the replacement of the faulty values by the measurements performed with several methods, for example, ANNs [14]. The data reconciliation methods, for example, PCA, are used for the calculation of a minimal correction to the measured variables, according to the several constraints of the model [13].

Thirdly, the other assisting techniques or tools are, namely, the checking of the status of the sensors, the checking

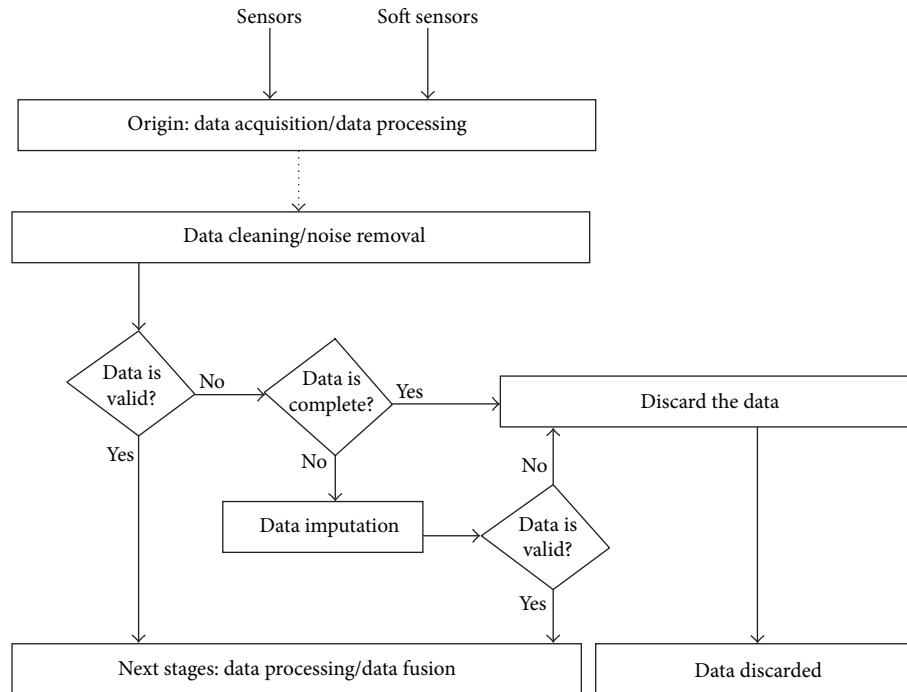


FIGURE 1: Sequence of activities performed during the data validation process.

of the duration after sensor maintenance, data context classification, the calibration of measuring systems, and the uncertainty consideration [6, 7, 10].

Several research studies have been performed, using data validation techniques. In [15], PCA is used for the compression of linearly correlated data. The authors compared the Auto-Associative Neural Network (AANN) and the Kernel PCA (KPCA) methods for data validation, creating a new approach named as Hybrid AANN-KPCA that uses these two methods. When compared with AANN and KPCA methods, the Hybrid AANN-KPCA achieves better performance results for the prediction or correction of inconsistent data.

In [16], the authors proposed that the data validation may be performed with Kalman filtering and linear predictive coding (LPC), showing that the results using Kalman filtering are better than LPC using several types of data, but the LPC reported a smaller energy consumption.

Several studies proposed the use of ANNs, for example, the Multilayer Perceptron (MLP), that can be trained to perform the identification of faulty sensors using prototype data and used to determine the near optimal subset of sensor data to produce the best results [2, 17–19]. Besides, the sensor data validation may be performed with other probabilistic methods, such as Bayesian Networks, Propagation in Trees, Probabilistic Causal Methods, and Learning Algorithms [20]. The authors of [20] proposed the anytime sensor validation algorithms that combine several probabilistic methods. On the contrary, [21] proposed the validation of data using the Sparse Bayesian Learning and the Relevance Vector Machine (RVM), which are a specialization of SVM.

For the estimation of the values during data validation, the authors of [22] analysed the use of the Kalman filter,

which was implemented in two methods: Algorithmic Sensor Validation (ASV), and Heuristic Sensor Validation (HSV). The ASV method implements different statistical methods, for example, mean, standard deviation, and sensor confidence that represent the uncertain nature of sensors. HSV identifies faulty sensor readings as attributable to a sensor or system failure. As an example, the authors of [23] proposed the use of the Kalman filter for the validation of the GPS data.

Other used methods are the grey models, which consists of differential equations describing the behaviour of an accumulated generating operation (AGO) data sequence. As an example, [4] presented a novel self-validating strategy using grey bootstrap method (GBM) for data validation and dynamic uncertainty estimation of self-validating sensor. The GBM can evaluate the measurement uncertainty due to poor information and small sample.

In [2], the Autoregressive Moving Averages (ARMA) transform the process for determining the validity of the acquired data, evaluating the levels of noise and providing a timely warning from the expected signals. The model created for ARMA includes linear regression techniques to predict the invalid values with Autoregressive (AR) and Moving Average (MA) models. Sensor Data Validation in Aeroengine Vibration Tests also implements the Autoregressive (AR) Model, complemented with the Empirical Mode Decomposition (EMD) [24]. Another method presented is the sensor validation and fusion of the Nadaraya-Watson statistical estimator [25], using a Fuzzy Logic model [26]. These methods and others, including the use of Gaussian distributions and error detection methods, may be also used to improve the quality of the measurements [27, 28].

Intelligent sensor systems are able to perform the capture and validation of the sensors' data. Staroswiecki [29] argues that the data validation is important to increase the confidence level of these systems, proposing two types of validation, such as technological and functional. Technological validation consists on the analysis of the conditions of the hardware resources of the sensors, but it does not guarantee that the estimation produced by the sensor is correct, but only that the operating conditions were not against possible correctness. On the contrary, functional validation consists of Fault Detection and Isolation (FDI) procedures, which consists of the use of algorithms to complement the Technological Validation. The authors of [30] also agreed with Staroswiecki in the separation of the data validation in two types, presenting a real time algorithm based on probabilistic methods. Other studies have been researched and developed, including the data validation techniques using intelligent sensor systems [31].

Another powerful technique for data validation consists of the use of self-validating (SEVA) sensors, which provide an estimation of the error bounds during the measurements [32]. SEVA are widely researched in literature. An example, using a Back-Propagation (BP) model, is applied into a system to obtain an estimated value and then a fault detection method called SPRT (sequential probability ratio test), identifying the validity of the system [33]. For the use of SEVA technologies, the authors of [34] also proposed the validated random fuzzy variable (VRFV) based uncertainty evaluation strategy for the online validated uncertainty (VU) estimation. In [35], the authors presented a novel strategy of using polynomial predictive filters coupled with VRFV which is proposed for the online measurements validation and validated uncertainty estimation of multifunctional self-validating sensors. These authors also performed a research about the use of some fuzzy logic rules, comparing the predicted values with the actual measurements to obtain the confidence evaluation [36]. In [37], the authors proposed an approach of sensor data validation using self-reporting, including the measurement based on the data quality, that is, validating the data loss measured by periodic sensors, the timing of data collection, and the accuracy of the detection of changes. ANNs may be used for SEVA with self-organizing maps (SOM) [38], which are trained using unsupervised learning techniques to produce a low-dimensional, discretized representation of the input space of the training samples [39].

The use of valid data is important for the developments of intelligent sensor systems, which may be used for health purposes and, consequently, for the detection of the ADLs [40–45]. The use of mobile devices allows the data acquisition anywhere and at anytime, but these devices have several constraints, such as low memory, processing power, and battery resources, but data validation may help for increasing of the performance of the measurements, reducing the resources needed [46–48]. In general, these systems use probabilistic methods to detect the failures at real-time to obtain better results.

Table 1 presents a summary of the data validation methods included on each category. The methods that are mainly implemented use statistical and artificial intelligence

techniques, such as PCA, RVM, ANNs, and others, increasing the reliability of the data acquisition and data processing algorithms. In spite of the SVM and the ANN working in a slightly different manner, their foundations are quite similar. In fact the SVM without kernel is a single neural network neuron with a different cost function. Congruently, when the SVM has a kernel it is comparable with a 2-layer ANN.

Following the methods presented at Table 1, the most studied scenarios for data validation are mainly related to health sciences, laboratory experiments, and other undifferentiated tasks. However, only a minor part of studies is related to the use of mobile devices, smart sensors, and other devices used daily. Besides, the development of healthcare solutions based on the sensors available on the mobile devices increases the requirement of the validation of the data collected by the sensors available on the mobile devices. Depending on the types of the data, for some complex data acquired, such as images, videos, GPS signal, and other complex types of data, the validation of the data should be accomplished by other auxiliary systems working at the same time, validating the data at the server-side, but a constant network connection must be available. Other topologies of systems may be susceptible for the implementation of data validation techniques. The Wireless Sensor Networks (WSN) are an example of systems where the different nodes of the network may perform the validation of the data collected for the neighbourhood nodes, and these nodes may be composed of different types of sensors. However, the main topology for the implementation with mobile devices is the self-validation using only the sensors and the data available on the mobile device.

The data validation may be executed automatically and transparently for the mobile devices' user and, commonly, at least one of the methods for each stage should be implemented in a system to perform the validation of the sensors' data. Firstly, for faulty data detection methods, the ANNs are the most used methods for the training of the data and for the detection of the inconsistent values. Secondly, for data correction methods, the most used method is the Kalman filter. Thirdly, the other assisting techniques that are commonly applied are the data context classification, the checking of the status of sensors, and the uncertainty considerations. Applying the data validation techniques correctly, the reliability and acceptability of the systems may be increased.

### 3. Classification of Data Validation Methods

Data validation methods may be classified in three large groups [5] as follows: faulty data detection methods, data correction methods, and other assisting techniques or tools.

The faulty data detection methods and the data correction methods may be executed sequentially in a multisensor system in order to obtain the results based on valid data. The other assisting techniques or tools mainly consist of the validation of the working state of the sensors, and this validation may be executed at the same time of the execution of faulty data detection and data correction methods, because these types of failures invalidated the results of the algorithms.

TABLE 1: Classification of the data validation methods by functionality.

Groups of data validation methods	Methods included	Description
Faulty data detection methods	ANNs (i) MLP; AANN; BP algorithm; SVM;	Consisting of the detection of faulty or incorrect values discovered during the data acquisition and processing stages
	Instance based (i) SOM Gaussian distributions Statistical methods (i) ASV; HSV Probabilistic methods (i) Bayesian Networks; Propagation in Trees; Probabilistic Causal Methods; Learning Algorithms; Sparse Bayesian Learning; RVM; SPRT Dimensionality Reduction (i) Fuzzy logic; PCA; KPCA; others (i) Hybrid AANN-KPCA	
Data correction methods	Kalman filter LPC ARMA (i) AR; MA; EMD Nadaraya-Watson statistical estimator Interpolation Smoothing Data mining techniques Data reconciliation techniques	Consisting of the estimation of faulty or incorrect values obtained during the data acquisition and processing stages
Other assisting techniques or tools	Checking of the status of the sensors Checking of the duration after sensor maintenance Data context classification Calibration of measuring systems Uncertainty consideration Grey models (i) GBM; dynamic uncertainty estimation of self-validating sensor VRFV method	These are different approaches created for the correct validation of the data

These different approaches are based on either mathematical methods, for example, statistical or probabilistic methods, or complex analysis, for example, artificial intelligence methods. According to [49], the data validation methods may be classified in several types of methods, which are presented in Figure 2.

As depicted in Figure 2, the faulty data detection methods, used to detect failures on the sensors' signal, may include ANNs, dimensional reduction methods, instance based methods, probabilistic and statistical methods, and Bayesian methods. On the contrary, the data correction methods include the following methods: filtering, regression, estimation, interpolation, smoothing, data mining, and data reconciliation. These methods work specifically with the sensors' data and the selection of the methods that can be applied by a system should consider the system's usage scenarios.

Finally, the other assisting techniques or tools are mainly related to detection of problems originated by either hardware components or its working environment. In addition, on real-time systems, these problems should be verified constantly to prevent the existence of failures in the data captured.

#### 4. Applicability of the Sensor Data Validation Methods

Mobile devices have a plethora of sensors available for the measurement of several parameters, including the identification of the ADLs. Examples of these sensors include the accelerometer, the gyroscope, the magnetometer, the GPS, and the microphone.

The data acquisition using accelerometers may fail because of several problems, including problems related with the internal electronic amplifier of the Integrated Electronic Piezoelectric (IEPE) device, the exposure to temperatures beyond the accelerometer working range, failure related with electrical components, capture of environmental noise, the multitasking and multithreading capabilities of the mobile devices that may cause irregular sampling rates, the positioning of the accelerometer, the low processing and memory power, and the battery consumption [50]. The causes of failure of an accelerometer are similar to the causes of the failure of a gyroscope, a magnetometer, or a microphone [51]. In addition, the GPS has another failure cause, which consists of the low connectivity of satellites in indoor environments [52].

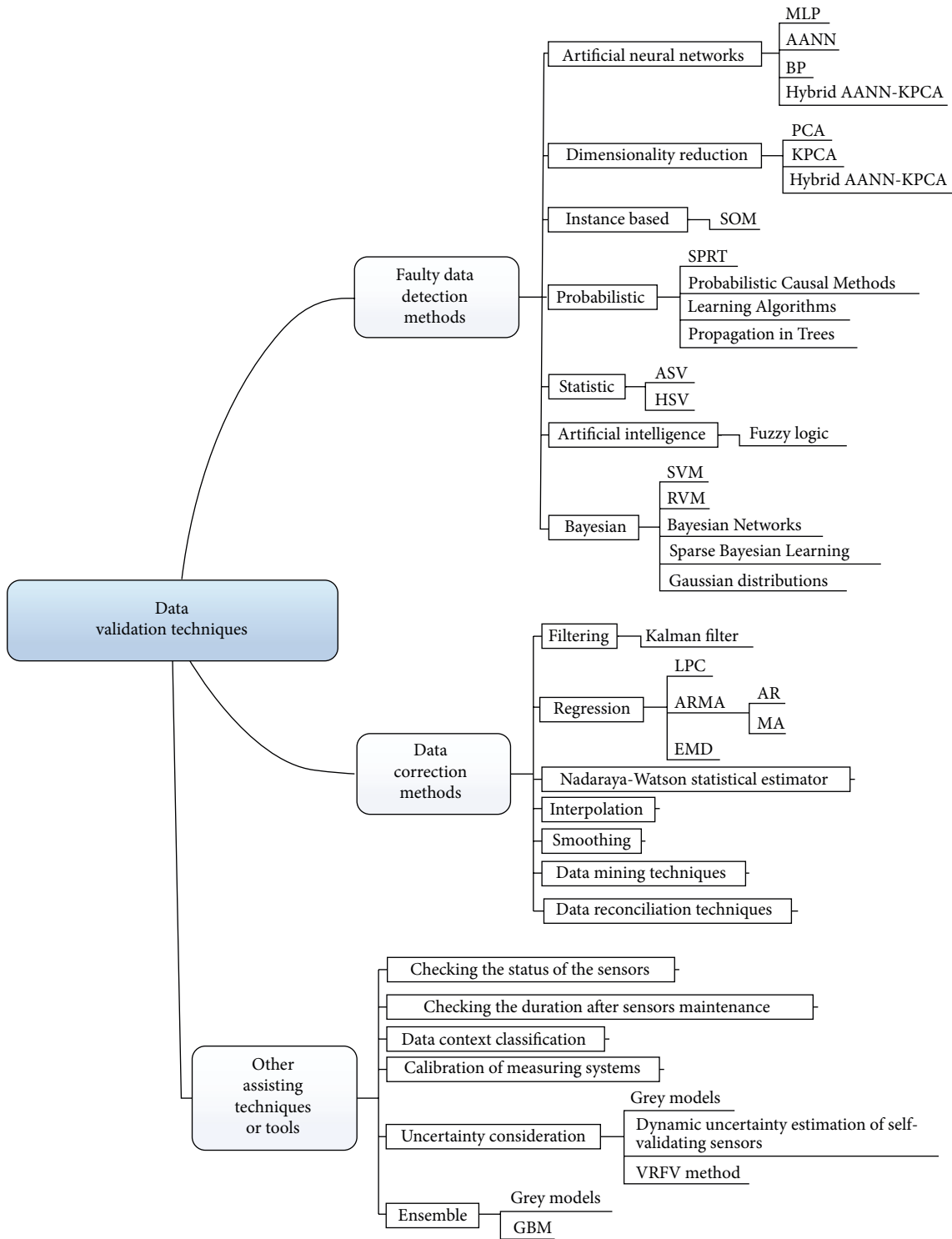


FIGURE 2: Different categories of the data validation methods.

The validation of the data is important, but, for critical systems, for example, clinical systems, not only should the input data be validated, but also the results should be validated to guarantee the reliability, accuracy and, consequently, acceptance of the system. The validation of the system may consist of the detection of failures and the methods that may

be applied are the faulty detection methods. As presented in Section 3, the methods that may be included in this category are probabilistic and statistical methods, among others, which may be used to validate the results of the system. This validation can be performed by comparing the results obtained by an equivalent system which is considered

to be a *gold standard* [53] with the results obtained by the developed methods implemented by different sensors or devices, for example, a mobile device.

Once estimated the initial error of the system, that is, how different the obtained results are from the results obtained by the gold standard system, the validation of the results of the system consist of three steps, such as the definition of the confidence level needed for the acceptance of the system, the determination of the minimum number of experiments needed to validate the application with confidence level defined, and the validation of the results when compared to a golden-standard [54]. The definition of the degree of confidence of the system is a choice of the development team. The system design leader may define what system needs to have a maximum 5% error 95% of the times. Using these parameters, a minimum number of calibration experiments need to be performed to allow the fine tuning of the algorithm. The minimum number of experiments may be measured by several statistical tests, for example, Student's *t*-test [55].

After the calibration of the algorithms in the system, further tests and comparison with golden-standard systems can be done to insure that the results reported by the system have a 5% maximum error when compared to the golden standard results, for 95% of the time. Note that the 5% and 95% values are merely indicative. Moreover, the data collection stage must hold into consideration the limits for the optimal functioning of the sensors. As these limits are extremely dependent on the task the sensors must perform, we do not discuss them in this paper, for example, if the application is supposed to track the movements of a sportsperson in an open environment, it is possible that a thermal sensor reports an environment temperature of  $-5^{\circ}\text{C}$ , yet, for an application that tracks the indoor activity of an elder, such value should raise an alarm. In this extreme case, it is even possible that more robust systems need to contain different types of sensors.

## 5. Conclusion

The validation of the data collected by sensors in a mobile device is an important issue for two main reasons: the first one is the increasing number of devices and the applications that make use of the devices' sensors; the other is that also increasingly users rely on these devices and applications to collect information and make decisions that may be critical for the user's life and well-being.

Despite the fact that there is a wide array and types of data validation algorithms, there is also a lack of published information on the validity of many mobile applications. Also, it is impossible to present a critical comparison of the discussed methods, even within their respective categories, as their efficiency is extremely dependent on their particular usage; for example, the efficiency of a specific method may be very dependent on the number and type of features the algorithm selects on the signal to be processed, and of course these features are chosen in view of the intended purpose of the application. Additionally, it is possible that even with the same chosen method and the same chosen set of

features, different authors report different efficiency ratios; for example, their base population sample varies in size and/or type using different population sizes or using populations that are homogenous in age (elders or youngsters).

This paper has presented a discussion on the different types of data validation methods such as faulty data detection, data correction, and assisting techniques or tools. Furthermore, a classification of these methods in accordance with its functionalities was provided. Finally, the relevance of the data validation methods for critical systems in terms of its reliability, accuracy and acceptance was highlighted. Complementary studies should be addressed aiming at providing an overview on the use of valid data for the identification of the ADLs.

## Competing Interests

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

## Acknowledgments

This work was supported by FCT project UID/EEA/50008/2013 (Este trabalho foi suportado pelo projecto FCT UID/EEA/50008/2013). The authors would also like to acknowledge the contribution of the COST Action IC1303 Architectures, Algorithms and Protocols for Enhanced Living Environments (AAPELE).


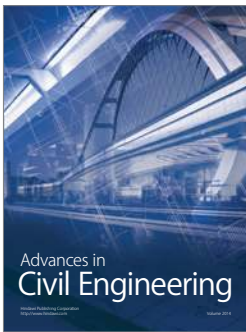
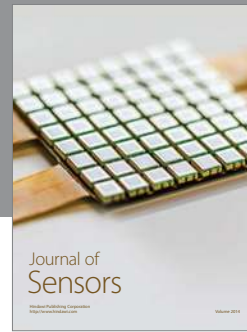
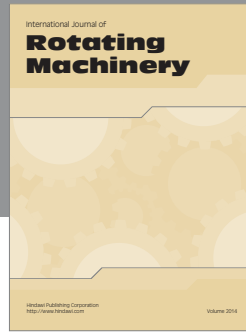
## References

- [1] A. Braun, R. Wichert, A. Kuijper, and D. W. Fellner, "A benchmarking model for sensors in smart environments," in *Ambient Intelligence: European Conference, (AmI '14), Eindhoven, The Netherlands, November 2014. Revised Selected Papers*, E. Aarts, B. de Ruyter, P. Markopoulos et al., Eds., pp. 242–257, Springer, Cham, Switzerland, 2014.
- [2] J. Garza-Ulloa, H. Yu, and T. Sarkodie-Gyan, "A mathematical model for the validation of the ground reaction force sensor in human gait analysis," *Measurement*, vol. 45, no. 4, pp. 755–762, 2012.
- [3] M. A. Cugueró, M. Christodoulou, J. Quevedo, V. Puig, D. García, and M. P. Michaelides, "Combining contaminant event diagnosis with data validation/reconstruction: application to smart buildings," in *Proceedings of the 22nd Mediterranean Conference on Control and Automation (MED '14)*, pp. 293–298, June 2014.
- [4] Y. Chen, J. Yang, and S. Jiang, "Data validation and dynamic uncertainty estimation of self-validating sensor," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC '15)*, pp. 405–410, Pisa, Italy, May 2015.
- [5] S. Sun, J. Bertrand-krajewski, A. Lynggaard-Jensen et al., "Literature review for data validation methods," *SciTechnol*, vol. 47, no. 2, pp. 95–102, 2011.
- [6] J.-L. Bertrand-Krajewski, S. Winkler, E. Saracevic, A. Torres, and H. Schaar, "Comparison of and uncertainties in raw sewage COD measurements by laboratory techniques and field UV-visible spectrometry," *Water Science and Technology*, vol. 56, no. 11, pp. 17–25, 2007.

- [7] N. Branisavljević, Z. Kapelan, and D. Prodanović, "Improved real-time data anomaly detection using context classification," *Journal of Hydroinformatics*, vol. 13, no. 3, pp. 307–323, 2011.
- [8] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*, MIT Press, 1995.
- [9] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp. 144–152, ACM Press, Pittsburgh, Pa, USA, 1992.
- [10] M. Mourad and J.-L. Bertrand-Krajewski, "A method for automatic validation of long time series of data in urban hydrology," *Water Science and Technology*, vol. 45, no. 4-5, pp. 263–270, 2002.
- [11] G. Olsson, M. K. Nielsen, Z. Yuan, and A. Lynggaard-Jensen, *Instrumentation, Control and Automation in Wastewater Systems*, IWA, London, UK, 2005.
- [12] F. Edthofer, J. Van den Broeke, J. Ettl, W. Lettl, and A. Weingartner, "Reliable online water quality monitoring as basis for fault tolerant control," in *Proceedings of the 1st Conference on Control and Fault-Tolerant Systems (SysTol '10)*, pp. 57–62, October 2010.
- [13] S. J. Qin and W. Li, "Detection, identification, and reconstruction of faulty sensors with maximized sensitivity," *AIChE Journal*, vol. 45, no. 9, pp. 1963–1976, 1999.
- [14] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE Journal*, vol. 37, no. 2, pp. 233–243, 1991.
- [15] R. Sharifi and R. Langari, "A hybrid AANN-KPCA approach to sensor data validation," in *Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications*, vol. 7, pp. 85–91, 2007.
- [16] C. C. Castello, J. R. New, and M. K. Smith, "Autonomous correction of sensor data applied to building technologies using filtering methods," in *Proceedings of the 1st IEEE Global Conference on Signal and Information Processing (GlobalSIP '13)*, pp. 121–124, December 2013.
- [17] M. Kasinathan, B. S. Rao, N. Murali, and P. Swaminathan, "An artificial neural network approach for the discordance sensor data validation for SCRAM parameters," in *Proceedings of the 1st International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications (ANIMMA '09)*, pp. 1–5, Marseille, France, June 2009.
- [18] E. V. Zabolotskikh, L. M. Mitnik, L. P. Bobylev, and O. M. Johannessen, "Neural networks based algorithms for sea surface wind speed retrieval using SSM/I data and their validation," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '99)*, vol. 2, pp. 1010–1012, Hamburg, Germany, June-July 1999.
- [19] E. Gaura, R. Newman, M. Kraft, A. Flewitt, and W. de Lima Monteiro, *Smart MEMS and Sensor Systems*, World Scientific, Singapore, 2006.
- [20] P. H. Iburguengoytia, *Any Time Probabilistic Sensor Validation*, University of Salford, Salford, UK, 1997.
- [21] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [22] S. Alag, A. M. Agogino, and M. Morjaria, "A methodology for intelligent sensor measurement, validation, fusion, and fault detection for equipment monitoring and diagnostics," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 15, no. 4, pp. 307–320, 2001.
- [23] E. Shi, "An improved real-time adaptive Kalman filter for low-cost integrated GPS/INS navigation," in *Proceedings of the International Conference on Measurement, Information and Control (MIC '12)*, vol. 2, pp. 1093–1098, May 2012.
- [24] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [25] S. J. Wellington, J. K. Atkinson, and R. P. Sion, "Sensor validation and fusion using the Nadaraya-Watson statistical estimator," in *Proceedings of the 5th International Conference on Information Fusion*, vol. 1, pp. 321–326, Annapolis, Md, USA, 2002.
- [26] K. E. Holbert, A. S. Heger, and N. K. Alang-Rashid, "Redundant sensor validation by using fuzzy logic," *Nuclear Science and Engineering*, vol. 118, no. 1, pp. 54–64, 1994.
- [27] Y. Ai, X. Sun, C. Zhang, and B. Wang, "Research on sensor data validation in aeroengine vibration tests," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '10)*, vol. 3, pp. 162–166, March 2010.
- [28] F. Pfaff, B. Noack, and U. D. Hanebeck, "Data validation in the presence of stochastic and set-membership uncertainties," in *Proceedings of the 16th International Conference on Information Fusion (FUSION '13)*, pp. 2125–2132, Istanbul, Turkey, July 2013.
- [29] M. Staroswiecki, "Intelligent sensors: a functional view," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 4, pp. 238–249, 2005.
- [30] P. H. Iburguengoytia, L. E. Sucar, and S. Vadera, "Real time intelligent sensor validation," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 770–775, 2001.
- [31] J. Rivera-Mejía, E. Arzabala-Contreras, and Á. G. Leyn-Rubio, "Approach to the validation function of intelligent sensors based on error's predictors," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference (I2MTC '10)*, pp. 1121–1125, IEEE, May 2010.
- [32] M. P. Henry and D. W. Clarke, "The self-validating sensor: rationale, definitions and examples," *Control Engineering Practice*, vol. 1, no. 4, pp. 585–610, 1993.
- [33] B. Mounika, G. Raghu, S. Sreelekha, and R. Jeyanthi, "Neural network based data validation algorithm for pressure processes," in *Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCCCT '14)*, pp. 1223–1227, July 2014.
- [34] Q. Wang, Z. Shen, and F. Zhu, "A multifunctional self-validating sensor," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC '13)*, pp. 1283–1288, Minneapolis, Minn, USA, May 2013.
- [35] Z. Shen and Q. Wang, "Data validation and validated uncertainty estimation of multifunctional self-validating sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 7, pp. 2082–2092, 2013.
- [36] Z. Shen and Q. Wang, "Data validation and confidence of self-validating multifunctional sensor," in *Proceedings of the Sensors*, pp. 1–4, Taipei, Taiwan, October 2012.
- [37] J. Doyle, A. Kealy, J. Loane et al., "An integrated home-based self-management system to support the wellbeing of older adults," *Journal of Ambient Intelligence and Smart Environments*, vol. 6, no. 4, pp. 359–383, 2014.
- [38] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.



- [39] B. Lamrini, E.-K. Lakhali, M.-V. Le Lann, and L. Wehenkel, "Data validation and missing data reconstruction using self-organizing map for water treatment," *Neural Computing and Applications*, vol. 20, no. 4, pp. 575–588, 2011.
- [40] 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.
- [41] S. Helal, J. W. Lee, S. Hossain, E. Kim, H. Hagrais, and D. Cook, "Persim—simulator for human activities in pervasive spaces," in *Proceedings of the 7th International Conference on Intelligent Environments (IE '11)*, pp. 192–199, Nottingham, UK, July 2011.
- [42] H. Eren, "Assessing the health of sensors using data historians," in *Proceedings of the IEEE Sensors Applications Symposium (SAS '12)*, pp. 208–211, February 2012.
- [43] S. Oonk, F. J. Maldonado, and T. Politopoulos, "Distributed intelligent health monitoring with the coremicro Reconfigurable Embedded Smart Sensor Node," in *Proceedings of the IEEE AUTOTESTCON*, pp. 233–238, Anaheim, Calif, USA, September 2012.
- [44] C.-M. Chen, R. Kwasnicki, B. Lo, and G. Z. Yang, "Wearable tissue oxygenation monitoring sensor and a forearm vascular phantom design for data validation," in *Proceedings of the 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN '14)*, pp. 64–68, June 2014.
- [45] F. J. Maldonado, S. Oonk, and T. Politopoulos, "Enhancing vibration analysis by embedded sensor data validation technologies," *IEEE Instrumentation and Measurement Magazine*, vol. 16, no. 4, pp. 50–60, 2013.
- [46] E. Miluzzo, *Smartphone Sensing*, Dartmouth College, Hanover, New Hampshire, 2011.
- [47] F. J. Maldonado, S. Oonk, and T. Politopoulos, "Optimized neuro genetic fast estimator (ONGFE) for efficient distributed intelligence instantiation within embedded systems," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '13)*, pp. 1–8, August 2013.
- [48] B. Wallace, R. Goubran, F. Knoefel et al., "Automation of the validation, anonymization, and augmentation of big data from a multi-year driving study," in *Proceedings of the IEEE International Congress on Big Data*, pp. 608–614, New York, NY, USA, June 2015.
- [49] J. Brownlee, *A Tour of Machine Learning Algorithms*, Machine Learning Mastery, 2013.
- [50] R. Denton, "Sensor Reliability Impact on Predictive Maintenance Program Costs," 2010, Wilcoxon Research.
- [51] O. Brand, G. K. Fedder, C. Hierold, J. G. Korvink, O. Tabata, and T. Tsuchiya, *Reliability of MEMS: Testing of Materials and Devices*, John Wiley & Sons, New York, NY, USA, 2013.
- [52] G. Heredia, A. Ollero, M. Bejar, and R. Mahtani, "Sensor and actuator fault detection in small autonomous helicopters," *Mechatronics*, vol. 18, no. 2, pp. 90–99, 2008.
- [53] G. J. Prescott and P. H. Garthwaite, "A simple Bayesian analysis of misclassified binary data with a validation substudy," *Biometrics*, vol. 58, no. 2, pp. 454–458, 2002.
- [54] V. R. Basili, R. W. Selby Jr., and T.-Y. Phillips, "Metric analysis and data validation across fortran projects," *IEEE Transactions on Software Engineering*, vol. 9, no. 6, pp. 652–663, 1983.
- [55] K. R. Sundaram and A. Jose, "Teaching: estimation of minimum sample size and the impact of effect size and altering the type-I & II errors on IT, in clinical research," in *Data and Context in Statistics Education: Towards an Evidence-Based Society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia, C. Reading, Ed., International Statistical Institute, Voorburg, The Netherlands, 2010.*



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

