# Neural Networks for Automated Smart Health Platforms oriented on Heart Predictive Diagnostic Big Data Systems

Alessandro Massaro
Dyrecta Lab IT Research Laboratory
*Via Vescovo Simplicio 45 -70014-*
Conversano (BA)- Italy-

Vincenzo Maritati
Dyrecta Lab IT Research Laboratory
*Via Vescovo Simplicio 45 -70014-*
Conversano (BA)- Italy-

Nicola Savino
Dyrecta Lab IT Research Laboratory
*Via Vescovo Simplicio 45 -70014-*
Conversano (BA)- Italy-

Angelo Galiano
Dyrecta Lab IT Research Laboratory
*Via Vescovo Simplicio 45 -70014-*
Conversano (BA)- Italy-

***Abstract*—The proposed study is related to an experimental application of a multilayer perceptron (MLP) artificial neural network (ANN) oriented to predictive diagnostic and to automated patient monitoring in homecare assistance. By means of a control room it is possible to check periodically the heart rate of patient sending data from home. The ANN-MLP model is designed and developed by a KNIME workflow providing heart rate prediction with a good data processing performance. The paper describes all the steps of the data processing by highlighting the procedures to read the workflow outputs according with optimal heart rate values. The ANN-MLP predictive diagnostic model is suitable to de-hospitalization process enabling telemedicine and smart sensors able to measure patient physiological data. Finally, the paper provides solutions for big data system integration.**

***Keywords—Artificial Neural Networks, Predictive Diagnostic, Smart Health, Heart Rate Prediction, Homecare Assistance, Big Data.***

## I. Introduction and Related Works

An important issue in telemedicine is the predictive diagnostic by means of artificial intelligence (AI). Different open source tools [1]-[4] such as RapidMiner Studio, Weka, KNIME, Orange Canvas, Keras, TensorFlow and Theano are suitable to AI, in particular for the implementation of multilayer perceptron (MLP) artificial neural networks (ANN). These tools are in general appropriate to the design of decision support systems (DSS) implementing artificial intelligence algorithms [5]-[13]. Specifically ANN could be adopted to predict patient status [14]-[15] and heart problems [16]. In particular, continuous heart monitoring represents an important aspect for critical patients assisted at home. In this direction the use of smart sensors transmitting data in a cloud network could support periodically the patient which can be remotely monitored by a control room [17]. Following this requirements is designed the architecture of Fig.1 made by the following components:

- Heart rate smart sensor able to measure the data from patient at home (patient data can be transmitted by a tablet or a mobile phone behaving as a router);
- A central database (DB) available in cloud and collecting all patient data and measurements (for the monitoring of a lot of patient a big data system is required);
- An ANN–MLP workflow model behaving as graphical user interface (GUI) able to predict heart rate values by alerting in case of possible threshold overcoming;
- A control room which displays in real time the patient data [17] and the predicted ones temporizing the ANN-MLP data processing by a chronological (Cron) setting.

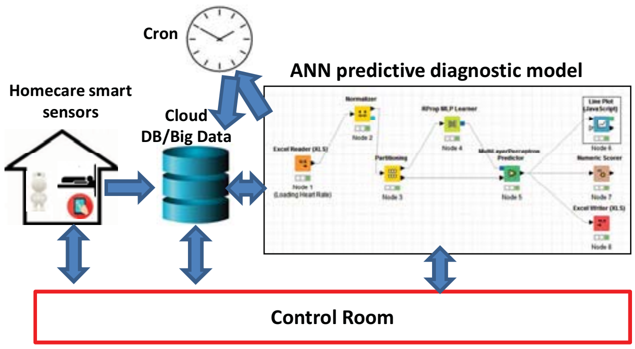


Fig. 1. Architecture of the automated smart heatlh platform improving predictive diagnostic.

From Fig. 1 it is possible to distinguish the different locations of the actors (patient and nurse at home, and doctor in the control room), and the direction of the data flow. The arrows indicated into the ANN predictive diagnostic model are related to the ANN-MLP data processing. The paper is focused on heart rate monitoring and on ANN-MLP workflow development. In order to estimate possible alert thresholds, in Table I and Table II are reported two chart tables of heart rate typical values versus the patient age. Table I is most general then table II and indicates also heart rate values for people having less of 18 years old [18]. It is important to note that the heart rate measurements should be detected in different days because there are many conditions that can affect the heart pulse. Conditions may include:

- fast pulse (exercise or athletic activity, medication, anemia condition, heart disease, use of stimulants, alcohol consumption, feverish state, stress state, etc.);
- slow pulse (bradycardia, hypothyroidism, etc.);
- weak pulse (blood vessels disease, chronic heart disease , blood clots, etc.).

The heart rate values concerning different conditions can be processed by a trained ANN able to learn by labeled

978-8-8872-3740-5 ©2018 AEIT

examples, by processing historical data, the heart behavior of a patient. In this paper are shown the procedures and the approches suitable to the prediction of the heart reate behavior by means of ANN-MLP [19]-[20] implemented in KNIME [19], thus supporting the predictive diagnostic.

TABLE I. OPTIMAL HEART RATE VALUES VERSUS AGES

| *AGE (YEARS)* | *HEART RATE (beats/min)* |
|---|---|
| Less than 1 | 100 to 160 |
| 1 to 2 | 90 to 150 |
| 2 to 5 | 80 to 140 |
| 6 to 12 | 70 to 120 |
| Greather than 12 | 60 to 100 |
| Well-Trained Athletes | 40 to 60 |

TABLE II. OPTIMAL HEART RATE VALUES VERSUS AGE RANGES

| *AGE (YEARS)* | *18-25* | *26-35* | *36-45* | *46-55* | *56-65* | *65+* |
|---|---|---|---|---|---|---|
| Athlete | 49-55 | 49-54 | 50-56 | 50-57 | 51-56 | 50-55 |
| Excellent | 56-61 | 55-61 | 57-62 | 58-63 | 57-61 | 56-61 |
| Good | 62-65 | 62-65 | 63-66 | 64-67 | 62-67 | 62-65 |
| Above Average | 66-69 | 66-70 | 67-70 | 68-71 | 68-71 | 66-69 |
| Average | 70-73 | 71-74 | 71-75 | 72-76 | 72-75 | 70-73 |
| Below Average | 74-81 | 75-81 | 76-82 | 77-83 | 76-81 | 74-79 |
| Poor | 82+ | 82+ | 83+ | 84+ | 82+ | 80+ |

## II. OPERATING CONDITIONS

### A. KNIME ANN Workflow

As for all the self-learned data mining algorithms, MLP ANN needs of different phases for data processing. Figure 2 showa illustrated the KNIME workflow used for heart rate prediction, where it is possible to distinguish the following phases:

- Cron network: 5 Cron blocks connected in series ("Wait…" modules) improve a total delay of 5 days (each KNIME Cron block can delay up to a maximum of 24 hours), so after 5 days will be processed data stored into the big data (the second block will be activated after the expiry of the delay of the first block and so on);
- data input: a "Phyton Source" behaves as a big data connector enabling data processing;
- data pre-processing: input data are normalized in order to breakdown the error due to the use of a no defined scale of values; data are partitioned into two datasets named dataset 1 (train dataset) and dataset 2 (test dataset);
- data processing: dataset 1 is processed by the training node RProp MLP Learner [21]-[22] (RProp algorithm performs a local adaptation of the weight-updates according to the behavior of the error function), and the partitioned dataset 2 is processed by the MLP algorithm [19]-[20] (MultilayerPerceptronPredictor node);
- data output: predicted data are plotted by a line plot node, the numeric scorer provides algorithm performance scoring, and the excel writer stores data output into an excel file useful for a post-processing analysis.

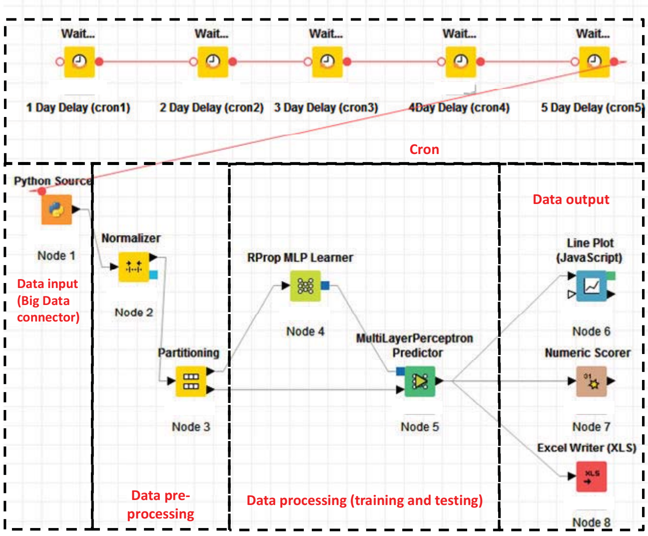


Fig. 2. KNIME ANN workflow processing heart rate signals.

### B. Dataset Description

Figure 3 is plots the experimental dataset related to data of 30 consecutive days. In order to increase the initial data of the training model, the dataset is increased by adding pseudo-random values (low value variations) in the first 56 records and in the last 56 records of the dataset input table, for a total of 168 records. The increase of records is necessary in order to create a consistent training model. The used procedure will be not adopted in case of a large training dataset containing a lot of heart rate historical data. A minimum value of 57 beats/min and a maximum one of 86 beats/min are stored into the experimental dataset. A preliminary calibration of the sphygmomanometer has been performed before all the measurements. More measurements are acquired during a day in order to reduce the error propagation during the data processing. All the measurements have been detected by the same device and are related to the same middle age patient.

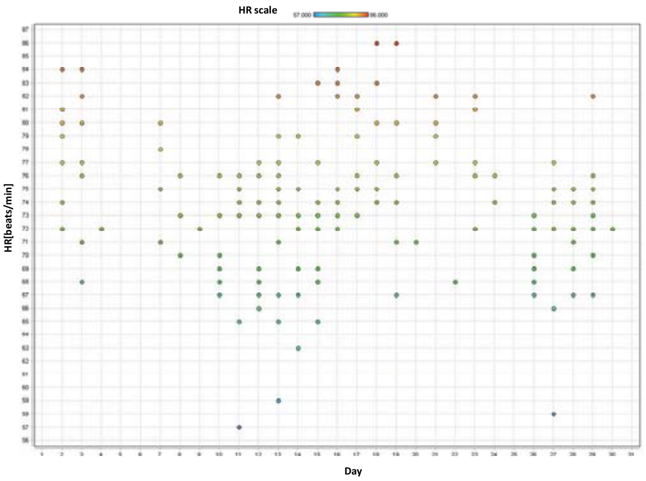


Fig. 3. Experimental dataset (RapidMiner scattering plot libraries).

Table III lists the specifications of the Contec 08A-BT sphygmomanometer used for the dataset acquisition:

TABLE III. SPHYGMOMANOMETER SPECIFICATIONS

| *Specification* | *Description* |
|---|---|
| Input Voltage | AC 100V - 240V ; 60Hz/60Hz; 150mA |
| Output voltage | DC 6.0V +/- 0.2V ; 1.0A |
| IP rank | IP21 ingress of liquids rank |
| Display type | 2.8" color LCD display |
| Measurement Method | Oscillometric |
| Pressure | -range: 0-290 mmHg (0-38.67 kPa);<br><br>-adult pressure: SYS 40-270 mmHg / DIA 10-215 mmHg / Inflation 160+/-5 mmHg;<br>-pediatric pressure: SYS 40-200 mmHg / DIA 10-150 mmHg / Inflation 120 +/-5 mmHg;<br>-neonatal pressure: SYS 40-135 mmHg / DIA 10-100 mmHg / Inflation 70+/-5 mmHg;<br>-resolution: 1 mmHg;<br>-accuracy: +/-3 mmHg static pressure. |
| PR (pulse rate) | -measurement range: 30-250 bpm<br>-error +/- 2bpm |
| Cuff | -range of limb circumference is 6-11cm / 10-19cm / 18-26cm / 32-43 cm<br>-service life 1 year (measuring 6 times a day, everyday) |
| Operating temperature | 5 - 40 °C |
| Power supply | 4 AA alkaline batteries, AC Adapter (AC, 100V-240V optional) |
| Dimension | 130mm x 110mm x 80mm |
| Weight | 300g |
| Safety classification | class II type BF applied equipment |

## III. ANN-MLP RESULTS

A computational cost of only two second is checked for the data processing of the workflow of Fig. 2. The KNIME workflow has been executed by a laptop having the following characteristics: Intel(R) Core(TM) i3-403U CPU, 1.9 GHz, 8 GB RAM, 64 bit operative system. In Fig. 4 is shown the output of the MultiLayerPerceptron Predictor node (node 5) obtained by fixing the following best parameters: first partition of 80 (absolute value), 2 hidden layers, 100 as maximum number of iterations, 10 hidden neurons per layer, linear sampling. The choice of the parameters is provided mainly by the analysis of the mean absolute error versus different parameters, finding the minimum value of 0,141. We observe that, taking into account the total number of 168 records to process, a good compromise is to set the first partitioning to 80 as absolute number of rows. All the 80 rows entered into the first table have been processed by the Learner node (node 4) enabling the training process of the model, while the second table (second partition) contains the remaining rows which have been processed for the testing process directly by the node 5. The predicted results of Fig. 4 oscillate around a value of 0.596 which corresponds to a value of about 75 beats/min: observing Table I and Table II the predicted values match with average condition. The ANN-MLP algorithm takes into account also the incremental information of the measurements: the increase of the day number provides also information about the time series forecasting approach of the model orienting the prediction on the next 30 days. These predicted values are significant for the predictive diagnostic because it is possible to define in the normalized scale different thresholds indicating different risk levels, thus observing if predicted values overcomes the thresholds. The thresholds could be defined by taking into account patient age and pathology. The normalization of the heart rate –HR- values has been performed by changing the scale of the range between minimum and maximum detected value into a scale having a range between 0 and 1 and measured in arbitrary units. The normalization is useful in order to reduce the data processing error due when are considered different attributes having different scale values. The data processing automatism of the workflow in Fig. 2 can be improved by modeling a workflow with KNIME "Wait…" blocks in cascade configuration, enabling time delay and timing (Cron). In Fig. 5 is illustrated the error plot trend thus proving that after yet 14 iterations the error is braked down, and then a fast solution convergence. All the performance scores of the ANN-MLP processing are listed in Table IV thus proving the good convergent solution. The results of Table IV are the output of the Numeric Scorer node.

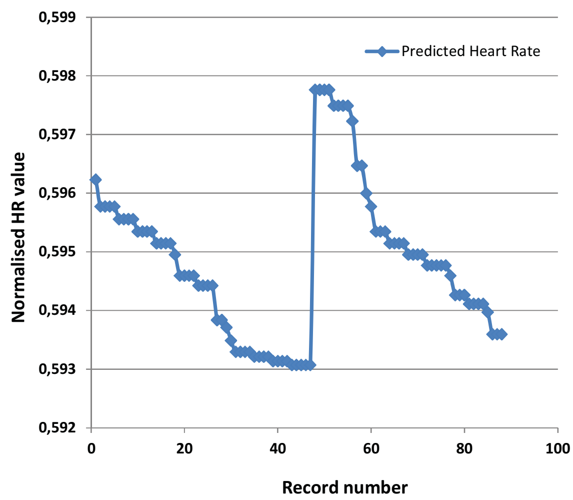


Fig. 4. Predicted heart rate values.

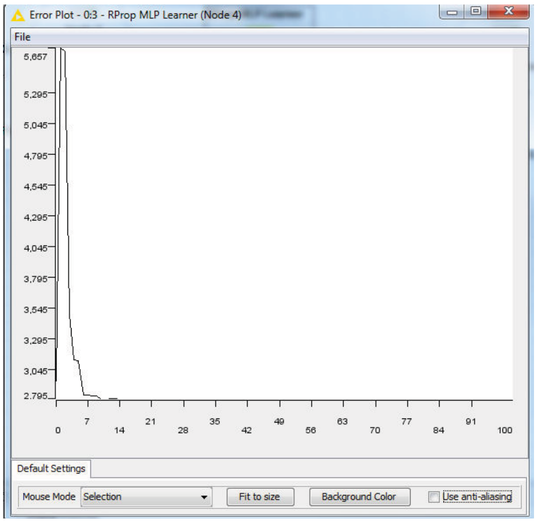


Fig. 5. Error plot trend.

The Numeric Scorer (node 7) computes certain statistics between the numeric column's values ($r_i$) and predicted ($p_i$) values. In particular it computes $R^2=1-\Sigma(p_i-r_i)^2/\Sigma(r_i-1/n*\Sigma r_i)^2$, mean absolute error ($1/n*\Sigma|p_i-r_i|$), mean squared error ($1/n*\Sigma(p_i-r_i)^2$), root mean squared error ($sqrt(1/n*\Sigma(p_i-r_i)^2)$), and mean signed difference ($1/n*\Sigma(p_i-r_i)$). All the listed parameters are useful in order to check the algorithm performance.

TABLE IV. OUTPUT ANN-MLP PERFORMANCE

| ***PARAMETER*** | ***VALUE*** |
|---|---|
| $R^2$ | -0,002 |
| Mean absolute error | 0,141 |
| Mean squared error | 0,033 |
| Root mean squared error | 0,182 |
| Mean signed difference | 0,011 |

### *A. Big Data Connection*

The workflow of Fig. 2 can be connected to a big data system by means of an Enterprise Service Bus –ESB- network [19] able to integrate different data systems and data protocols, or by a Python script enabling connectivity. As example we list below a Python script which can be executed in the “Python Source Node” block of KNIME enabling Cassandra big data linking (see node 1 of Fig. 2):

```
from Cassandra.cluster import Cluster
from Cassandra.auth import PlainTextAuthProvider
from pandas import DataFrame
auth_provider = PlainTextAuthProvider(username=’***’, password=’***’)
cluster = Cluster([“188.166.***.***”], auth_provider=auth_provider)
session = cluster.connect()
session.yet_keyspace(‘test’)
cluster.connect()
query = “SELECT * FROM heart_rate”
rows = session.execute(query)
my_data = []
for d rows:
my_data.append (d)
print (d)
output_table = DataFrame(my_data)
```

## IV. CONCLUSION AND FUTURE WORKS

The paper provides an innovative approach based on a KNIME MLP-ANN model concerning hearth predictive diagnostic. The proposed self-learned workflow can break down the probabilistic error by processing more patient data including historical ones. The model can be applied for other physiological data by setting different alert thresholds. For a capillary homecare assistance network it is possible to connect the timed workflow directly to a big data system. The proposed workflow is suitable to de-hospitalization processes and can be integrated into enterprise resource planning –ERP- platforms. The study of ANN-MLP processing with multiple attributes and comparisons with linear regression approach are under investigation.

## ACKNOWLEDGMENT

The work has been developed in the frameworks of the project: “Piattaforma B.I. intelligente di management risorse e di monitoraggio costi di assistenza sanitaria ‘Healthcare Assistance Platform: Management and Resources Allocation’”. Authors gratefully thanks the researchers: V. Calati, D. Carella, A. Colonna, R. Cosmo, G. Fanelli, R. Guglielmi, A. Leogrande, A. Lombardi, A. Lorusso, N. Malfettone, F. S. Massari, G. Meuli, L. Pellicani, R. Porfido, D. Suma, F. Tarulli, and E. Valenzano.

## REFERENCES

[1] H. Wimmer, and L. M. Powell, "A comparison of open source tools for data science," *J. of Inf. Sys. Appl. Res.*, vol. 9, no. 2, pp. 4-12, 2016.

[2] A. Al-Khoder, and H. Harmouch, "Evaluating four of the most popular open source and free data mining tools," *Int. J. Acad. Scient. Res.*, vol. 3, no. 1, pp. 13-23, 2015.

[3] Gulli, A., and S. Pal, "Deep learning with Keras- implement neural networks with Keras on Theano and TensorFlow," Birmingham-Mumbai Packt book, ISBN 978-1-78712-842-2, 2017.

[4] Kovalev V., Kalinovsky A., and Kovalev S., “Deep Learning with Theano, Torch, Caffe, TensorFlow, and deeplearning4j: which one is the best in speed and accuracy?,” In Proceeding of XIII Int. Conf. on Pattern Recognition and Information Processing, 3-5 October, Minsk, Belarus State University, pp. 99-103, 2016.

[5] Li, J.-S., Yu, H.-Y. and Zhang, X.-G. “Data mining in hospital information system,” Intech, New Fundamental Technologies in Data Mining, Edited by Prof. Kimito Funatsu, 2011.

[6] Goodwin, L.,VanDyne, M., and Lin, S. “Data mining issues and opportunities for building nursing knowledge,” *J. of Biomed. Inf.*, vol. 36, pp. 379-388, 2003.

[7] Belacel N., and Boulassel M.R. “Multicriteria fuzzy assignment method: a useful tool to assist medical diagnosis,” *Art. Intell. in Medicine*, vol. 21, pp. 201-207, 2001.

[8] Demšar, J., Zupan, B., Aoki, N., Wall, M. J., Granchi, T. H. and Beck, J. R., “Feature mining and predictive model construction from severe trauma patient's data,” *Int. J. of Med. Infs.*, vol. 41, no. 50, 2001.

[9] Kusiak, A., Dixon, B., and Shital S., “Predicting survival time for kidney dialysis patients: a data mining approach,” *Comp. In Biol. and Med.*, vol. 35, pp. 311-327, 2005.

[10] Yu, H.-Y., Li, J.-S. “Data mining analysis of inpatient fees in hospital information system,” Proceeding of IEEE International Symposium on IT in Medicine & Education ITME2009, Jinan, China, 14-16 August 2009.

[11] Chae, Y.M., and Kim, H.S. “Analysis of healthcare quality indicator using data mining and decision support system,” *Exp. Sys. with Appl*., vol. 24, pp. 167-172, 2003.

[12] Morando, M., Ponte, S., Ferrara, E., and Dellepiane, S. “Definition of motion and biophysical indicators for home-based rehabilitation through serious games,” *Information*, vol. 9, pp. 1-20, 2018.

[13] Ozcan Y.A, “Quantitative methods in health care management,” California: Josey-Bass 2nd edition, .pp.10-44, ch. 2, 2009.

[14] Ghavami, P., and Kapur, K. “Artificial neural network-enabled prognostics for patient health management. Prognostics and Health Management (PHM),” in IEEE Proceeding of Conference on, Proceeding of Prognostics and Health Management (PHM), Denver, CO, USA, 18-21 June 2012.

[15] Grossi, E. Artificial Neural Networks and Predictive Medicine: a Revolutionary Paradigm Shift, In Artificial Neural Networks - Methodological Advances and Biomedical Applications, 1sth ed.; Suzuki, K. ed.; InTech: University Campus STeP Ri- Slavka Krautzeka 83/A- 51000 Rijeka, Croatia, 2011, Vol. 1, pp. 130- 150.

[16] Adhikari, N. C. D. , “Prevention on hearth problem using artificial intelligence,” *Int. J. of Art. Intell. and Appl. (IJAIA)* , vol. 9, pp. 21-35, 2018, 9.

[17] Galiano, A., Massaro, A. et al. "Improvements in haematology for home health assistance and monitoring by a web based communication system," Proceedings of IEEE International Symposium on Medical Measurements and Applications, MeMeA 2016.

[18] http://www.newhealthadvisor.com/Normal-Heart-Rate-Chart.html

[19] Massaro, A., Maritati, V., Galiano, A., Birardi, V., and Pellicani, L. “ESB platform integrating KNIME data mining tool oriented on Industry 4.0 based on artificial neural network predictive maintenance,” *Int. J. of Art. Intell. and Appl. (IJAIA)*, vol. 9, pp.1-17, 2018.

[20] Al-Shayea, Q. K., “Artificial neural networks in medical diagnosis,” *Int. J. of Comp. Science Issue (IJCSI)* , vol. 8, pp. 150-154, 2011.

[21] Braun, H. A. “Direct adaptive method for faster backpropagation learning: the RPROP algorithm,” Proceedings of the IEEE International Conference on Neural Networks (ICNN), San Francisco, CA, USA, 28 March-1April 1993.

[22] Igel, C., Toussaint, M., and Weishui, W. “RProp using the natural gradient,” In: Mache D.H., Szabados J., de Bruin M.G. (eds) Trends and Applications in Constructive Approximation. ISNM International Series of Numerical Mathematics, Birkhäuser Basel, vol. 151, 259-272, 2005.