Conceptual design of a machine learning-based wearable soft sensor for non-invasive cardiovascular risk assessment

# A machine learning-based wearable soft sensor for real-time non-invasive accurate assessment of cardiovascular risk

Pasquale Arpaia[a,b], Renato Cuocolo[a,c], Francesco Donnarumma[a,d], Antonio Esposito[a,e], Nicola Moccaldi[a,b], Angela Natalizio[a,f], Roberto Prevete[a,b,f]

[a] *Augmented Reality for Health Monitoring Laboratory (ARHeMLab).*
[b] *Department of Electrical Engineering and Information Technology (DIETI), Universita' degli Studi di Napoli Federico II, Naples, Italy.*
[c] *Department of Advanced Biomedical Sciences, Universita' degli Studi di Napoli Federico II, Naples, Italy.*
[d] *Institute of Cognitive Sciences and Technologies, National Research Council (ISTC-CNR), Rome, Italy.*
[e] *Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy.*
[f] *CRdC Tecnologie Scarl, Naples, Italy.*

**Abstract**

The concept design of a soft sensor for measuring cardiovascular risk of a patient in real time is reported. Biosignals such as electrocardiogram, blood oxygenation, and body temperature are acquired through non-invasive wearable transducers. Together with data acquired from patientsâĂŹ interviews, these are processed to extract features characterizing each patient. These are finally classified in order to assess the cardiovascular risk. The soft sensor design relies on a publicly available dataset. Several classifiers were evaluated. Experimental results show that patients' classification accuracy can be as high as 80% when employing a random forest classifier, even with few data employed for training. This result is compared with a previous study on the same dataset, where report accuracy was 73%, achieved by also exploiting invasive measures. Finally, a noise robustness test was carried out and the results show that the classification accuracy remains above 73% even when 10% of noise is added.

*Keywords*: Machine learning, cardiovascular status, soft sensor, non-invasive measurements, wearable sensor.

## 1. Introduction

Worldwide, the number of elderly people is increasing. According to the World Health Organization, the population expecting to live into their sixties and beyond will nearly double between 2015 and 2050 [1]. With the aim of contributing to a healthy ageing of population, heart health is a major issue. Current clinical practice indicates risk calculators as the key tools for cardiovascular risk assessment. They are recommended to select the most appropriate primary prevention measures for each specific patient [2, 3].

Different risk calculators rely on different models built from data acquired and stored over time, such as age, gender, blood pressure, smoking status, and cholesterol [4–7]. Therefore, a model is derived by studying the impact of risk factors on the surveyed population. Typically, the long-term risk is assessed through the conditional probability of mortality from cardiovascular risk disease over a 10-years period. The output of different calculators often varies significantly according to the data set, owing to the specific characteristics of the population, risk factors, mathematical formulations, and events to be predicted [8]. Moreover, the main role in determining the risk is usually undertaken by few parameters, with a poor resolution due to the consideration of only severe events.

A recent trend in cardiovascular risk assessment foresees the employment of further parameters to enhance the resolution [9]. A wearable system helps in recording secondary risk markers, non-detectable in the current clinical practice. To this aim, several systems have been proposed to collect biosignals directly at patient's home [10–15]. On the other hand, the strength of the currently adopted calculators relies on the big data amount. Therefore, wearable systems should not only aim to measure the cardiovascular risk in real time, but also to create a database of biosignals over time.

The literature reports several contributions with artificial intelligence algorithms for diagnosing or predicting cardiovascular diseases. Machine learning techniques are generally applied to imaging examinations [16–19], such as com-

puter tomography, or to the classification of cardiac diseases [20–24]. Conversely, with particular reference to wearable systems, machine learning is usually employed to process data. Measurements and patient records are collected and sent to a central platform, so that their analysis allow the physician to telemonitor the patient status [12–14, 25, 26]. For instance, in 2010, a wearable platform based on smartphones was proposed for real-time detection of cardiovascular diseases trough an electrocardiogram (ECG) [27]. The platform only exploits the ECG analysis through a feed-forward multilayer perceptron for the detection of four types of arrhythmia by using a single-channel transducer connected to a smartphone. In 2013, a detector of atrial fibrillation was also proposed. It exploits signals related to cardiac vibration, namely a ballistocardiogram complementary to an ECG analysis. The system is integrated into the patient's bed for telemonitoring a cardiovascular disease, and signals are analyzed by means of time-frequency-domain or time-domain features and a random forest classifier [28]. Both works refer only to a particular type of cardiovascular disease, while the measurement of cardiovascular risk involves multiple factors. As a final example, in 2018, a wearable system was proposed for the cardiovascular risk assessment [29]. Data from multiple transducers are here exploited, namely electrocardiography, heart rate, blood oxygen, body temperature, and the galvanic skin response for the emotional state assessment . The dataset produced by the authors is available on request. However, the risk is assessed by emotional dynamics, and the patient is required to watch some videos and carry out a self-assessment of the emotive response to these videos.

In this paper, a soft transducer is proposed for measuring the cardiovascular risk in real time, by exploiting biosignals from non-invasive wearable sensors. Results from single patient interviews are also taken into account in order to enhance the accuracy. Measurement non-invasiveness guarantees both everyday usability and rare event detection. Cardiovascular risk is assessed by extending an algorithm of a previous feasibility study reported in [30]. The dataset in [29] is exploited, but the above-mentioned drawbacks are avoided by non considering emotional dynamics.

In particular, Section 2 discusses the basic ideas, the architecture of the soft transducer, the features extraction, and the classification step. Then, Section 3 reports the experimental procedure and the results of the classifier comparison, together with a noise robustness test of the optimal classifier.

## 2. Proposal

### *2.1. Basic Ideas*

The present study is part of a broader project with two main aims: (i) assess the cardiovascular risk in real-time, and (ii) telemonitoring an eventual rehabilitation protocol. A soft transducer is here proposed for the real-time accurate assessment of cardiovascular risk by means of non-invasive measurements. The assessment is based on a model for classifying patients in risk classes.

The input data are obtained by both wearable transducers and patient interview. The classification of the patient according to the heart health status helps both the patient and the physician to take proper countermeasures. Non-invasive measures are processed to obtain features given as input, together with the patient interview, to a decision forest classifier. In particular, [29] demonstrated that this is an optimal machine learning technique to classify these data. The model is trained and calibrated by reference data labeled by an expert cardiologist. The advantage would thus be the avoidance of crowded hospitals and, consequently, of psychological stress for both the patients and the physicians. This will in turn enhance patient care possibilities.

In comparison with the state of the art, this study considers only signals from wearable transducers, without any self-assessment of emotive response to videos watching. Furthermore, the case in which the patient interview is employed was compared to the case in which no clinical data is exploited. Then, a noise robustness test was also carried out in order to generalize the proposed model in absence of new data.

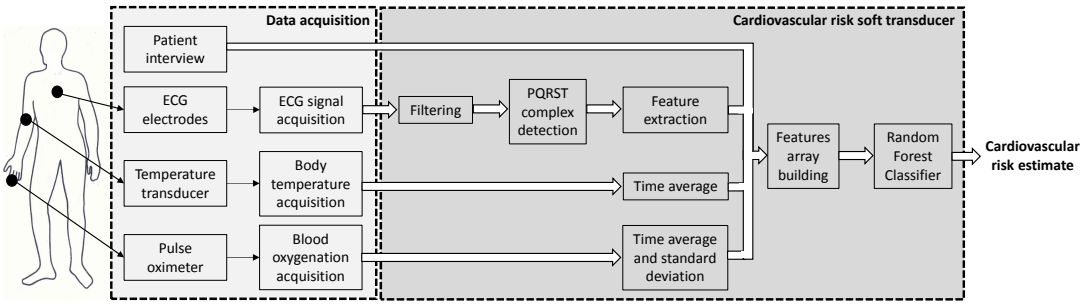


Figure 1: Architecture of the soft transducer for assessing cardiovascular risk.

### *2.2. Architecture*

The architecture of the soft transducer is illustrated in Fig. 1. In the data acquisition section, ECG electrodes, a temperature transducer, and a pulse oximeter collect biosignals. Furthermore, the patient has to answer to some questions regarding his/her life style. These data are the input of the cardiovascular risk soft transducer. In this main block, each signal follows a different path with the aim to extract relevant features: (i) the interview results are directly employed as features, (ii) the ECG signal is filtered, the complex "Preview, Question, Read, Summary, Test" (PQRST) is detected, and the features are extracted, (iii) a time average of the body temperature is carried out, and (iv) the time average of blood oxygenation is calculated together with its standard deviation. Then, the features go in input to the classifier that, once trained, is able to assess the cardiovascular risk of the patient by assigning each patient to a risk class.

### *2.3. Features extraction*

A crucial aspect in the soft transducing is the extraction of features from the available data. Features resulting from each patient's interview are 18: age, gender ('M' or 'F'), body mass index, number of stents (0-7), bypass ('yes' or 'no'), smoking status (0-4), metabolic syndrome ('yes' or 'no'), diabetes status

(0-2), angina or heart attack in a 1st degree relative (0-3), chronic kidney disease (0-3), atrial fibrillation ('yes' or 'no'), on blood pressure treatment ('yes' or 'no'), migraines ('yes' or 'no'), rheumatoid arthritis ('yes' or 'no'), three measures of systolic blood pressure, and diastolic pressure. Features resulting from wearable transducers are instead extracted by processing the electrocardiogram (ECG), the blood oxygenation (SpO2), and the body temperature. These features are all well-established factors influencing the patient cardiovascular risk. They are already used in currently available cardiovascular risk calculators and are discussed in detail in the most recent clinical practice guidelines [31].

The biosignals are acquired when the patient is in a resting state, over a time window of about 10 min. The total number of features from wearable transducers is 35: 32 features are extracted from the ECG, while the remaining 3 features are the mean SpO2 with its standard deviation, and the mean body temperature. The features extracted from the ECG signal are related to the PQRST complex of the ECG signal (Fig. 2) [32], namely the mean and standard deviation of the R-peaks, P-peaks, Q-peaks, S-peaks, J-peaks, and T-peaks, of the segments $\overline{JS}$, $\overline{QRS}$, $\overline{QT}$, and $\overline{PR}$, the heart rate, the mean, standard deviation, minimum and maximum of $\overline{RR}$ distances, the quantities RMSSD, NN50, pNN50, HRVI, SD1, SD2, and the ratio SD1/SD2 related to the $\overline{RR}$ distances (eq. (1)),

$$
\begin{aligned}
RMSSD &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(\overline{RR}_{i+1}-\overline{RR}_i)^2} \\
NN50 &= count(|\overline{RR}_{i+1}-\overline{RR}_i| > 50\,\mathrm{ms}) \\
pNN50 &= 100\cdot\frac{NN50}{N-1} \\
HRVI &= \frac{N}{\max\left[D(\overline{RR})\right]} \\
SD1 &= \sqrt{\frac{1}{2}\sigma\left(\overline{RR}_{i+1}-\overline{RR}_i\right)} \\
SD2 &= \sqrt{\frac{1}{2}\sigma\left(\overline{RR}_{i+1}+\overline{RR}_i\right)}
\end{aligned}
\tag{1}
$$

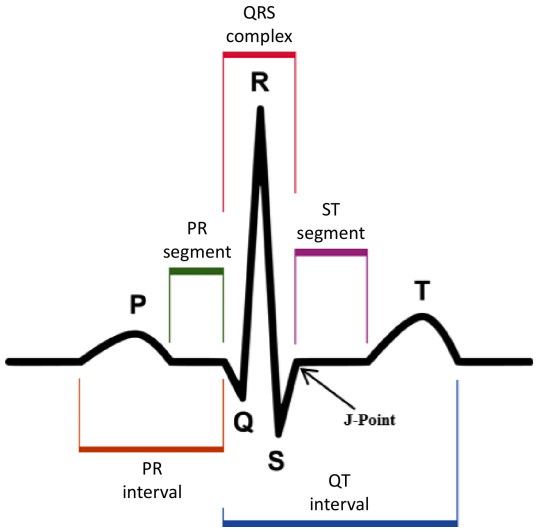


Figure 2: Points of measurement on the ECG curve [32].

In eq. (1), $N$ is the number of $\overline{RR}$ distances extracted from the ECG record, $count(\,)$ is a function that counts the element satisfying the condition specified in the argument, $D(\overline{RR})$ is the density distribution of the $\overline{RR}$ distances, and $\sigma(\,)$ is the function that calculates the standard deviation of the specified argument.

The described features are computed by finding the points of measurement on the ECG signal, as shown in Fig. 2. They are extracted from the ECG curve as reported in [33]. In particular, the baseline of the ECG signal is firstly removed, and the noise is smoothed by a low-pass and high-pass filtering. Then, the peaks of the complex are detected. The R peaks are detected first, because they should assume the maximum value in the PQRST complex. False peaks are rejected by considering the distance of adjacent peaks and their amplitude. Subsequently, starting from the R peaks, the other peaks can be detected, and the distances of interest between peaks are calculated.

### *2.4. Classification*

The extracted features constitute the input of a classifier. Different classifiers were evaluated for discriminating the cardiovascular risk classes: A *multiclass*

*support vector machine (SVM)* with linear kernel or Gaussian kernel, a *random forest*, and *a shallow neural network*.

The *SVM* classifier relies on a hyperplane that guarantees the best separation between features of two different classes [34]. The input features are usually mapped to a higher dimension space through a "kernel function" to deal with non-linear separability of classes. In the present case, a linear kernel and a Gaussian kernel are taken into account. Errors are allowed during separation, but minimizing the separation error is the key for the SVM training. Although the SVM is normally employed for binary classification, multiclass extensions are possible.

A *random forest* is a collection of decision trees [35]. Each tree decides the class of a data according to a random subset of features, selected from the overall features set. The strength of the classifier derives both from this random selection and from the usage of "bagging" (bootstrap aggregating). The scope is to create a forest of uncorrelated trees, whose prediction is more accurate than the prediction of an individual tree itself.

*Shallow neural networks* are universal approximators [36], consisting of three layers: input, with the input features, hidden, and a single output layer. The hidden and output layer are composed of elementary computing units, usually said nodes or neurons. Each neuron receives connections from all the elements of the previous layer only. These connections are weighted, and the weights are found during training. Also, the only hidden layer nodes contain non-linear transfer functions, while a post-processing is possible at the output. Hence, the output can be interpreted as a weighted linear combination of parametric non-linear functions [36]. During the training phase both the parameters (weights) of the non-linear functions and the weights of the linear combination are learned.

In building the soft transducer, the classifier is trained thanks to the labels associated to each patient by a cardiologist. According to their risk class, these labels are "healthy" (0), "unhealthy non-critical" (1), or "unhealthy critical" (2) patient. The above classifiers are selected by means of a repeated stratified k-fold procedure [37]. Each classifier has also a number of specific hyperpa-

rameters to be set (for example, the number of hidden nodes of the shallow neural network). Thus, for each of them, suitable hyperparameters values were selected among a specified subset of the hyperparameters space. Classification metrics such as accuracy, accuracy per class, precision, and recall are taken into account. However, the best performing classifier is chosen according to the only classification accuracy. The type A uncertainty, associated to the accuracy, is also taken into account for the best model selection. This is calculated as the ratio between the standard deviation resulting from the cross-validation and the square root of the number of iterations.

The design was completed by investigating also features normalization and principal component analysis (PCA) in trying to enhance the performance of each classifier. The performance of the classifiers for the present case study were compared experimentally. The metrological performance of the optimal classifier was assessed not only in terms of the classification performance metrics, but also with a noise robustness test.

## 3. Experimental procedure and results

In this section, after a brief description of the dataset, the experimental comparison between different classifiers for the cardiovascular risk assessment is reported. For each benchmark classifier, the optimal hyper-parameters are found by also considering features normalization and Principal Component Analysis (PCA). These hyper-parameters are optimal in the sense of maximizing the classification accuracy, namely the classifier's ability to distinguish the 3 classes of patients. Finally, a noise robustness test was carried out for the best classifier in order to assess the capability of classifying new incoming data. These results are reported and discussed in the following. For comparison purposes, the only classification accuracy was taken into account as performance metric. Moreover, metrics like accuracy per class, precision, and recall were considered also to complete the performance analysis of the optimal classier.

### *3.1. Dataset description*

Data from 30 patients were considered [29]: 16 healthy patients, 10 unhealthy non-critical patients, and only 4 unhealthy critical patients. Their health status was labeled by an expert cardiologist. For each patient, the data acquired through an interview (clinical data) and the data recorded from wearable transducers are furnished. Data are stored into files associated to the different patients. From each file, the ECG signal, blood oxygenation, and body temperature are extracted. Files related to 4 different states per patient are present: namely, 10 minutes resting state, 6-minutes walking test (6MWT), and watching two 6 minutes videos. However, in the present study, only the resting state was taken into account. Among clinical data, some invasive measures were also furnished, such as the results of blood examination. But, with the aim of building a non-invasive system, they were not taken into account.

In the classification step, only 12 patients were considered with the aim to exploit a balanced dataset. In particular, balanced classes were obtained by taking into account 4 healthy subjects, 4 unhealthy non-critical subjects, and the 4 unhealthy critical subjects. For each of them, 10 minutes of registered biosignals and non-invasive clinical data are available. These data were processed in Python according to the features extraction steps described in subsection 2.3. For the classifier identification, the features can be arranged in a matrix where each row of features corresponds to a different patient.

### *3.2. Classifiers comparison*

The selected classifiers were compared by a repeated stratified k-fold, with 4 splits and 16 repetitions, thus exploiting all the 64 possible combinations when one patient per class is considered for the test set. A grid-search method was adopted for selecting the values of the classifier hyperparameters. The above steps are schematically represented in Fig. 3. The hyperparameters taken in account for the Gaussian SVM were the regularization parameter $C$ and the kernel coefficient $\gamma$. Three values were chosen for $C$, namely 1, 10, and 100. Meanwhile, two options were possible for $\gamma$, the inverse of the number of features

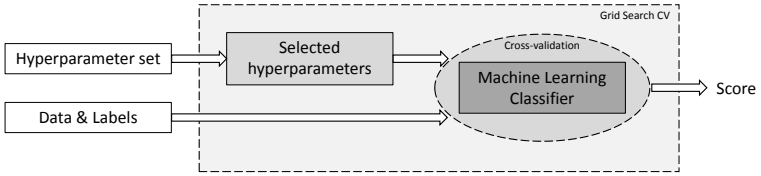


Figure 3: Procedure of classifier model identification: all the combinations of hyperparameters values are selected for the cross-validation and the set leading to the optimal performance score is identified.

("auto"), or the inverse of the number of features times their variance ("scale"). Instead, the only $C$ is taken into account for the linear SVM.

The hyperparameters for the random forest classifier and their ranges are reported in Tab. 1. The "bootstrap" is a re-sampling technique where data are

| **Hyperparameters** | **Ranges/values** |
|---|---|
| Trees in the forest | from 50 to 400 with a step of 50 |
| Maximum depth of the tree | from 10 to 100 with a step of 10 |
| Minimum number of samples required to split an internal node | from 2 to 10 with a step of 1 |
| Number of features to consider when looking for the best split | All the features, or its square root, or again its logarithm with base 2 |
| Bootstrap | True/False |
| Class weight | Balanced/ Balanced subsample |
| Function to measure the quality of a split | Gini/Entropy |

Table 1: Hyperparameters of the random forest classifier and their ranges or specific possible values.

sampled with replacement to estimate statistics on a population. In accordance to that, the class weight has twofold modes: (i) "balanced", where the labels are used to automatically adjust weights that are inversely proportional to class frequencies, and (ii) "balanced_subsample", where weights are computed according to the bootstrap sample for every grown tree.

The ranges for the number of trees, the maximum depth, and the minimum number of samples were established after some preliminary simulations, in order to estimate the order of magnitude of each parameter.

Finally, the shallow neural network was built with a hyperbolic tangent activation function ("tanh") for the hidden nodes, and identity function at the output nodes. The output layer is composed of three nodes, namely one node per class. "Cross-entropy" was employed as cost function, and "soft-max" as post-processing of the network output, while the training algorithm was the resilient propagation (RProp) [38].

The hyperparameters taken into account are the learning rates $\eta^-$ and $\eta^+$, and the number of hidden nodes. The learning rates assess the change between subsequent steps of the network training. The first was set equal to 0.50, while an optimal value was searched for $\eta^+$. In particular, this was varied between 1.05 and 1.30, with step of 0.05. Lastly, the number of hidden nodes was varied between 20 and 220 with step of 40.

For each classifier, the hyperparameter values are searched and the optimal models are reported in Tab. 2. The performance score taken into account to

| **Classifier** | **Optimal hyperparameters** | **Accuracy (%)** | **Accuracy (%) (normalized features)** |
|---|---|---|---|
| linear SVM | C = 1.0 | 65 ± 3 | 60 ± 2 |
| Gaussian SVM | C = 1.0<br>$\gamma$: scale | 52 ± 3 | 59 ± 2 |
| random forest | n_estimators = 250<br>max_depth = 30<br>min_samples_split = 7<br>max_features = None<br>bootstrap = True<br>class_weight = balanced<br>criterion = entropy | 80 ± 2 | 80 ± 2 |
| shallow neural network | number of hidden nodes = 220<br>$\eta^+$ = 1.20 | 51 ± 3 | 61 ± 3 |

Table 2: Performance of different cardiovascular risk models: for each classifier, optimal hyperparameters, classification accuracy, and results with normalized features are reported.

choice the best classifier is the mean classification accuracy among the cross-validation iterations and its associated type-A uncertainty.

The performance of each model was also assessed with features normalized

between 0 and 1. This was done because some features change in a range that is much wider than others. The results show that the best classifier is the random forest, which is also the only classifier not affected by the input normalization.

### *3.3. PCA Analysis*

The performance of the classifiers different from the random forest was improved by processing the features by means of Principal Component Analysis (PCA) [39] prior to the classification step. In particular, the first $p$ principal components were calculated on the training data for each iteration of the cross-validation. Then, both training and test data were projected according to the new components space. Finally, the reduced representations of both datasets go in input to the classifier. The number of principal components $p$ was varied ($p \in \{1, 2, \ldots, 11\}$), and the results are reported in Tab. 3 and in Tab. 4, for non-normalized and normalized features, respectively. Best results are highlighted in gray. In case of non-normalized features and PCA (Tab. 3), there is an enhancement for the only shallow neural network, while the linear SVM performance remains the same and the Gaussian SVM are even worse. In the mean time, if features are normalized (Tab. 4), PCA enhances the performance

| | **number of PCA Components** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Classifiers** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| linear SVM | 45 ± 3 | 56 ± 3 | 57 ± 3 | 64 ± 3 | 64 ± 3 | 65 ± 3 | 65 ± 3 | 65 ± 3 | 65 ± 3 | 65 ± 3 | 65 ± 3 |
| Gaussian SVM | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 | 48 ± 4 |
| shallow neural network | 46 ± 2 | 54 ± 3 | 53 ± 2 | 63 ± 3 | 66 ± 3 | 64 ± 3 | 63 ± 3 | 65 ± 2 | 62 ± 3 | 59 ± 3 | 62 ± 3 |

Table 3: Cardiovascular risk models (Accuracy and type A uncertainty %) obtained by pre-processing input features with Principal Component Analysis (PCA).

| | **number of PCA Components** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Classifiers** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| linear SVM | 49 ± 2 | 52 ± 2 | 51 ± 2 | 58 ± 2 | 62 ± 2 | 63 ± 2 | 64 ± 2 | 60 ± 2 | 60 ± 2 | 60 ± 2 | 60 ± 2 |
| Gaussian SVM | 41 ± 2 | 47 ± 2 | 56 ± 2 | 56 ± 2 | 54 ± 1 | 56 ± 2 | 58 ± 2 | 61 ± 2 | 61 ± 2 | 61 ± 2 | 61 ± 2 |
| shallow neural network | 52 ± 2 | 56 ± 2 | 60 ± 3 | 66 ± 2 | 70 ± 3 | 70 ± 3 | 65 ± 2 | 69 ± 3 | 61 ± 3 | 62 ± 3 | 63 ± 3 |

Table 4: Cardiovascular risk models performance (Accuracy and type A uncertainty %) obtained by pre-processing input normalized features with Principal Component Analysis (PCA).

for each classifier. Notably, the shallow neural network classification accuracy reaches 70%. Nonetheless, the better performance is still achieved with the random forest (without PCA).

### *3.4. Discussion*

The comparison between different classifiers has shown that the best performance in classifying patient into three cardiovascular risk classes is achieved with the random forest. In particular, the mean cross-validation accuracy is 80% with a 2% type A uncertainty. This result is to be compared to a recent study employing the same dataset [29]. This study considered features extracted from heart rate variability analysis in resting state condition, demographic data, biochemical and blood tests results, and other physiological values. The prediction of cardiovascular health status for the patients was possible with a 73% accuracy [29].

Hence, an improvement in terms of performance has been achieved even by employing only non-invasive data. It is also worth noting that misclassified patients can belong to any class, namely there is not a bias in terms of misclassified class. This is reported in Tab. 5, where the accuracy performance is reported per class, together with precision and recall.

| | **Class 0** | **Class 1** | **Class 2** |
|---|---|---|---|
| **Accuracy (%)** | $(95 \pm 2)$ | $(68 \pm 5)$ | $(78 \pm 5)$ |
| **Precision (%)** | $(95 \pm 2)$ | $(58 \pm 5)$ | $(61 \pm 5)$ |
| **Recall (%)** | $(95 \pm 2)$ | $(69 \pm 5)$ | $(78 \pm 5)$ |

Table 5: Random forest performance scores for each class.

The random forest is well-known in the recent scientific literature not only for its classification performance, but also for its advantage to be interpretable [40]. In particular, the importance of each feature in classification can be derived. In the present case, the 15 most important features are reported in Fig. 4.

The figure highlights that the most important feature is "diabetes status", followed by "age". This is in accordance with the current clinical practice be-

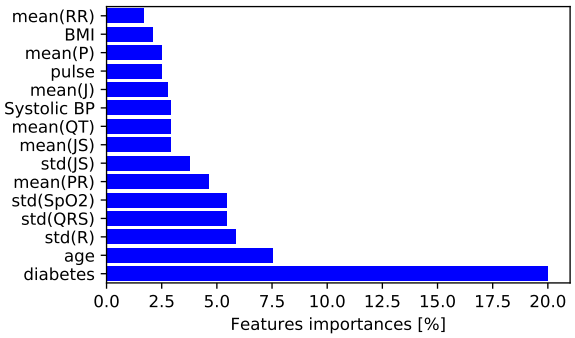


Figure 4: The top 15 most important features for the random forest.

cause, as discussed in the introduction, the two highlighted features are also employed by classical cardiovascular risk calculators. On the other hand, the results of Fig. 4 give some hints about other important factors to assess the cardiovascular risk. Nevertheless, further investigations are needed, especially by exploiting data from more patients.

### *3.5. Noise robustness*

Training a machine learning algorithm with a small dataset can cause the algorithm to merely memorize training samples, thus leading to over-fitting and poor performance on new uncorrelated data. Nevertheless, the model must be capable of properly classifying new data in order to be useful. It is thus necessary to validate the model for new incoming data. If there is no further data for this validation, the creation of synthetic data could be a solution. However, if the starting sample is too small, the synthetic data would result highly correlated to the starting one. A possible solution is to add noise to input variables during training in order to possibility to generalize the model. Such a noise robustness test is executed by adding noise only on test data, while no noise must be added on training data [36]. In this work, a Gaussian noise was added. This was generated with the *randn* Python function, which gives back normally distributed random numbers. Five noise levels were investigated, namely 2%, 4%, 6%, 8%, and 10%. Each level fixes the amplitude of the added

noise. Thirty iterations were performed for each noise level, and in each iteration random noise is added to continuous features, while discrete features remain unchanged. Finally, for each noise level, the mean and standard deviation of the thirty resulting classification accuracies were calculated. The results are shown in Fig. 5. As expected, by adding noise the classifier performance decreases, but

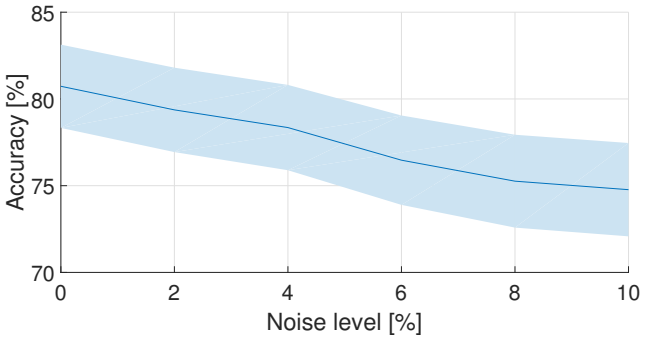


Figure 5: Results of the noise robustness test in terms of classification accuracy and related uncertainty as a function of noise level.

the accuracy is still 75% with 10% of superimposed noise, which is even higher than the abovementioned reference result.

## 4. Conclusion

The concept design of a soft sensor for accurate assessment of the cardiovascular risk of a patient in real time was reported. The system processes measures from wearable transducer and results of patients' interviews. In order to deal with the few available data, different classifier models were compared by a k-fold procedure [37]. The data consisted of measures from just 12 patients, equally divided between healthy, non-critical unhealthy, and critical unhealthy cardiovascular status.

It has been shown that the classification accuracy goes up to 80% with 2% type A uncertainty when a random forest is employed for classification, and this was obtained by merely considering the measures acquired during 10 minutes of patients monitoring in resting state together with other non-invasive data.

These results are also to compare with a recent study reporting 73% accuracy by employing invasive and non-invasive measures. Furthermore, a noise robustness test has shown that, even by adding Gaussian noise to training data, accuracy still remains around 75% with 10% noise level. Thus, the results of this work can be employed for a non-invasive telemonitoring of patients, directly at their home.

In the next future, the system will have the two-fold advantages of both furnishing a real-time measure of the cardiovascular risk of a patient and providing data to create a database for further studies. A mong the other advantages, accuracy improvements will be possible with more data for the model identification.

**Acknowledgment**

This work was funded by the Italian project "Indago" on screening of cardiovascular diseases, PON MISE HORIZION 2014-2020, D.M. 01/06/2016. The authors also thank the "Excellence Department project" (LD n. 232/2016) whose support is gratefully acknowledged.

**References**

[1] World Health Organization (WHO), Ageing and health, 2018. URL: `https://www.who.int/news-room/fact-sheets/detail/ageing-and-health`.

[2] M. F. Piepoli, A. W. Hoes, S. Agewall, C. Albus, C. Brotons, A. L. Catapano, M.-T. Cooney, U. Corra, B. Cosyns, C. Deaton, et al., 2016 european guidelines on cardiovascular disease prevention in clinical practice: The sixth joint task force of the european society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the european association for cardiovascular prevention & rehabilitation (eacpr), European heart journal 37 (2016) 2315–2381.

[3] D. C. Goff, D. M. Lloyd-Jones, G. Bennett, S. Coady, R. B. D'Agostino, R. Gibbons, P. Greenland, D. T. Lackland, D. Levy, C. J. O'Donnell, et al., 2013 acc/aha guideline on the assessment of cardiovascular risk: a report of the american college of cardiology/american heart association task force on practice guidelines, Journal of the American College of Cardiology 63 (2014) 2935–2959.

[4] R. B. D'Agostino, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, W. B. Kannel, General cardiovascular risk profile for use in primary care, Circulation 117 (2008) 743–753.

[5] R. Conroy, K. Pyörälä, A. e. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetiere, P. Jousilahti, U. Keil, et al., Estimation of ten-year risk of fatal cardiovascular disease in europe: the score project, European heart journal 24 (2003) 987–1003.

[6] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, M. May, P. Brindle, Derivation and validation of qrisk, a new cardiovascular disease risk score for the united kingdom: prospective open cohort study, Bmj 335 (2007) 136.

[7] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, R. Minhas, A. Sheikh, P. Brindle, Predicting cardiovascular risk in england and wales: prospective derivation and validation of qrisk2, Bmj 336 (2008) 1475–1482.

[8] G. M. Allan, F. Nouri, C. Korownyk, M. R. Kolber, B. Vandermeer, J. McCormack, Agreement among cardiovascular disease risk calculators, Circulation 127 (2013) 1948–1956.

[9] A. Wierzbicki, New directions in cardiovascular risk assessment: the role of secondary risk stratification markers, International journal of clinical practice 66 (2012) 622–630.

[10] G. Eysenbach, What is e-health?, Journal of medical Internet research 3 (2001) e20.

[11] F. Lamonaca, G. Polimeni, K. Barbé, D. Grimaldi, Health parameters monitoring by smartphone for quality of life improvement, Measurement 73 (2015) 82–94.

[12] E. Kańtoch, Recognition of sedentary behavior by machine learning analysis of wearable sensors during activities of daily living for telemedical assessment of cardiovascular risk, Sensors 18 (2018) 3219.

[13] S. Majumder, M. J. Deen, Smartphone sensors for health monitoring and diagnosis, Sensors 19 (2019) 2164.

[14] F. Landreani, A. Faini, A. Martin-Yebra, M. Morri, G. Parati, E. G. Caiani, Assessment of ultra-short heart variability indices derived by smartphone accelerometers for stress detection, Sensors 19 (2019) 3729.

[15] I. Jabłoński, Integrated living environment: Measurements in modern energy efficient smart building with implemented the functionality of telemedicine, Measurement 101 (2017) 211–235.

[16] G. Singh, S. J. Al'Aref, M. Van Assen, T. S. Kim, A. van Rosendael, K. K. Kolli, A. Dwivedi, G. Maliakal, M. Pandey, J. Wang, et al., Machine learning in cardiac ct: basic concepts and contemporary data, Journal of cardiovascular computed tomography 12 (2018) 192–201.

[17] C. X. Tang, Y. N. Wang, F. Zhou, U. J. Schoepf, M. van Assen, R. E. Stroud, J. H. Li, X. L. Zhang, M. J. Lu, C. S. Zhou, et al., Diagnostic performance of fractional flow reserve derived from coronary ct angiography for detection of lesion-specific ischemia: A multi-center study and meta-analysis, European journal of radiology 116 (2019) 90–97.

[18] E. D. Nicol, B. L. Norgaard, P. Blanke, A. Ahmadi, J. Weir-McCall, P. M. Horvat, K. Han, J. J. Bax, J. Leipsic, The future of cardiovascular computed tomography: Advanced analytics and clinical insights, JACC: Cardiovascular Imaging 12 (2019) 1058–1072.

[19] L. Itu, S. Rapaka, T. Passerini, B. Georgescu, C. Schwemmer, M. Schoebinger, T. Flohr, P. Sharma, D. Comaniciu, A machine-learning approach for computation of fractional flow reserve from coronary computed tomography, Journal of Applied Physiology 121 (2016) 42–52.

[20] J.-H. Eom, S.-C. Kim, B.-T. Zhang, Aptacdss-e: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction, Expert Systems with Applications 34 (2008) 2465–2479.

[21] P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, D. S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, Journal of clinical epidemiology 66 (2013) 398–407.

[22] M. Motwani, D. Dey, D. S. Berman, G. Germano, S. Achenbach, M. H. Al-Mallah, D. Andreini, M. J. Budoff, F. Cademartiri, T. Q. Callister, et al., Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis, European heart journal 38 (2016) 500–507.

[23] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, N. Qureshi, Can machine-learning improve cardiovascular risk prediction using routine clinical data?, PloS one 12 (2017) e0174944.

[24] P. Pławiak, Novel methodology of cardiac health recognition based on ecg signals and evolutionary-neural system, Expert Systems with Applications 92 (2018) 334–349.

[25] S. I. Chaudhry, J. A. Mattera, J. P. Curtis, J. A. Spertus, J. Herrin, Z. Lin, C. O. Phillips, B. V. Hodshon, L. S. Cooper, H. M. Krumholz, Telemonitoring in patients with heart failure, New England Journal of Medicine 363 (2010) 2301–2309.

[26] A.-L. Bleda, F.-M. Melgarejo-Meseguer, F.-J. Gimeno-Blanes, A. García-Alberola, J. L. Rojo-Álvarez, J. Corral, R. Ruiz, R. Maestre-Ferriz, En-

abling heart self-monitoring for all and for aal-portable device within a complete telemedicine system, Sensors 19 (2019) 3969.

[27] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, A. C. Cheng, A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing, IEEE Transactions on Information Technology in Biomedicine 14 (2010) 734–740.

[28] C. Bruser, J. Diesel, M. D. Zink, S. Winter, P. Schauerte, S. Leonhardt, Automatic detection of atrial fibrillation in cardiac vibration signals, IEEE journal of biomedical and health informatics 17 (2013) 162–171.

[29] F. P. Akbulut, A. Akan, A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics, Measurement 128 (2018) 237–246.

[30] P. Arpaia, R. Cuocolo, D. D'Andrea, F. Donnarumma, A. Esposito, N. Moccaldi, A. Natalizio, R. Prevete, Feasibility of cardiovascular risk assessment through non-invasive measurements, 2019 IEEE International Workshop on Metrology for Industry 4.0 and IoT, 2019.

[31] D. K. Arnett, R. S. Blumenthal, M. A. Albert, A. B. Buroker, Z. D. Goldberger, E. J. Hahn, C. R. Dennison-Himmelfarb, A. Khera, D. Lloyd-Jones, J. W. McEvoy, et al., 2019 acc/aha guideline on the primary prevention of cardiovascular disease: executive summary: a report of the american college of cardiology/american heart association task force on clinical practice guidelines, Journal of the American College of Cardiology 74 (2019) 1376–1414.

[32] ECG interpretation part 1: definitions, criteria, and characteristics of the normal ECG (EKG) waves, intervals, durations and rhythm, 2018. URL: https://ecgwaves.com/ecg-topic/ecg-normal-p-wave-qrs-complex-st-segment-t-wave-j-point/.

[33] Development of an ECG risk score system applied for detection and quantification of myocardial damage with a focus on morphological features and de-trended fluctuation analysis, 2017. URL: https://github.com/nichealpham/MY_THESIS.

[34] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., A practical guide to support vector classification (2003).

[35] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.

[36] C. M. Bishop, et al., Neural networks for pattern recognition, Oxford university press, 1995.

[37] S. Arlot, A. Celisse, et al., A survey of cross-validation procedures for model selection, Statistics surveys 4 (2010) 40–79.

[38] M. Riedmiller, H. Braun, Rprop: A fast adaptive learning algorithm, in: Proc. of the Int. Symposium on Computer and Information Science VII, 1992.

[39] I. Jolliffe, Principal component analysis, Springer, 2011.

[40] Interpretability and Random Forests, 2019. URL: https://towardsdatascience.com/interpretability-and-random-forests-4fe13a79ae34.