Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication

Konstantia Zarkogianni, *Member, IEEE*, Maria Athanasiou, Anastasia C. Thanopoulou, and Konstantina S. Nikita, *Senior Member, IEEE*

***Abstract*—The estimation of long-term diabetes complications risk is essential in the process of medical decision making. Guidelines for the management of Type 2 Diabetes Mellitus (T2DM) advocate calculating the Cardiovascular Disease (CVD) risk to initiate appropriate treatment. The objective of this study is to investigate the use of sophisticated machine learning techniques toward the development of personalized models able to predict the risk of fatal or nonfatal CVD incidence in T2DM patients. The important challenge of handling the unbalanced nature of the available dataset is addressed by applying novel ensemble strategies. Hybrid Wavelet Neural Networks (HWNNs) and Self-Organizing Maps (SOMs) constitute the primary models for building ensembles following a subsampling approach. Different methods for combining the decisions of the primary models are applied and comparatively assessed. Data from the 5-year follow up of 560 patients with T2DM are used for development and evaluation purposes. The highest discrimination performance (Area Under the Curve (AUC): 71.48%) is achieved by taking into account both the HWNN- and SOM- based primary models' outputs. The proposed method is superior to the Binomial Linear Regression (BLR) model justifying the need to apply more sophisticated techniques in order to produce reliable CVD risk scores.**

***Index Terms*—Cardiovascular disease, diabetes, machine learning, UKPDS, unbalanced data.**

I. Introduction

TYPE 2 Diabetes Mellitus (T2DM) is the most common form of diabetes affecting 90% of people with diabetes worldwide. Optimal management of T2DM requires a deep understanding of the predisposing factors associated with the disease, early diagnosis and treatment before the occurrence of complications, and tight glycemic control. Early identification of patients at an increased risk of developing diabetes complications is of utmost importance to select appropriate treatment.

Cardiovascular disease (CVD) is the most serious long-term diabetes complication being the major cause of death in people with diabetes, accounting for 50% or more of all diabetes fatalities and severe disabilities. Mortality rates due to CVD in patients with T2DM are 2–4 times higher than the corresponding ones in patients without diabetes.

Taking into account that the medical treatment of CVD related risk factors reduces the occurrence of CVD, along with the increased prevalence of CVD and its economic burden, clinical practice guidelines focus on primary prevention and recommend care givers to evaluate patients for CVD risk factors that may warrant medical treatment [1]. However, it has been shown that care providers cannot estimate the CVD risk on their own [1]. Thus, computational models able to predict the CVD risk can provide valuable tools for treatment planning/selection. International guidelines for T2DM management advocate estimating the CVD risk to initiate appropriate treatment [2], [3].

The European Association for the study of Diabetes (EASD) recommends using FRAMINGHAM [4] and DECODE [5] as preferred prediction models for calculating the CVD risk. However, these models are applicable to the general population and underestimate the risk in diabetes patients [6], [7]. On the other hand, International Diabetes Federation guidelines recommend using the UKPDS risk engine [8] which is dedicated to the T2DM population, but results in varying discriminative performance (c-statistic: 65%–86%) and poor calibration. The methodologies used for the development of these risk engines, are usually based on survival analysis (e.g., cox hazard regression) and regression equations.

In general, a variety of methodologies in the area of statistics and machine learning, such as logistic regression, decision trees, artificial neural networks (ANNs), and Bayesian networks have been applied towards the development of computational models to predict clinical outcomes [8]–[14]. Among these, logistic

Manuscript received March 17, 2017; revised August 11, 2017 and October 1, 2017; accepted October 13, 2017. Date of publication October 22, 2017; date of current version August 31, 2018. This work was supported in part by the EU-FP7 MOSAIC Project (FP7-2011-600914). *(Corresponding author: Konstantia Zarkogianni.)*

K. Zarkogianni and M. Athanasiou are with the Biomedical Simulations and Imaging Laboratory, National Technical University of Athens, Athens 15780, Greece (e-mail: kzarkog@biosim.ntua.gr).

A. C. Thanopoulou is with Diabetes Center, 2nd Department of Internal Medicine, Medical School, University of Athens, Athens 12462, Greece (e-mail: mathanasiou@biosim.ntua.gr; a_thanopoulou@hotmail.com).

K. S. Nikita is with the Faculty of Electrical and Computer Engineering, National Technical University of Athens, Athens 15780, Greece (e-mail: knikita@ece.ntua.gr).

Digital Object Identifier 10.1109/JBHI.2017.2765639

regression and ANNs gained the most widespread acceptance in the field of risk prediction models, due to their simplicity and good predictive ability. ANNs are particularly useful for capturing complex relationships in the data and have been successfully applied in several domains of medical diagnosis and prognosis. A hybrid decision support system for the risk assessment of retinopathy development in patients with Type 1 Diabetes Mellitus has been proposed [9], [14]. The system has been based on the combined use of a back-propagated Feed forward Neural Network (FFN), a Classification and Regression Tree (CART) and a Hybrid Wavelet Neural Network (HWNN). The three primary classifiers have been comparatively assessed and the superiority of the HWNN over both CART and FNN has been demonstrated. Self Organizing Maps (SOMs) constitute another class of ANNs that has been extensively used in a variety of applications for clustering and classification purposes. However, limited SOM-based models have been proposed for assisting disease prognosis and diagnosis [15]–[19]. Recently, the use of SOM has been investigated for the diagnosis of DM [18] as well as diabetic neuropathy [19].

A major issue that is frequently encountered in the field of diagnosing a disease or predicting disease risks is the unbalanced nature of the available medical data [20]. Several techniques have been developed to overcome this problem, such as under- and over –sampling, cost-sensitive learning, and ensemble learning [21].

The present study aims at applying and comparatively assessing sophisticated machine learning techniques towards the development of personalized risk prediction models for the fatal or non-fatal CVD incidence in T2DM, focusing on Coronary Heart Disease (CHD) and Stroke. Taking as input CVD related risk factors, the proposed models output the 5-year risk score for a T2DM patient to experience a CVD incidence for the first time. HWNNs' and SOMs' classification and regression abilities have motivated their application in the present work. In order to handle the unbalanced nature of the data, novel ensemble learning approaches have been deployed. To the best of authors' knowledge, this is the first work proposing the use of HWNNs and SOMs to calculate the CVD risk in T2DM. The performance of the proposed models has been assessed against a Binomial Linear Regression (BLR) model which has been chosen due to its simplicity and wide usage towards the development of risk prediction models [22]. Moreover, a comparison with commonly used machine learning techniques for the development of CVD risk prediction models in T2DM has been conducted.

II. Dataset

The development and evaluation of the proposed risk prediction models was based on data from the medical records of 560 T2DM patients, which were collected from a 5 year follow up at the Hippokration General Hospital of Athens during the period 1996–2007. In this dataset, 41 out of the 560 T2DM patients (7.32%) developed fatal or non-fatal CVD during their 5 year follow up period. Out of the 41 patients with CVD incidents, four patients experienced stroke and the rest experienced CHD. The incidence of fatal or non fatal CVD (positive instances) was encoded to 1 and the non-occurence of CVD (negative instances) to 0. The considered risk factors composing the input space are summarized in Table I along with their descriptive statistics. Baseline data related to demographics, lifestyle, laboratory examinations, and treatment are included, providing adequate information with regard to the clinical status of a T2DM patient [23]. The use of data collected from screening visits to build the proposed risk prediction models renders their adoption in clinical practice feasible.

TABLE I
Risk Factors for the Incidence of Fatal or Non-Fatal CVD in T2DM

Continuous variables	
Risk Factor	Average ± Standard Deviation
Age	58.56 ± 10.70 (*years*)
Diabetes duration	7.67 ± 7.37 (*years*)
Body Mass Index (BMI)	29.49 ± 5.54
Glycosylated Hemoglobin	7.43 ± 1.81 (%)
Pulse Pressure	56.75 ± 15.80 (*mmHg*)
Fasting Glucose	165.15 ± 56.15 (*mg/dL*)
Total Cholesterol	226.64 ± 50.04 (*mg/dL*)
Triglycerides	167.39 ± 110.81 (*mg/dL*)
HDL Cholesterol	48.35 ± 16.46 (*mg/dL*)
Categorical variables	
Risk Factor	Number of patients (Percentage)
Smoking Habit	
Non smokers	289 (51.61%)
Current smokers	146 (26.07%)
Ex- smokers	125 (22.32%)
Sex	
Male	263 (46.96%)
Female	297 (53.04%)
Hypertension	260 (46.42%)
Lipid-lowering therapy	
No	469 (83.75%)
Statins	74 (13.21%)
Fibrates	17 (3.04%)
Aspirin	
No	509 (90.89%),
100 *mg*	44 (7.85%),
325 *mg*	7 (3.03%)
Insulin Therapy	
No	494 (88.21%),
Yes	66 (11.79%)
Parental History of Diabetes	
No	304 (54.28%)
Yes	256 (45.72%)

Each considered factor influences the CVD risk as evidenced by several studies. In particular, age constitutes an important risk factor by approximately tripling the CVD risk with each decade [24]. Duration of diabetes and elevated BMI increase the risk [25]. As an indicator of the average blood glucose concentrations over the preceding 2–3 months, glycosylated haemoglobin (HbA1c) level has been proven to be an independent risk factor for CVD events [26]. The relation between Pulse Pressure (PP) and CHD is nonlinear in patients with T2DM. Patients with PP outside the range of 45 and 55 *mmHg* are at increased risk of future CHD [27]. Elevated fasting glucose levels have been also significantly associated with CVD and all-cause mortality [28]. Abnormal cholesterol levels, including high LDL and low HDL

cholesterol, along with high triglycerides levels indicate poor lipid counts which often occur in patients with premature CHD [29]. Evidently, active smoking is associated with increased risks, up to 50%, of CVD events in T2DM [30]. Although men are at greater risk of heart disease than pre-menopausal women, females with diabetes are twice as likely as males with diabetes to develop heart disease [31]. Hypertension is quantitatively the most important risk factor for premature CVD accounting for an estimated 54 percent of all strokes and 47 percent of all ischemic heart disease events globally [32]. Lipid-lowering therapy and anti-thrombotic agents (e.g., aspirin) are considered to be protective factors for the incidence of CVD [33]. Several large trials have shown that insulin does not increase the CVD risk despite evidence of enhanced atherosclerosis obtained from *in vitro* studies [34]. Parental history of diabetes has been associated with the incidence of CVD in T2DM [35].

III. Methods

A. Conceptual Framework

Taking into account the multifactorial nature of T2DM, HWNNs and SOMs were applied due to their ability to exploit the complex interactions between the risk factors. HWNNs demonstrated increased input-output mapping capabilities by identifying nonlinearities and heterogeneities in diabetes related data [9]. SOM's capacity to capture both density and topology of input data and map them in a bidimesional representation along with their ability to cluster input data in the absence of class memberships' knowledge motivated their use towards detecting and recognizing nonlinearities and heterogeneities inherent to diabetes data facilitating representation of relationships between the risk factors and their interactions.

In order to handle the unbalanced nature of the available dataset, the use of different ensemble methods was investigated. The primary models for building the ensembles were based on HWNN and SOM. Over-fitting to the majority class (negative instances) was avoided by applying a sub-sampling approach [36] during the training stage, as described in detail in Section III-D. Following this approach, m sub-samples from the initial training dataset were created and used for training equally numbered primary models. The outputs of the trained primary models were combined by means of four different schemes resulting in corresponding HWNN- and SOM- based ensembles (Fig. 1). The level of complexity of the combination schemes 1 to 4 ranged from simple (e.g., average) to more sophisticated (e.g., selection based on the best performance achieved on the nearest neighbor training instance). Moreover, a hybrid ensemble was developed by applying a voting scheme to the outputs of the HWNN-based ensemble 4 and SOM-based ensemble 4 (Fig. 2).

The hybrid ensemble integrates the inherent advantages and features of both the HWNN- and the SOM- based ensembles offering the opportunity to produce more flexible relationships between risk factors and the CVD risk by identifying and representing nonlinearities and differentiations in the correlations between certain risk factors in the dataset. The primary models, combination schemes, HWNN- and SOM-based ensembles, as well as the hybrid ensemble are described in detail in the following Sections.

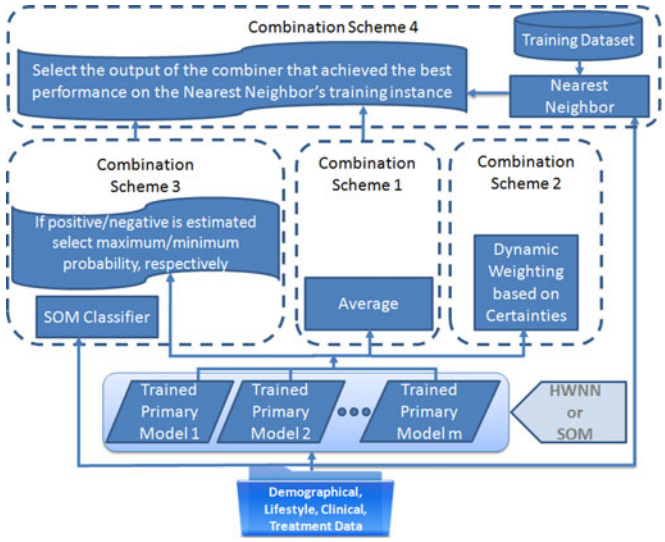

Fig. 1. Conceptual framework. Different combination schemes were developed and applied on the HWNN- and SOM- based primary models towards producing the final CVD risk.

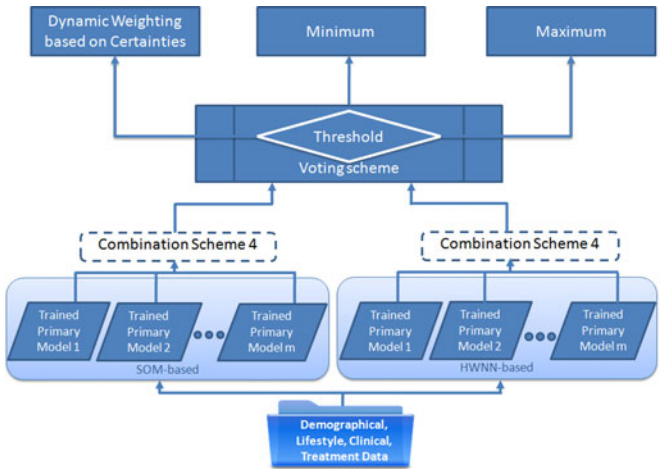

Fig. 2. Outline of the hybrid ensemble.

B. Hybrid Wavelet Neural Network (HWNN)

Wavelet Neural Networks (WNN) belong to a new class of Neural Networks with unique capabilities in system identification and classification [37]. The concept of WNN is inspired by both the technologies of wavelet decomposition and Neural Networks. Particularly, wavelets are obtained from a single prototype wavelet $\psi(t)$ called mother wavelet by dilations and translations,

$$\Omega_c = \left\{ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \right\} \tag{1}$$

where Ω_c is a family of wavelets, α is the dilation parameter and b is the translation parameter, that are real numbers in R and R^+ respectively.

In the present study, a structure of a feed forward Hybrid WNN (HWNN) was used employing the wavelet (1) and the sigmoid function as the activation function of the hidden and output layer units, respectively [9], [37]. A bias term was added in the hidden layer. In order to handle back propagation's problem of local minima, two momentum terms were used in its learning algorithm. Moreover, a modified cost function was defined as follows:

$$J=\sum_{i=1}^{N}\frac{1}{2}\Gamma_i\cdot(y_{iHWNN_i}-y_{real_i})^2,\quad \Gamma_i=\begin{cases}1\ if\ y_{real_i}=0\\ \gamma\ if\ y_{real_i}=1\end{cases} \tag{2}$$

where y_{iHWNN_i} and y_{real_i} are the HWNN's output and the target output, respectively, and N is the number of patterns of the training data set. By applying the sub-sampling approach, the HWNN was trained with a representation of positive instances equal to 33.33%. Given that the CVD incidence rate in the general T2DM population is lower, training the HWNN with this dataset would avoid over-fitting but would cause the prediction of high risk scores which would, in turn, result in bad calibration performance. To overcome this problem, a weighting coefficient (Γ_i) was applied to the cost function in order to decrease the contribution of the positive instances to the weights' adjustment. Oscillations were avoided and the convergence was accelerated by using an adaptive learning rate, based on Vogl's algorithm [38]. Furthermore, a particular initialization procedure was followed to ensure that the wavelets covered the entire input space [37]. Initial weights were randomly set to small values and the bias term was initialized to the average output.

C. Self Organizing Map (SOM)

In general, SOMs belong to the category of competitive learning networks and are widely used for data clustering and visualization of high dimensional data. Their philosophy is based on the notion of unsupervised learning, according to which a system can learn to represent input data in the absence of any information related to the target outputs. However, SOMs can be trained to learn input-output mappings and can be, thus, effectively applied for function approximation [39], [40]. Through a competitive-cooperative learning scheme, the neurons of the SOM are driven to capture the spatial relationships of input data and finally perform a vector quantization of the input space.

In the present study, a technique that generated multiple local linear models following a SOM-based vector quantification was deployed [41]. In particular, the SOM consisted of a two dimensional grid of neurons (j), each of which was associated with a weight vector (w_{in}) of identical dimension to that of the input, a coefficient vector (v) of a linear autoregressive model with exogenous inputs and a weight output value (w_{out}). All weights were randomly initialized and subsequently updated at every iteration of the training stage, based on the Euclidean distance between the current input instance vector (x_{in}) and the weight vectors. The winner neuron $(j^*(t))$ was selected as the one that minimized the Euclidean distance:

$$j^*(t)-\arg\min\{\|x_{in}(t)-w_{in}(t)\|\}. \tag{3}$$

Input and output weights were updated according to the equations explicitly presented in [39], [42]. In this way, the degree to which each neuron was affected depended on its proximity to the input, and the inputs' area of influence was restricted as the training procedure evolved, so that eventually only the winner neuron's weights were updated.

The estimation of the final CVD risk score for every testing input instance was performed by deployment of a novel method, according to which the neurons' outputs of the whole grid were taken into account. Particularly, for each testing input instance the CVD probabilities $(\widehat{y_j})$ were calculated by applying the local linear models associated to each neuron (j), as follows:

$$\hat{y}_J=v_jT\cdot x_{in} \tag{4}$$

The weighted average of the probabilities produced by the local linear models constituted the final CVD risk. The corresponding weights assigned to each individual probability were based on the Euclidean distance of the testing input instance from each neuron, according to the following equation:

$$w_{neuron}=\begin{cases}\frac{d_k-d_j}{d_k-d_1}, & \text{if } d_k\neq d_1\\ 1, & \text{if } d_k=d_1\end{cases} \tag{5}$$

where d_j represents the Euclidean distance of the testing input instance to the weight vector of neuron j, d_1 and d_k denote the input's distance from the nearest and the furthest neuron, respectively.

D. Ensemble Learning Based on Sub-Sampling Approach

The dataset used for the development and the evaluation of the personalized risk prediction models, includes a small number (7.32%) of CVD incidents. Since most classifiers are built with the assumption that the testing data are drawn from the same distribution as the training data, this unbalanced dataset would cause over-training the models. To reduce the over-fitting, appropriate testing and training datasets should be created. Therefore, an ensemble learning technique was followed, in which multiple primary models were individually trained using subsets of the original data while their outputs were combined to produce the final CVD risk prediction of a patient. The training subsets were created using a sub-sampling approach so that the minority representation in each subset was equal to 33.33% [36]. For the creation of the sub-samples, the initial training dataset was firstly divided into positive and negative instances. Sub-samples were created using all positive instances from the initial training dataset and double sized randomly selected negative instances. Following this approach, 6 different sets of subsamples $(m=6)$ were created and used for training 6 primary models. The adopted ensemble learning approach is presented in Fig. 3.

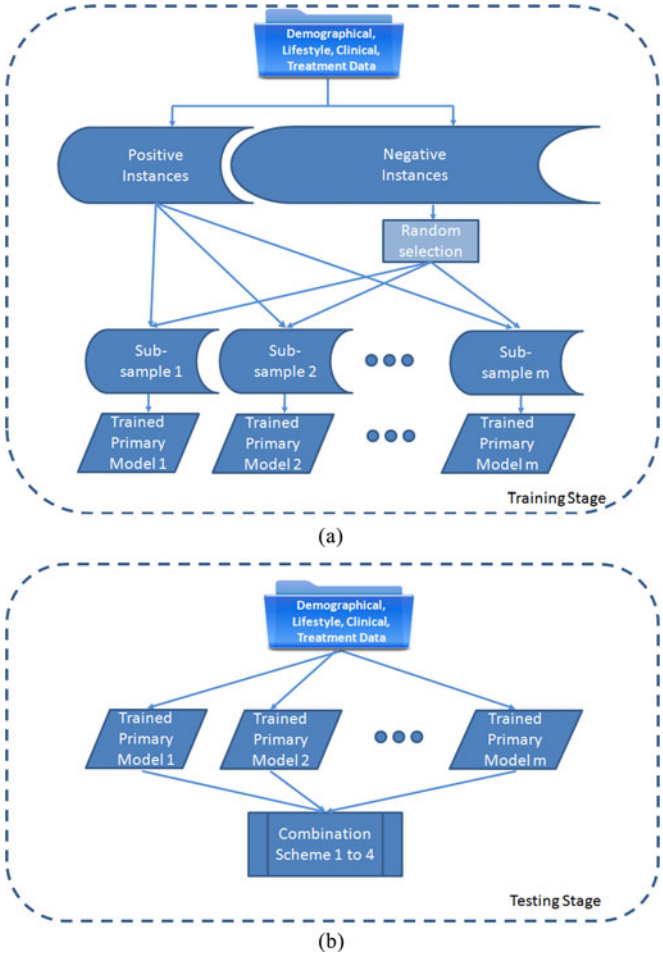

Fig. 3. Upper panel: Ensemble learning based on sub-sampling approach. Lower panel: The application of combination schemes 1 to 4 on the trained primary models led to corresponding HWNN- and SOM-based ensembles.

E. Combination Schemes and Ensembles

Four methods to combine the outputs of each individually trained primary model (HWNN or SOM) were applied, resulting in four HWNN- and four SOM- based ensembles.

1) *Combination scheme 1:* The first scheme was based on the Ensemble Averaging (EA), in which the outputs of the individually trained models were simply averaged (*ensemble 1*).
2) *Combination scheme 2:* The second scheme followed the dynamic weighted average approach to obtain the final risk score (*ensemble 2*), as

$$\bar{y}(x) = \frac{1}{m} \cdot \sum_{net=1}^{m} w_{net} \cdot y_{net}(x) \tag{6}$$

where $\overline{y(x)}$ is the final risk score, m denotes the number of the individually trained models $(m = 6)$, w_{net} and $y_{net}(x)$ represent the weights and the outputs of the trained primary models, respectively.

For the calculation of the weights (w_{net}), the Dynamic Weighting based on Certainties (DWC) [43], [44] was used, according to which the weights were set proportional to the certainties (c_{net}) of the respective trained primary model output, using the following equations:

$$w_{net} = \frac{C_{net}}{\sum_{net=1}^{m} C_{net}} \tag{7}$$

$$c_{net} = \begin{cases} y_{net}, & \text{if } y_{net} \geq 0.5 \\ 1 - y_{net}, & \text{if } y_{net} < 0.5. \end{cases} \tag{8}$$

The values of the weights depended on the outputs of the trained primary models. The certainty of a specific output increased when the output is closer to 0 or 1.

3) *Combination scheme 3:* Selection of the minimum or maximum estimated probability by the trained primary models was performed according to the output produced by a SOM classifier concerning positivity or negativity for CVD. Particularly, if the input instance was classified as negative, the minimum output value was selected, otherwise the maximum output value was set as the final predicted probability (*ensemble 3*).

The SOM classifier was built following an approach able to handle highly unbalanced datasets [45]. The approach resulted in the creation of clusters labeled as positive or negative. The general architecture included two phases in the testing stage. During the first phase, the new testing input instance was classified as negative if it was assigned to a negative cluster, otherwise it was forwarded to the second phase, where it received the label of the winner neuron's cluster. For this purpose, the following training procedure was implemented:

1) *Phase 1:* Negative instances were firstly filtered out of the training dataset and used to train the first SOM (SOM_1). The trained SOM's neurons were considered as the clusters' centers. The clusters were created taking into account the minimum Euclidean distance of both the positive and negative training instances from each cluster center. A cluster was labeled as negative if all the training instances that were assigned to it belonged to the negative class, otherwise, it was labeled as positive.
2) *Phase 2:* Negative training instances that were assigned to positive clusters during phase 1 were used to train the second SOM (SOM_2). After the training procedure was completed, these negative training instances along with the positive ones were compared (Euclidean distance) with the clusters' centers (e.g., SOM's neurons) in order to create the final clusters. The challenge of assigning labels to inhomogeneous clusters was addressed by using the 0–1 knapsack approach [45]. Appropriate labels were assigned to the clusters so that the False Positive Rate was lower than a predefined value. The training and testing algorithm are summarized by the pseudo code given in Algorithms 1 and 2, respectively.

During the training stage of each phase, the neighborhood function was modified so that at every iteration only the winner neuron's weights were updated [46].

The dimension of the SOM classifier's input space was decreased in order to avoid the creation of too many

Algorithm 1: Training stage of the SOM Classifier.

1. Input**:** A matrix **I** containing the training instances
2. Output: Two trained SOM nets $(\mathrm{SOM}_1, \mathrm{SOM}_2)$, labeled clusters of the training instances

//Phase 1

3. Use the negative training instances to train SOM_1
4. Apply SOM_1 on the training instances (both negative and positive) in order to form clusters of all instances included in **I**
5. **for all** clusters **do**
6. **if** cluster contains only negative instances
7. assign a negative label
8. **else**
9. assign a positive label
10. **end if**
11. **end for**

//Phase 2

12. Create a matrix $\mathbf{I_2}$ containing the training instances that were assigned to clusters with positive labels through Phase 1.
13. Use the negative training instances included in $\mathbf{I_2}$ to train SOM_2.
14. Apply SOM_2 on the training instances (both negative and positive) included in $\mathbf{I_2}$ in order to form clusters
15. **for all** clusters **do**
16. **if** cluster contains only negative instances
17. assign a negative label
18. **else**
19. assign a mixed label
20. **end if**
21. **end for**
22. Apply the 0–1 knapsack approach to the clusters with mixed label in order to classify each for positive or negative label

Algorithm 2: Testing stage of the SOM Classifier.

1. Input: A matrix **X** containing testing input instances, trained SOM nets $(\mathrm{SOM}_1, \mathrm{SOM}_2)$, labeled clusters of the training instances
2. Output: A vector containing labels of the testing input instances
3. **for all** testing input instances in **X do**

//Phase 1

4. apply SOM_1 to assign the testing input instance to a cluster
5. **if** cluster is negative
6. assign a negative label to the instance
7. **else**

//Phase 2

8. apply SOM_2 to assign the testing input instance to a cluster
9. **if** cluster is negative
10. assign a negative label to the instance
11. **else**
12. assign a positive label to the instance.
13. **end if**
14. **end if**
15. **end for**

inhomogeneous clusters which would lower its performance. For this reason, an identical set of risk factors from the input space to those used by the UKPDS risk engine were used. According to Table I, the input domain consisted of: i) Age, ii) Duration of Diabetes, iii) Sex, iv) Smoking Habit, v) HbA1c, vi) Systolic Blood Pressure, vii) Total Cholesterol, and viii) HDL Cholesterol.

4) *Combination scheme 4:* The fourth method included a selection between the decisions produced by ensemble versions 1 and 3, based on their corresponding performance achieved on the nearest neighbor training input instance. More specifically, by following the k nearest neighbors approach, for every testing input instance, the nearest neighbor in the training dataset was determined. If the nearest neighbor had developed CVD, the combination scheme that produced the highest probability was selected to be applied for the testing instance, otherwise the scheme that produced the lowest probability was used (*ensemble 4)*.

F. Hybrid Ensemble

A hybrid ensemble was developed by applying a voting scheme to the outputs of the HWNN-based ensemble 4 and the SOM-based ensemble 4 (Fig. 2). Following the sub-sampling approach, several HWNNs were trained while the SOM-based vector quantification technique was deployed as described in Section III-C, towards generating multiple local linear models in order to create the trained primary SOM models. As depicted in Fig. 2 the outputs of the trained primary HWNNs and SOMs were combined using the combination scheme 4 (Section III-E) towards producing two decisions of the CVD risk, respectively. The final decision was provided by applying a voting scheme on the two estimated CVD risks. In particular, for every testing input instance, the maximum or minimum decision was selected, depending on whether both decisions were greater or lower than a certain threshold value, respectively. In all other cases, decisions were combined by applying the DWC method. The threshold value was chosen equal to 0.1.

G. Evaluation Criteria

In order to assess the generalization ability of the developed risk prediction models, 10-fold cross-validation was used. The division of the dataset into 10 sets of identical size was based on the DUPLEX data splitting method with the aim of achieving effective coverage of the training input space [47]. According to DUPLEX, data samples are drawn to be added to a set based on their mutual Euclidean distance. The predictive performance of the models was measured for both discrimination and calibration. Discrimination refers to the ability of the model to separate

the patients who developed the disease from those who did not, by providing higher risk scores to the former case. One of the most reliable and popular measures of discrimination is the Area Under the Curve (AUC), also known as c-statistic [48]. An AUC of 100% indicates perfect discrimination ability, while an AUC of 50% provides worthless performance. In addition to AUC, the classification accuracy, sensitivity and specificity were also used as a measure of the models' discriminative ability. Accuracy corresponds to the percentage of correctly predicted outcomes. Sensitivity reflects the model's ability to correctly identify the positive incidents, while specificity represents the percentage of the correctly predicted negative for CVD outcomes. Calibration measures how close the predictions are to the actual probability. A commonly used criterion of calibration is the Brier score [49], which measures the mean square difference between the estimated and the actual risks of CVD incidence.

Further to the above criteria, the clinical impact of the model, which achieved the best performance, was also assessed by applying the Net Benefit criterion [50],

$$Net\ Benefit = \frac{TPC}{N_t} - \frac{FPC}{N_t} \cdot \left(\frac{p_t}{1-p_t}\right) \qquad (9)$$

where p_t, N_t, TPC, and FPC represent the threshold probability, the total number of patients, the true positive counts and the false positive counts, respectively. Assuming that the threshold probability of developing CVD, at which a patient would initiate treatment, is informative of how the patient weighs the relative harms of a false-positive and a false-negative prediction, the net benefit criterion was calculated across different threshold probabilities towards the creation of the "decision curve". Based on decision curve analysis, the range of threshold probabilities in which a model is of value was identified.

IV. Results

A. Parameters Tuning

1) *HWNN-based primary model*: In the present study, the Morlet wavelet was selected as the mother wavelet for the HWNN, because of its high resolution both in time and frequency domains. The initial values of the learning rate (η) and the two momentum terms (α_1, α_2), as defined in [37], were set to 0.1, 10^{-4} and 10^{-4}, respectively. Parameter γ in (2), was set to 0.5.
2) *SOM-based primary model:* A grid of 3×3 was found to be large enough to quantize the input data. The number of epochs was selected equal to 100. Parameters a_0, a_T, and s_T, which were associated with the learning rate as defined in [39], [42], were set equal to 0.9, 0.01, and 1, respectively. Before training the models, a normalization of the data within the range $[-1, 1]$ was performed. For this reason, the initial values of the weights were randomly selected within the range $[-1, 1]$.
3) *SOM classifier:* A grid of 5×5 and 3×3 neurons was selected for the first and second phase of the SOM classifier, respectively. The false positive rate in the 0–1 knapsack problem algorithm, was set equal to 35%. The values of all the other tuning parameters were identical to those used in the SOM-based primary model.

B. Evaluation of Models' Performance

Table II summarizes the results obtained by the HWNN- and SOM-based ensembles 1 to 4 and the hybrid ensemble. Accuracy, sensitivity and specificity were calculated by setting the probability threshold equal to 10%.

The simple combination schemes 1 and 2 achieved low discrimination ability for both the HWNN- based ensembles (AUC values: 59.97 ± 15.65% and 60.03 ± 15.70%, respectively) and the SOM- based ensembles (AUC values: 61.85 ± 8.12% and 61.46 ± 8.68%, respectively). The higher values of specificity (68.44 ± 26.76 % and 75.18 ± 24.27%) with respect to sensitivity values (39.50 ± 29.67% and 30.00 ± 32.91%) obtained by HWNN-based ensembles 1 and 2 indicate over-fitting to the majority class. On the other hand, the SOM-based ensembles 1 and 2 achieved low specificity (20.43 ± 4.38% and 23.32 ± 5.10%) and high sensitivity values (90.00 ± 17.48% and 90.00 ± 17.48%).

The application of more sophisticated combination schemes increased the ensembles' performance. In particular, the minimum - maximum selection via the SOM classifier (combination scheme 3) demonstrated higher average AUC values for both the HWNN- (61.83 ± 19.90%) and the SOM- based (67.55 ± 14.61%) ensembles 3. Taking also into account the performance of ensemble 1 and 3 on the nearest neighbor training instance in order to choose the best final decision between ensemble 1 and 3 (combination scheme 4), the AUC value was further increased up to 67.64 ± 15.09% and 70.54 ± 13.72% for HWNN- and SOM-based ensemble 4, respectively. A low sensitivity and high specificity value was obtained by applying the HWNN-based ensemble 4, while the opposite was achieved by the SOM- based ensemble 4. This outcome motivated the development of the hybrid ensemble, which resulted in the highest AUC (71.48 ± 15.73) and acceptable levels of both sensitivity (61.00 ± 26.65%) and specificity (72.64 ± 9.44%) for this unbalanced dataset. As it is demonstrated by the low Brier scores, all ensembles were well-calibrated.

For reference purposes, the discriminative ability of the SOM classifier was also assessed (Table II). The obtained results indicated low sensitivity (51.50 ± 27.79%) and acceptable accuracy (74.46 ± 7.05%) and specificity (76.31 ± 7.12%).

In order to compare the averaged criteria obtained by applying the 10-fold cross validation, the pairwise t-test was used. Table III presents the p-values obtained by comparing the hybrid ensemble with all the other ensembles and the BLR model. Statistical significant differences were observed for most of the cases demonstrating the hybrid ensemble's superiority.

A boxplot presenting the distributions of the hybrid ensemble's estimated risks for the patients who developed CVD and those who did not is depicted in Fig. 4. In can be seen that significantly higher probabilities (p-value < 0.05) were estimated for the patients who developed CVD. The hybrid ensemble's clinical impact was also assessed by applying the Net Benefit criterion. The obtained decision curve presented in Fig. 5

TABLE II
DISCRIMINATION AND CALIBRATION PERFORMANCE FOR EACH ENSEMBLE, THE SOM CLASSIFIER, THE BLR MODEL, THE FNN, THE CART, THE RANDOM FOREST AND THE NAÏVE BAYES

		Mean ± SD (%)				
		AUC	ACC	SPC	SENS	Brier Score
HWNN-based ensemble	***1***	59.97 ± 15.65	66.25 ± 23.18	68.44 ± 26.76	39.50 ± 29.67	0.07 ± 0.01
	2	60.03 ± 15.70	71.19 ± 21.28	75.18 ± 24.27	30.00 ± 32.91	0.07 ± 0.01
	3	61.83 ± 19.90	77.68 ± 8.72	81.52 ± 9.71	29.50 ± 25.98	0.09 ± 0.03
	4	67.64 ± 15.09	83.04 ± 8.22	87.30 ± 9.73	29.50 ± 23.15	0.08 ± 0.01
SOM-based ensemble	***1***	61.85 ± 8.12	25.54 ± 3.77	20.43 ± 4.38	90.00 ± 17.48	0.08 ± 0.01
	2	61.46 ± 8.68	28.21 ± 4.19	23.32 ± 5.10	90.00 ± 17.48	0.08 ± 0.01
	3	67.55 ± 14.61	48.21 ± 5.19	46.25 ± 5.88	73.00 ± 18.59	0.08 ± 0.02
	4	70.54 ± 13.72	49.11 ± 5.00	46.44 ± 5.77	83.00 ± 16.70	0.07 ± 0.01
Hybrid ensemble		71.48 ± 15.73	71.79 ± 9.06	72.64 ± 9.44	61.00 ± 26.65	0.07 ± 0.02
SOM classifier		-	74.46 ± 7.05	76.31 ± 7.12	51.50 ± 27.79	-
BLR model		55.11 ± 9.81	41.25 ± 12.70	39.08 ± 14.36	67.50 ± 20.58	0.09 ± 0.02
FFN		60.09 ± 19.70	55.54 ± 12.89	55.49 ± 13.93	57.00 ± 31.90	0.09 ± 0.03
CART		47.99 ± 8.64	78.57 ± 10.10	83.84 ± 11.10	11.50 ± 15.47	0.12 ± 0.03
Random Forest		60.80 ± 18.65	68.04 ± 10.96	70.53 ± 11.97	37.00 ± 37.95	0.07 ± 0.01
Naive Bayes		67.19 ± 8.71	68.04 ± 3.99	68.79 ± 3.90	58.50 ± 11.80	0.10 ± 0.06

The discrimination ability was measured using the C-statistic (AUC), specificity (SPC) and sensitivity (SENS). The calibration ability was measured by calculation of the brier score. criteria values were calculated by averaging the corresponding values to the testing sets generated by the 10-fold cross-validation.

TABLE III
P-VALUES OBTAINED BY APPLYING T-TEST TO COMPARE THE PREDICTIVE PERFORMANCE OF HYBRID ENSEMBLE WITH THAT OF HWNN-BASED ENSEMBLES 1 TO 4, SOM-BASED ENSEMBLES 1 TO 4 AND THE BLR MODEL, IN TERMS OF THE C-STATISTIC (AUC), ACCURACY (ACC), SPECIFICITY (SPC), AND SENSITIVITY (SENS) CRITERIA

		Hybrid ensemble			
		AUC	ACC	SPC	SENS
HWNN-based ensemble	***1***	**0.04**	0.45	0.62	0.13
	2	**0.04**	1.00	0.75	**0.03**
	3	0.07	**0.00**	**0.00**	**0.01**
	4	0.16	**0.00**	**0.00**	**0.01**
SOM-based ensemble	***1***	**0.03**	**0.00**	**0.00**	**0.00**
	2	**0.02**	**0.00**	**0.00**	**0.00**
	3	0.24	**0.00**	**0.00**	0.15
	4	0.55	**0.00**	**0.00**	**0.00**
BLR model		**0.02**	**0.00**	**0.00**	0.47

depicts the magnitude of the Net Benefit across different threshold probabilities including also the cases of assuming that all patients are positive (dashed line) and negative (dash-dot line) for CVD, respectively. The hybrid model is of value within the range of 4% and 14%.

C. Comparison With the State of the Art

The results obtained by the best performing hybrid ensemble were comparatively assessed against those obtained by applying the UKPDS risk engine on the same dataset [53]. The development of the UKPDS risk engine was based on multivariate logistic regression while data from 5102 T2DM patients were used. The publicly available executable file of the UKPDS risk engine v2 was downloaded, set up and applied on the entire Hippokration dataset in order to produce the estimated probabilities for each patient to develop CHD and Stroke within the 5-year time

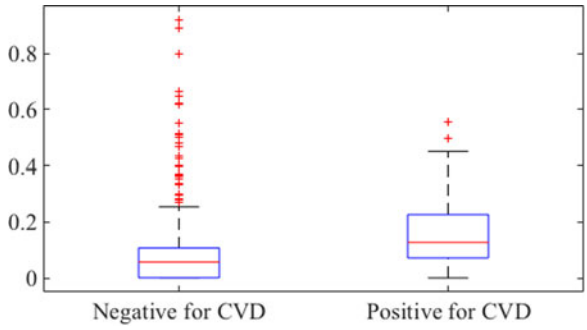

Fig. 4. Distributions of the CVD risks estimated by applying the hybrid ensemble.

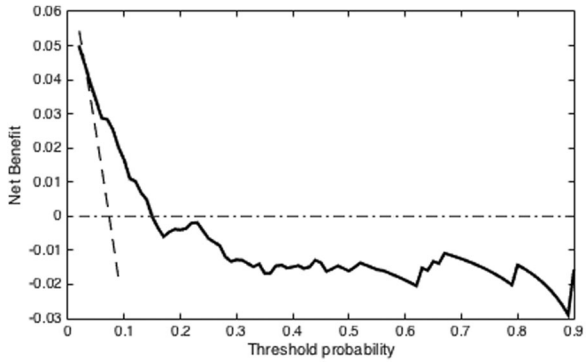

Fig. 5. Decision curve for the hybrid ensemble. Net benefit amplitude calculated across different threshold probabilities based on the risk produced by the hybrid ensemble (solid line) and on the assumption that all patients are positive (dashed line) and negative (dash-dot line) for CVD.

period. For comparison purposes, the maximum value between the estimated CHD and Stroke risk was taken into consideration while, for those patients who developed either CHD or Stroke, the corresponding estimated risk was considered. In order to

provide a fair comparison, the multiple test sets generated from the 10-fold cross validation were merged together into one large test set, which included all the patients of the Hippokration dataset. This test set was used to produce the AUC obtained by applying the hybrid ensemble. The hybrid ensemble provided much higher discriminative ability than the UKPDS (70.00% versus 58.74%).

Since logistic regression is one of the most commonly used methods towards the development of CVD risk prediction models, for comparison purposes, a BLR model was also implemented, applying the logit link function,

$$\log\left(\frac{\mu}{1-\mu}\right) = X \cdot b_B \tag{10}$$

where μ, X and b_B represent the CVD risk, a vector including the CVD risk factors (Table I) and the coefficients, respectively. Following the 10-fold cross validation criterion, the obtained results (Table II), demonstrated low discrimination performance (AUC: 55.11 $\pm$ 9.81%) and proved BLR model's inability to handle the unbalanced nature of the dataset. As presented in Table III, the hybrid ensemble achieved significantly higher discrimination performance in terms of AUC (p-value $= 0.02$), ACC (p-value $= 0.00$) and SPC (p-value $= 0.00$). Although the BLR model provided a measure of the association between the CVD risk factors and the CVD onset in a clinically interpretable way, it failed to produce reliable CVD risk scores. On the other hand, the hybrid ensemble did not require pre-specification of a model structure and resulted in a better CVD prediction model at the sacrifice of interpretability of how risk factors relate to the CVD incidence.

A comparison between the obtained results and those reported in the literature was, also, carried out. Although a direct and fair comparison was not feasible due to different datasets, input spaces and evaluation frameworks, substantial inferences could be obtained. The most well-studied 5-year CVD risk prediction models dedicated to T2DM population were based on survival analysis and regression models while the reported AUCs ranged from 67% to 85% (average 72.80%) [51]. However, in all cases the validation method was based on split sampling which usually resulted in biased analysis since the choice of the training and testing dataset strongly affected the models' performance. This becomes of greater importance when considering the unbalanced nature of the data used in all studies. The limitations of the State of the Art (SoA) models were also evidenced by the low validated performance achieved on external datasets [52]. In the present study, the application of the 10-fold cross validation provided a reliable measure of the models' generalization capabilities. In this sense, although the AUC obtained from the hybrid ensemble was slightly lower than the average by the SoA models (71.48% vs 72.80%), it highlighted the ensemble's potential to produce more accurate risk scores.

Having recognized the increased capabilities of machine-learning approaches to improve the risk accuracy by exploiting complex interactions between risk factors, recently reported studies have been focused on applying machine learning techniques for the calculation of the CVD risk in T2DM [53]. The most commonly used approaches were based on neural networks, Bayesian, and decision trees. In order to provide a fair comparison with the SoA, the performance of the hybrid ensemble was comparatively assessed with those obtained by applying a Feedforward Neural Network (FNN), a Classification and Regression Tree (CART), a Random Forest, and a Naïve Bayes classifier, on the Hippokration dataset. The results presented in Table II justify the superiority of the hybrid ensemble over the other models.

V. Discussion

Existing CVD risk prediction models in T2DM are mainly based on advanced statistics. In the present study, the use of advanced machine learning approaches was investigated and compared with the SoA justifying the need to apply more sophisticated techniques towards producing reliable CVD risk scores. The proposed ensembles of HWNNs and SOMs addressed the commonly encountered challenge of handling the unbalanced nature of the available datasets as well as issues associated with the complex interactions and nonlinearities inherent in diabetes related data.

The deployment and comparison of several combination schemes towards the creation of the ensembles highlighted the increased effectiveness of the advanced combination schemes against simpler ones, with respect to the ensembles' discriminative ability. Moreover, the SOM- based ensembles' superiority over the HWNN-based ensembles was shown. The combined use of both ensembles (hybrid ensemble) further improved the overall performance, which was attributed to the diversity among the members of the ensemble.

Evidence regarding the hybrid ensemble's potential to produce more reliable CVD risk scores than those reported in the literature was provided. The comparison between the results obtained by applying the 10-fold cross validation to evaluate the hybrid ensemble's performance and those reported in the literature, which, in most cases, were produced following the split sampling validation method, illustrated the higher generalization capabilities of the hybrid ensemble over the existing CVD risk prediction models for patients with T2DM. Moreover, the application of the UKPDS risk engine on the Hippokration dataset resulted in lower discrimination performance than the hybrid ensemble. In order to provide a fair comparison with the SoA, commonly used statistics (e.g., BLR) and machine learning strategies (e.g., FFN, CART, Random Forest, and Naïve Bayes) were applied and evaluated on the Hippokration dataset. The obtained results indicated the hybrid ensemble's ability to produce more accurate CVD risk scores.

Although the main scope of the present study was to provide evidence of the effectiveness of the proposed machine learning techniques and not to propose a final CVD risk prediction model to be used in clinical practice, the hybrid ensemble's clinical assessment highlighted its potential to support clinical decision making. Furthermore, its usability was ensured since it only required information from patients' medical history and easily available laboratory tests.

Potential limitations of the present study refer to the nature of the used dataset and the increased inherent complexity of the

proposed models. Particularly, data corresponding to homogeneous population in terms of race and ethnicity were used for development and evaluation purposes. This was beneficial to the proposed models as compared to the UKPDS risk engine, which was built based on data corresponding to cohorts of T2DM patients with larger variance regarding their race/ethnicity and validated on the population used in this study. Furthermore, the proposed machine learning approaches were characterized by higher complexity than logistic and cox hazard regression, making thus the representation of each risk factor's impact and interpretation of the predicted outputs more difficult.

Future work concerns the validation of the proposed methods on datasets of larger size and heterogeneity with the ultimate goal to provide a well validated CVD risk prediction model to be adopted in clinical practice. The integration of the model into a web-based medical decision support system that provides user-friendliness though interfaces incorporating simplified charts and tables, is expected to facilitate the interpretation of the results.

VI. Conclusion

Within the framework of the present study, the use of advanced machine learning techniques based on HWNNs, SOMs and ensemble learning towards producing CVD risk scores for the T2DM population was investigated. The application of a sub-sampling learning approach led to the creation of multiple HWNN- and SOM-based primary models. Several combination schemes were deployed to merge the outputs of the individually trained primary models and comparatively assessed. The models were tested using data from the medical records of 560 T2DM patients, and the best discrimination performance was up to 71.48% in terms of AUC. The obtained results indicate that a hybrid ensemble integrating both the HWNN- and SOM- based primary models performs well even if the dataset used includes a small number of CVD incidents. The introduction of larger datasets corresponding to patients with different ethnicity and race will extend the applicability of the model to other cohorts of patients.

References

[1] M. Matheny *et al.*, "Systematic review of cardiovascular disease risk assessment tools," Evidence Synthesis no. 85, AHRQ Publication 11-05155-EF-1, Rockville, MD, USA: Agency Healthcare Res. Qual., May 2011.

[2] L. Rydén *et al.*, "Diabetes, pre-diabetes and cardiovascular diseases developed with the EASD," *Eur. Heart J.*, vol. 34, pp. 3035–3087, 2013.

[3] National Collaborating Centre for Chronic Conditions, "Type 2 diabetes: National clinical guideline for management in primary and secondary care (update)," London: Royal College of Physicians, 2008.

[4] K. Eichler, M. Puhan, J. Steurer, and L. Bachmann, "Prediction of first coronary events with the Framingham score: A systematic review," *Amer. Heart J.*, vol. 153, pp. 722–731, 2007.

[5] B. Balkau, G. Hu, Q. Qiao, J. Tuomilehto, K. Borch-Johnsen, K. Pyorala, and European Diabetes Epidemiology Group, "Prediction of the risk of cardiovascular mortality using a score that includes glucose as a risk factor. The DECODE Study," *Diabetologia*, vol. 47, no. 12, pp. 2118–2128, 2004.

[6] P. Chamnan, R. K. Simmons, S. J. Sharp, S. Griffin, and N. Wareham, "Cardiovascular risk assessment scores for people with diabetes: A systematic review," *Diabetologia*, vol. 52, pp. 2001–2014, 2009.

[7] P. Brindle, A. Beswick, T. Fahey, and S. Ebrahim, "Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: A systematic review," *Heart*, vol. 92, pp. 1752–1759, 2006.

[8] R. J. Stevens, V. Kothari, A. I. Adler, and I. Stratton, "The UKPDS risk engine: A model for the risk of coronary heart disease in type 2 diabetes," *Clin. Sci.*, vol. 101, pp. 671–679, 2001.

[9] M. Skevofilakas, K. Zarkogianni, B. Karamanos, and K. Nikita, "A hybrid decision support system for the risk assessment of retinopathy development as a long term complication of type 1 diabetes mellitus," in *Proc. 32nd Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2010 pp. 6713–6716.

[10] K. Valavanis, S. Mougiakakou, K. Grimaldi, and K. Nikita, "A multifactorial analysis of obesity as CVD risk factor: Use of neural network based methods in a nutrigenetics context," *BMC Bioinformat.*, vol. 11, no. 453, pp. 1–10, Sep. 2010.

[11] T. Ayer, J. Chhatwal, O. Alagoz, C. E. Kahn, R. Woods, and E. Burnside, "Informatics in radiology: Comparison of logistic regression and artificial neural network models in breast cancer risk estimation," *RadioGraphics*, vol. 30, pp. 13–22, 2010.

[12] K. Zarkogianni, A. Vazeou, S. Mougiakakou, and K. Nikita, "An insulin infusion advisory system based on autotuning nonlinear model-predictive control," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2467–2477, Sep. 2011.

[13] K. V. Dalakleidi, K. Zarkogianni, V. G. Karamanos, A. C. Thanopoulou, and K. S. Nikita, "A hybrid genetic algorithm for the selection of the critical features for risk prediction of cardiovascular complications in Type 2 Diabetes patients," presented at the *IEEE 13th Int. Conf. Bioinformatics Bioengineering*, Nov. 2013, pp. 1–4.

[14] S. Mougiakakou *et al.*, "SMARTDIAB: A communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. Special issue: New and emerging trends in bioinformatics and bioengineering," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 622–633, May 2010.

[15] G. S. Uttreshwar and A. A. Ghatol, "Hepatitis B diagnosis using logical inference and self-organizing map 1," *J. Comput. Sci.*, vol. 4, pp. 1042–1050, 2008.

[16] T. Faisal, F. Ibrahim, and M. N. Taib, "Analysis of significant factors for dengue infection prognosis using the self organizing map," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 5140–5143.

[17] M. A. Fekihal and J. H. Yousif, "Self-organizing map approach for identifying mental disorders," *Int. J. Comput. Appl.*, vol. 45, no. 7, pp. 25–30, 2012.

[18] T. Daghistani and R. Alshammari, "Diagnosis of diabetes by applying data mining classification techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 7, pp. 329–332, 2016.

[19] M. Luboschik *et al.*, "Supporting an early detection of diabetic neuropathy by visual analytics," in *Proc. EuroVis Workshop Vis. Anal.*, 2014, pp. 49–53.

[20] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *Med. Informat. Decision Making*, vol. 11 no. 51, pp. 1–13, Jul. 2011.

[21] G. Nguyen Hoang, A. Bouzerdoum, and S. Phung, "Learning pattern classification tasks with imbalanced data sets," in *Pattern Recognition*. P. Yin Ed. Vukovar, Croatia: In-Teh, 2009, pp. 193–208.

[22] I. Colombet, A. Ruelland, G. Chatellier, F. Gueyffier, P. Degoulet, M. C. Jaulent, "Models to predict cardiovascular risk: Comparison of CART, multilayer perceptron and logistic regression," in *Proc. AMIA Symp.*, 2000 pp. 156–60.

[23] A. Dagliati *et al.*, "A data gathering framework to collect type 2 diabetes patients data," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat.*, Jun. 2014, pp. 244–247.

[24] J. A. Finegold, P. Asaria, and D. P. Francis, "Mortality from ischaemic heart disease by country, region, and age: Statistics from World Health Organisation and United Nations," *Int. J. Cardiol.*, vol. 168, pp. 934–945, Dec. 2012.

[25] C. S. Fox, L. Sullivan, R. B. D'Agostino, Sr., P. W. Wilson, and Framingham Heart Study., "The significant effect of diabetes duration on coronary heart disease mortality: The Framingham Heart Study," *Diabetes Care*, vol. 27, pp. 704–708, Mar. 2004.

[26] I. Cavero-Redondo, B. Peleteiro, C. Álvarez-Bueno, F. Rodriíguez-Artalejo, and V. Martiínez-Vizcaiíno, "Glycosylated haemoglobin as a predictor of cardiovascular events and mortality: A protocol for a systematic review and meta-analysis," *BMJ Open*, vol. 6, pp. 1–5, 2016.

[27] A. Nargesi *et al.*, "Nonlinear relation between pulse pressure and coronary heart disease in patients with type 2 diabetes or hypertension," *J. Hypertension*, vol. 34, pp. 974–980, May 2016.

[28] M. Mongraw-Chaffina *et al.*, "A prospective study of low fasting glucose with cardiovascular disease events and all-cause mortality: The Women's Health Initiative," *Metabolism*, vol. 70, pp. 116–124, May 2017.
[29] American Heart Association, "Cardiovascular disease & diabetes," Apr. 2017. [Online]. Available: https://goo.gl/6kNtfW
[30] A. Pan, Y. Wang, M. Talaei, and F. B. Hu, "Relation of smoking with total mortality and cardiovascular events among patients with diabetes: A meta-analysis and systematic review," *Circulation*, vol. 136, pp. 1795–804, Aug. 2015.
[31] J. G. Regensteiner *et al.*, "Sex differences in the cardiovascular consequences of diabetes mellitus," *Circulation*, vol. 132, pp. 2424–2447, Sep. 2015.
[32] N. M. Kaplan, "Cardiovascular risks of hypertension," Mar. 2016. [Online]. Available: http://www.uptodate.com/contents/cardiovascular-risks-of-hypertension
[33] S. V. Ittaman, J. J. VanWormer, and S. H. Rezkalla, "The role of aspirin in the prevention of cardiovascular disease," *Clin. Med. Res.*, vol. 12, pp. 147–154, Dec. 2014.
[34] E. Onge, C. A. Motycka, and S. A. Miller, "A review of cardiovascular risks associated with medications used to treat type-2 diabetes mellitus," *Pharmacy Therapeutics*, vol. 34, pp. 368–378, Jul. 2009.
[35] J. R. Law *et al.*, "Association of parental history of diabetes with cardiovascular disease risk factors in children with type 2 diabetes," *J. Diabetes Complications*, vol. 29, pp. 534–539, Feb. 2015.
[36] P. Chan and S. Stolfo, "Toward scalable learning with non-uniform class and cost distribution: A case study in credit card fraud detection," in *Proc. 4th Int. Conf. Knowl. Discovery Data Mining*, 1998, pp. 164–168.
[37] N. Vuong, Y. Zhao, and X. Li. *Classification Using Improved Hybrid Wavelet Neural Networks*, vol. 5351. Berlin, Germany: Springer-Verlag, 2008, pp. 1092–1097.
[38] T. Vogl, A. Mangis, W. Zigler, "Accelerating the convergence of the back-propagation method," *Biological Cybern.*, vol. 59, pp. 257–263, 1998.
[39] K. Zarkogianni, E. Litsa, A. Vazeou, and K. S. Nikita, "Personalized glucose-insulin metabolism model based on self-organizing maps for patients with type 1 diabetes mellitus," in *Proc. IEEE 13th Int. Conf. Bioinformat. Bioeng.* , 2013 pp. 1–4.
[40] K. Zarkogianni *et al.*, "A review of emerging technologies for the management of diabetes mellitus," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 12, pp. 2735–2749, Dec. 2015.
[41] L. G. M. Souza and G. A. Barreto "Multiple local ARX modeling for system identification using the self-organizing map," in *Proc. II Eur. Symp. Time Series Prediction*, 2008, pp. 215–224.
[42] K. Zarkogianni *et al.*, "Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring," *Med. Biol. Eng. Comput.*, vol. 53, no. 12, pp. 1333–1343, 2015.
[43] R. Valdovinos, J. Sánchez, and R. Barandela, "Dynamic and static weighting in classifier fusion," *Pattern Recognit. Image Anal*, vol. 3523, pp. 59–66, 2005.
[44] D. Jimenez, "Dynamically weighted ensemble neural networks for classification," in *Proc. IEEE World Congr. Comput. Intell./IEEE Int. Joint Neural Netw.*, Anchorage, AK, USA, May 1998, pp. 753–756.
[45] V. Almendra and D. Enächescu, "Using self-organizing maps for binary classification with highly imbalanced datasets," *Int. J. Numerical Anal. Modeling*., vol. 5, no. 3, pp. 238–254, 2014.
[46] Z. M. Zin, R. Yusof, and E. Mesbahi, "Improving self organizing maps method for data clustering and classification," presented at the 10th Asian Control Conf., Kota Kinabalu, Malaysia, 2015.
[47] Z. Reitermanova, "Data splitting," *Proc. Contributed Papers*, vol. 10, pp. 31–36, 2010.
[48] G. Collins, S. Mallett, O. Omar, and L. Yu, "Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting," *BMC Med.*, vol. 9, pp. 1–14, 2011.
[49] E. W. Steyerberg *et al.*, "Assessing the performance of prediction models: A framework for some traditional and novel measures," *Epidemiology*, vol. 21, pp. 128–138, 2010.
[50] A. J. Vickers and E. B. Elkin, "Decision curve analysis: A novel method for evaluating prediction models," *Med. Decis. Making*, vol. 26, pp. 565–574, 2006.
[51] S. Dieren *et al.*, "Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: A systematic review," *Heart*, vol. 98, pp. 360–369, 2012.
[52] A. Al-Rawahi and P. Lee, "Applicability of the existing CVD risk assessment tools to type II diabetics in Oman: A review," *Oman Med. J.*, vol. 30, no. 5, pp. 315–319, 2015.
[53] K. Dalakleidi, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, "Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications," *Expert Systems*, 2017. [Online]. Available: https://doi.org/10.1111/exsy.12214