Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Prediction of Nocturnal Hypoglycemia by an aggregation of previously known prediction approaches: Proof of concept for clinical application

P. Tkachenko, G. Kriukova, M. Aleksandrova, O. Chertov, E. Renard, S. Pereverzyev

RICAM-Report 2016-06

Prediction of Nocturnal Hypoglycemia by an aggregation of previously known prediction approaches: Proof of concept for clinical application

Pavlo Tkachenko*[a], Galyna Kriukova[a], Marharyta Aleksandrova[b,c], Oleg Chertov[b], Eric Renard[d,e], Sergei V. Pereverzyev[a]

[a]*Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, 4040 Linz, Austria. Phone: +43 732 2468 5214. Fax: +43 732 2468 5212. pavlo.tkachenko@oeaw.ac.at*

[b]*National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine*

[c]*Université de Lorraine - LORIA, Vandoeuvre les Nancy, France*

[d]*Department of Endocrinology, Diabetes, Nutrition, and CIC INSERM 1411, Montpellier University Hospital, Montpellier, France.*

[e]*Institute of Functional Genomics, UMR CNRS 5203/INSERM U1191, University of Montpellier, Montpellier, France*

Abstract

Background and Objective: Nocturnal hypoglycemia (NH) is common in patients with insulin-treated diabetes. Despite the risk associated with NH, there are only a few methods aiming at the prediction of such events based on intermittent blood glucose monitoring data and none has been validated for clinical use. Here we propose a method of combining several predictors into a new one that will perform at the level of the best involved one, or even outperform all individual candidates.

Methods: The idea of the method is to use a recently developed strategy for aggregating ranking algorithms. The method has been calibrated and tested on data extracted from clinical trials, performed in the European FP7-funded project DIAdvisor. Then we have tested the proposed approach on other datasets to show the portability of the method. This feature of the method allows its simple implementation in the form of a diabetic smartphone app.

Results: On the considered datasets the proposed approach exhibits good performance in terms of sensitivity, specificity and predictive values. Moreover, the resulting predictor automatically performs at the level of the best involved method or even outperforms it.

Conclusion: We propose a strategy for a combination of NH predictors that leads to a method exhibiting a reliable performance and the potential for everyday use by any patient who performs self-monitoring of blood glucose.

Keywords: Prediction of Nocturnal Hypoglycemia, Type 1 Diabetes, Aggregation, Last Before Bed Measurement, LBGI.

---

1. Introduction

According to a report of a Workgroup of the American Diabetes Association and the Endocrine Society [1] hypoglycemia is defined as blood glucose (BG) level less than 70 mg/dl. Nocturnal hypoglycemia (NH) is the most feared type of hypoglycemia in patients with diabetes treated by insulin. Due to its time of occurrence it is usually asymptomatic [2, 3] but has negative impact on patients health. NH problem is less worrisome for the patients equipped with Continuous Glucose Monitors (CGM), but only about 2-3% of insulin-treated patients use such systems because of their high market price, frequent annoying false alarms and the lag behind actual glucose measures which impairs patients trust.

On the other hand, intermittent monitoring performed from finger sticks remains the most widely used blood glucose monitoring method (BGM). The main advantage of this method is that it provides fairly accurate results of BG concentration. Additionally, this type of BGM is marketed at very low prices compared to noninvasive systems or CGM. Therefore, it is attractive to develop a method for predicting NH which uses only limited discrete information on blood glucose level during daytime hours. The first attempt in this direction was done by Whincup and Milner [4] in 1987. They proposed a classification method which uses only one before-bed measurement for the prediction. After testing 6 threshold values (from 90 mg/dl to 180 mg/dl) they have recommended to use the values below the threshold of 126 mg/dl as announcing NH during the forthcoming night, i.e. if the bedtime BG concentration is lower than 126 mg/dl then one expects NH to occur, whereas one does not expect NH in case of higher bedtime value. However, this method has been criticized, for example by Davies [5] , because of its poor performance on other datasets. Moreover, the clinical tests of Davies [5] show that the threshold value of 126 mg/dl gives one of the worst predictions, and the best among used threshold-based predictors is the one using the threshold value of 90 mg/dl.

After these attempts of NH prediction, a diabetes advisory system (DIAS) based on causal probabilistic network was proposed in [6] as a tool to identify periods of unrecognized NH. In addition to BG concentration, DIAS handled data on insulin dose and carbohydrate intake to provide an indication of BG values between home blood tests. However, it should not be assumed that NH would always occur at the time predicted by DIAS. For example, in five of the six patients in whom NH was predicted by DIAS for four consecutive nights, the tests [6] confirmed hypoglycemia only on one of the nights.

Another method which aims at the prediction of severe hypoglycemia, and can be potentially also used for prediction of NH, is based on the low blood glucose index (LBGI) [7, 8]. The value of LBGI index cumulates all daily measurements of blood glucose and thus contains more information than the classifier of Whincup and Milner [4]. Moreover, LBGI can classify NH risk into more than 2 values reflecting true glycaemic control, where different risks of NH are relevant.

However, from the definition of the LBGI it follows that hypoglycemia can be predicted only in cases when a patient had numerous mild low BGs, a few extreme low BGs, or a mixture of both [7]. In other words, if, for example, all daily BG measurements of a patient were above 95 mg/dl, the resulting LBGI will be smaller than 1 indicating minimal risk of NH. Despite that, cases do occur when a patient did not have low BG-measurements during daytime and was affected by a NH during sleep hours. As a result, LBGI cannot be used as a predictor for such cases, but, interestingly, some of them can be caught by the classifiers of Whincup and Milner [4], which correspond to high threshold values (e.g., 180 mg/dl).

In the current study we present a general approach based upon combining several NH predictors that automatically work at the level of the best involved predictor, and may even outperform all of them. The proposed approach has been recently advocated in [9] in the context of ranking, which is relatively new problem of machine learning. Note that the use of ranking framework is natural for NH prediction, because, for example, Accu-Chek Connect [10] suggests the use of LBGI values for ranking NH risks into 4 categories: minimal, low, medium and high. Our approach requires solving a low-dimensional system of linear equations, and can be potentially implemented in on-line mode. The approach has been realized in the form of an app for Android smartphones, tested on clinical datasets and exhibited a secure level of predictive accuracy.

At the end of the introduction we want to point out that there is a vast

literature on applications of machine learning for predicting hypoglycemia. Some of the most recent publications are, for example, [11, 12, 13]. In addition to BG concentration, the inputs of machine learning algorithms may include also physical activity, medication use and nutritional data, which need to be manually inserted. On the other hand, as it has been mentioned in [14], some experts believe that the usability of the diabetes smartphone apps, for example, can be improved by eradicating needs for manual entries. Therefore, the main BG meters manufacturers follow the current digital trends and produce new devices equipped with Bluetooth technology, such that BG measurements of a patient can be immediately transferred to a patient's smartphone without a need of manual input. In view of this the algorithms enabling NH prediction with BG data alone become attractive. However, as it is reported in the literature [15, 12] such algorithms usually have a low specificity (below 70 %) that means that they may predict most hypoglycemia but produce many false positives (alarms) and not to be able to deliver meaningful interventions to patients. Our approach promises a remedy for this issue, because it combines individual NH predictors and automatically follows the one currently performing better than others. As a result, a good balance between specificity (above 80%) and sensitivity (approximately 70%) has been observed in all tests with different clinical datasets.

In this paper, for the purpose of results comparison, we use the same performance metrics as in the study on NH prediction [4] and in further publications [15, 12]. These metrics count the numbers or percentage of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) predictions. In the present context TP and TN mean the cases when NH appearance or absence, respectively, was correctly predicted. FP means that NH was predicted, but did not occur and FN means the opposite scenario.

Sensitivity SE=TP/(TP+FN) and Specificity SP=TN/(TN+FP) measure respectively the proportion of positives and negatives that were correctly identified as such.

The positive and negative predictive values, PPV=TP/(TP+FP) and NPV=TN/(TN+FN), respectively, are the proportions of positive and negative predictions, which were true.

For measuring the accuracy in the above mentioned terms the so-called f-scores are also used, since they represent a weighted average of the precision (PPV) and recall (SE). In the present study we use the traditional f-measure or balanced f-score (f1 score =2TP/(2TP+FN+FP)) that is the harmonic mean of precision and recall. Additionally, we measure the f2 score

=5TP/(5TP+4FN+FP), which weighs recall higher than precision. The importance of the latter one can be explained by the patients desire to be sure that the predictions of no hypoglycemia will be correct (SP is more important).

2. Materials and methods

2.1. NH prediction as a ranking problem

Similar to [6], the appearance of NH can be represented by a stochastic model operating in a discrete space of a finite number of NH risk levels y. For example, the value $y = 1, 0.5, -0.5, -1$ may mean respectively high, moderate, low and minimal NH risk levels.

Let $x = (x^1, x^2, x^3, \ldots, x^l) \in \mathbb{R}^l$ be a vector of daily BG measurements, where, for instance, x^l is the last before-bed (LBB) measurement that is used as input in NH predictors [4, 5]. Then as in [6] the relation between x and the value y for the night succeeding the day with BG measurements x is specified using conditional probabillity $\rho(y|x)$ of y given x.

As it is explained in the Introduction, we deliberately restrict ourselves to stochastic models including only two parameters y and x. Then let $\rho(x)$ be a marginal probability distribution for the remaining model variable x.

In spite of the assumption of a stochastic relationship between x and y, we are interested in synthesizing a deterministic predictor p that will assign NH risk levels $p(x)$ to the night succeeding the day with BG measurements, forming vector x. Note that the value $p(x)$ can be also used to predict whether or not NH appears.

For given true NH risk levels y and y', which correspond to daily BG measurements x and x', the value

$$(y - y' - (p(x) - p(x')))^2$$

is interpreted by a standard assessment methodology of machine learning [16, 17] as the loss of the predictor p in its risk ranking. Then the quality of a predictor p can be measured by the expected misranking error

$$\mathcal{E}(p) = \int\int (y - y' - (p(x) - p(x')))^2 d\rho(y'|x') d\rho(y|x) d\rho(x') d\rho(x),$$

and it is natural to minimize this in the space $L_{2,\rho}$ of all functions $p(x)$, which are square-integrable with respect to the marginal dictribution $\rho(x)$.

It is known that one of the expected error minimizers can be written as

$$p_\rho(x) = \int y d\rho(y|x) - \int\int y d\rho(y|x) d\rho(x),$$

but this ideal predictor cannot be used in practice, because neither the conditional probability $\rho(y|x)$ nor the marginal distribution $\rho(x)$ is known.

On the other hand, we can access clinical records of diabetic patients, which contain the historical data, such as daily BG measurements $x_j = (x_j^1, x_j^2, \ldots, x_j^l), j = 1, 2, \ldots, n$ collected within n different days, and retrospectively estimated NH risk levels y_j for the corresponding succeeding nights. In the simple situation, which we will deal later on this paper, the real case of NH in the night after the day j is coded as $y_j = 1$, while the night without NH corresponds to $y_j = -1$.

The set of pairs $Z_n = \{(x_j, y_j), j = 1, 2, \ldots, n\}$ will appear further under the name of training set. Assuming a stochastic relationship between x and y governed by an unknown probability distribution $\rho(x, y) = \rho(y|x)\rho(x)$, it is natural to think that a training set Z_n is independently drawn from $\rho(x, y)$, and this is the only information available to approximate the ideal predictor p_ρ. Therefore, in machine learning training datasets Z_n are used to construct prediction models $p(x)$ and access their performance.

2.2. Datasets

In the current study we use datasets *DIAdvisor* and *ChildrenData*. Both datasets contain BG measurements of patients with type 1 diabetes. Further details are given below.

DIAdvisor dataset containing the data of 34 patients with diabetes was collected within the framework of the European FP7-funded project DIAdvisor. The considered subjects have been treated with insulin for at least 12 months before data collection; their ages were between 18 and 65 years, with a BMI<35 kg/m^2. During the study, the CGM-values were sampled every 5-10 min using various CGM-sensors. These data were not used for a prediction, but allow a detection of NH during the nights. At the same time, 4 true BG measurements were performed daily in parallel with CGM estimations, and these BG data were used for training and testing the considered NH predictors. A total number of n=150 days has been chosen to test the proposed approach. NH has been observed in 40 cases (26.67%).

ChildrenData dataset was collected during three and a half months in children hospitals of the Kyiv city, Ukraine, according to a protocol that

is similar to [4]. The access to this dataset has been provided within the framework of the European Horizon 2020-funded MSC-project AMMODIT. The dataset contains information about 179 children. Each of n=476 records of this dataset contains 9 BG measurements, which were performed at the following time points of a 24-hour cycle: 08:00, 11:30, 13:30, 16:00, 18:00, 21:00, 00:00, 03:00 and 06:00. The measurements at 00:00, 03:00 and 06:00 were used to identify the occurrence of NH, while the measurements of other time points of the 24-hour cycle form the input for NH prediction. In this dataset the number of records with NH is 222 (46.64%).

2.3. Approximation of ideal NH predictor by a linear combination of known prediction models

Let us assume for now that we are able to employ m different predictors $p_1(x), p_2(x), \ldots, p_m(x)$. For example, as we already mentioned in Introduction, 6 NH predictors, which use only LBB measurement x^l, were discussed in [4, 5]. In our notation, they can be written as

$$p_k^{LBB} = p_k^{LBB}(x^l) = \begin{cases} 1, & \text{if } x^l \leq 72 + 18k \text{ (mg/dL)} \\ -1, & \text{if } x^l > 72 + 18k \text{ (mg/dL)} \end{cases}, k = 1, 2, \ldots, 6.$$

We can mention also hypoglycemia predictors [7, 8], which are based on LBGI. Recall that

$$LBGI(x) = LBGI(x^1, x^2, \ldots, x^l) = 10l^{-1} \sum_{i=1}^{l} R(x^i),$$

where $R(x^i)$ is the so-called quadratic risk function defined as follows

$$R(x^i) = \left[(5.381 - (\ln(x^i))^{1.084})_+\right]^2,$$

and $(b)_+ = \max\{b, 0\}$. Then NH prediction for the corresponding night can be made in the following way

$$p^{LBGI}(x) = \begin{cases} 1, & \text{if } LBGI(x) \geq 1 \\ -1, & \text{if } LBGI(x) < 1 \end{cases}.$$

Moreover, combining the idea of the quadratic risk function and LBGI with the observation [4] that LBB measurement x^l is an important NH indicator, we can introduce one more NH predictor

$$p^R(x) = p^R(x^l) = \begin{cases} 1, & \text{if } R(x^l) \geq 0.1 \\ -1, & \text{if } R(x^l) < 0.1 \end{cases}.$$

Such predictor can be seen as a threshold-independent analog of the predictors [4] in terms of LBGI.

Thus, having an ensemble of NH predictors $p_k(x), k = 1, 2, \ldots, m$ and keeping in mind that the ideal predictor $p_\rho(x)$ belongs to the Hilbert space $L_{2,\rho}$, it is natural to look for the optimal linear combination

$$p^{opt}(x) = \sum_{k=1}^{m} c_k^{opt} p_k(x)$$

such that

$$\left\| p_\rho - p^{opt} \right\|_{L_{2,\rho}} = \min_{c_k} \left\| p_\rho - \sum_{k=1}^{m} c_k p_k \right\|_{L_{2,\rho}}. \tag{1}$$

It is clear that the vector $c = c^{opt} = (c_1^{opt}, c_2^{opt}, \ldots, c_m^{opt})$ of the optimal coefficients of the linear combination p^{opt} solves the system of linear equations $Gc = g$ with the Gram matrix $G = (\langle p_k, p_\nu \rangle_{L_{2,\rho}})_{k,\nu=1}^m$ and the right-hand side vector $g = (\langle p_k, p_\rho \rangle_{L_{2,\rho}})_{k=1}^m$, where $\langle \cdot, \cdot \rangle_{L_{2,\rho}}$ is the standard inner product in $L_{2,\rho}$, but neither G nor g is accessible, since the ideal predictor p_ρ and the marginal distribution $\rho(x)$ are unknown.

At the same time, from [9] (see Proposition 9) it follows that under rather mild assumptions on ρ and p_k with high probability we have

$$\begin{aligned} \langle p_k, p_\rho \rangle_{L_{2,\rho}} &= n^{-2} \langle \mathbb{D}Y, S_{Z_n} p_k \rangle_{\mathbb{R}^n} + O(n^{-1/2}), \\ \langle p_k, p_\nu \rangle_{L_{2,\rho}} &= n^{-1} \langle S_{Z_n} p_k, S_{Z_n} p_\nu \rangle_{\mathbb{R}^n} + O(n^{-1/2}), \end{aligned} \tag{2}$$

where $Z_n = \{(x_i, y_i), i = 1, 2, \ldots, n\}$ is a training set, $Y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ denotes the standard inner product in n-dimensional Euclidean space $\mathbb{R}^n$, $\mathbb{D} = n\mathbb{I} - \mathbb{1} \times \mathbb{1}^T$, $\mathbb{I}$, $\mathbb{1}$ are the n-th order unit matrix and the vector of all ones, and S_{Z_n} is the so-called sampling operator

$$S_{Z_n} p_k = (p_k(x_i))_{i=1}^n \in \mathbb{R}^n.$$

Thus, (2) tells us that the quantities

$$\tilde{G}_{k,\nu} = n^{-1} \langle S_{Z_n} p_k, S_{Z_n} p_\nu \rangle_{\mathbb{R}^n} , \ \tilde{g}_k = n^{-2} \langle \mathbb{D}Y, S_{Z_n} p_k \rangle_{\mathbb{R}^n} ,$$

that can be easily calculated with the use of a dataset Z_n, approximate inaccessible values $\langle p_k, p_\nu \rangle_{L_{2,\rho}}, \langle p_k, p_\rho \rangle_{L_{2,\rho}}$ with an accuracy of order $O(n^{-1/2})$.

Consider now the linear system

$$\tilde{G}c = \tilde{g} \tag{3}$$

with the matrix $\tilde{G} = (\tilde{G}_{k,\nu})_{k,\nu=1}^m$ and the vector $\tilde{g} = (\tilde{g}_k)_{k=1}^m$. Then from [9] (see Theorem 10) it follows that with high probability for sufficiently large n the solution $c = \tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_m)$ of (3) exists and gives rise to the predictor

$$p^{ag}(x) = \sum_{k=1}^{m} \tilde{c}_k p_k(x) \tag{4}$$

aggregating the given ensemble $\{p_k\}$ such that

$$\|p_\rho - p^{ag}\|_{L_{2,\rho}} = \left\|p_\rho - p^{opt}\right\|_{L_{2,\rho}} + O(n^{-1/2}).$$

The latter means that, up to a quantity decreasing with the size of the training dataset Z_n, the effectively calculated aggregator p^{ag} is as good as the optimal p^{opt}.

3. Performance evaluation

Note that the predictors $p_k^{LBB}, k = 1, 2, \ldots, 6, p^{LBGI}, p^R$ mentioned above take the values -1 or 1. At the same time, the predictor (4) aggregating them may take other values as well. If one is going to use the aggregator (4) as a classifier predicting the "yes" or "no" answer, labeled respectively by 1 and -1, then the value $p^{ag}(x)$ can be interpreted as follows

$$\bar{p}^{ag}(x) = \begin{cases} 1, & \text{if } p^{ag}(x) \geq 0.5 \\ -1, & \text{if } p^{ag}(x) < 0.5 \end{cases} . \tag{5}$$

On the other hand, as it was mentioned in Introduction, in accordance with [10], the predictors p^{LBGI}, p^R, which are based on the quadratic risk function and LBGI, can be used for ranking NH risks into 4 categories: minimal, low, moderate, and high, labeled respectively by -1, -0.5, 0.5, 1. The

Table 1: The average performance (in percent) of NH predictors $p_k^{LBB}, k = 1, 2, \ldots, 6, p^{LBGI}, p^R$, and their aggregators $\bar{p}^{ag}, \bar{p}^{ag}_{LBB}$ on training sets (75 days) and testing sets (75 days) from *DIAdvisor* data

	Training set						Testing set					
NH predictor	SE	SP	PPV	NPV	f1	f2	SE	SP	PPV	NPV	f1	f2
p_1^{LBB}	49	99	95	84	64	55	51	99	95	85	66	55
p_2^{LBB}	70	91	75	89	72	71	70	92	76	90	73	71
p_3^{LBB}	80	71	50	91	62	71	80	71	50	91	61	71
p_4^{LBB}	85	53	40	90	54	68	85	53	39	91	54	69
p_5^{LBB}	98	38	37	98	53	72	97	38	36	98	53	73
p_6^{LBB}	98	31	34	97	50	70	97	31	34	97	50	71
p^{LBGI}	65	95	84	88	73	68	66	96	84	88	73	68
p^R	54	97	88	85	67	59	56	97	88	86	68	60
$\bar{p}^{ag}$	**82**	**93**	**81**	**93**	**81**	**82**	**79**	**92**	**79**	**93**	**79**	**79**
$\bar{p}^{ag}_{LBB}$	77	86	69	91	72	75	73	84	66	90	68	70

aggregator (4) can also be used for doing this. In this case the value $p^{ag}(x)$ is interpreted as minimal, low, moderate or high NH risk depending on an interval in which it falls. The intervals are defined by the above mentioned labels of risk ranks.

At first we evaluate the performance of $\bar{p}^{ag}$ for the case when the sets Z_n appearing in the formulas (2) are randomly taken from *DIAdvisor* dataset. Recall that the latter one contains 150 clinical records. In our first experiment we take Z_n with $n = 75$. Then the remaining 75 records (daily BG measurements and CGM traces) are used for testing $\bar{p}^{ag}$ against the performance metrics described in Introduction. Such random procedure is repeated 200 times.

This means that for each of 200 random simulations one needs to find a vector $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_8)$ from the system (3). Then (4) aggregates the predictors $p_k^{LBB}, k = 1, 2, \ldots, 6, p^{LBGI}, p^R$, and the corresponding predictor (5) denoted as $\bar{p}^{ag}$ is tested on the clinical data that have been not included in Z_n. The average values of the considered performance metrics over 200 random simulations are reported in Table 1.

From this table one can see that the predictor p_3^{LBB} suggested in [4] has a rather moderate SP and PPV. Moreover, the predictor p_1^{LBB}, suggested in [5], and the predictor p^R have low SE and f2-score. At the same time, their

Table 2: The performance (in percent) of the aggregator constructed with the use of *DIAdvisor* dataset on *ChildrenData*

	Performance					
Dataset	SE	SP	PPV	NPV	f1	f2
ChildrenData	73.4	87.8	84.0	79.0	78.4	75.3

aggregator $\bar{p}^{ag}$ performs well with respect to all of the considered metrics. Note also that in the considered tests the predictors p^{LBGI} and p^{R} exhibit a rather moderate performance compared to p_2^{LBB}, p_3^{LBB}. At the same time, if we exclude them from the aggregation, then, as it can be seen from last row of Table 1, the performance of the corresponding aggregator $\bar{p}^{ag}_{LBB}$ become worse than the one of $\bar{p}^{ag}$. This observation suggests us to aggregate all available NH predictors.

At the end of Introduction we have mentioned an implementation in the form of a smartphone app as a possible application of NH prediction algorithms. In view of such an application, a desirable feature of NH predictors would be their portability from individual to individual without readjustment. This means that an algorithm, which was constructed with the use of clinical data of one group of patients and implemented as a smartphone app, can be downloaded by other patients and used without recalibration and essential loss of prediction performance.

To demonstrate that the proposed aggregation of NH predictors allows the above mentioned portability, we construct the aggregator $\bar{p}^{ag}$ according to (2)–(5), where in (2) $p_k = p_k^{LBB}, k = 1, 2, \ldots, 6, p_7 = p^{LBGI}, p_8 = p^{R}$, and $Z_n, n = 150$, is chosen to be the whole *DIAdvisor* dataset. Then the constructed NH predictor (5) is applied without any adjustment to *ChildrenData* described above. Table 2 displays the corresponding evaluation results.

Table 2 shows a good balance between specificity and sensitivity of the integrator that has been demonstrated without any readjustment to previously unseen prediction inputs and can be considered as the evidence of the above mentioned portability.

4. Discussion

We have described an approach to the aggregation of several NH predictors. The approach is based on a recently developed strategy for aggregat-

ing ranking algorithms [9]. In the present paper we have demonstrated the proposed approach by aggregating NH predictors known from the literature [4, 5, 7, 8], and observed that the aggregators exhibit better prediction performance than the predictors used in the aggregation procedure (see Table 1). Note that the same approach can be used for aggregating NH predictors, which are not available yet, but which may be developed in future, and we expect that the observed effect of the improvement of prediction performance will also be demonstrated by the aggregators constructed with the use of new NH predictors.

In Section 3 we have mentioned that the aggregator constructed with the use of NH predictors p^{LBGI}, p^R, can be used for ranking NH risks into 4 categories. Of course, for constructing such an aggregator one needs a training set $Z_n = \{(x_i, y_i), i = 1, 2, \ldots, n\}$, where y_i are the true NH risk ranks taking the values $\pm 0.5, \pm 1$. *DIAdvisor* dataset containing CGM values sampled every 5-10 minutes also during night-time allows a retrospective NH risk ranking. For example, high NH risk ranks $y_i = 1$ can be assigned to the nights when CGM values below 30 mg/dL were observed. The duration of time intervals with CGM values below 70 mg/dL can be also taken into account in assigning NH risk ranks y_i.

As in our first experiment, the training set $Z_n, n = 75$, with the assigned NH risk ranks y_i have been randomly taken from *DIAdvisor* dataset and used in formulas (2), where $k = 1, 2$ and $p_1 = p^{LBGI}, p_2 = p^R$. The remaining clinical records have been used for the performance testing.

A natural metric for measuring the performance of algorithms p predicting NH risk ranks $p(x)$ is the fraction of misranked pairs in a testing set Z:

$$mis(p, Z) = \frac{\sum_{(x_i,y_i),(x_j,y_j)\in Z} 1_{\{y_i > y_j \wedge p(x_i) < p(x_j)\}}}{\sum_{(x_i,y_i),(x_j,y_j)\in Z} 1_{\{y_i > y_j\}}},$$

where $1_{\{s\}}$ is the indicator function of s. The avarage values of this metric over 200 random simulations for NH predictors p^{LBGI}, p^R and their aggregator p^{ag} are reported in Table 3.

Table 3 also allows a conclusion that the proposed aggregation approach improves the prediction performance.

In our experiments we have observed an important feature of the aggregated predictor p^{ag}. Namely, a portability from individual to individual without a necessity of recalibration and such that no essential loss appears in the prediction performance. This feature allows a simple implementation

Table 3: Performance of the predictors of NH risk ranks p^{LBGI}, p^R and their aggregator on *DIAdvisor* dataset

NH predictor p	$mis(p, Z)$
p^{LBGI}	0.4452
p^R	0.5755
p^{ag}	0.3884

of our prediction algorithm in the form of a smartphone app. A prototype of such diabetic smartphone has been developed for Android smartphones. Its interface displaying a paarticular NH prediction can be seen in Figure 1.

This application has been tested without retuning on *Single-patient's data* provided by an interested 42-year old volunteer, who is a patient with type 1 diabetes since his childhood. This dataset consists of 182 measurement vectors x_j sampled similar to *ChildrenData*. NH was attributed to 28.6% of measurement vectors and detected by self-assessment with the use, in particular, fasting BG levels at 7 AM suggested in [18] as a good indicator for NH in the previous night.

The aggregated predictor p^{ag} implemented in the smartphone app has been directly applied to this dataset and exhibited the following performance: SE=69.2%, SP=85.3%, PPV=65.4%, NPV=87.4%, f1=67.2%, f2=68.4%.

It is instructive to compare these values with the performance of NH-detecting device HypoMon® that has been highlighted [19] as the world's first non-invasive alarm system that identifies sleep-time hypoglycemia. In a special clinical trial [15] HypoMon® performance was reported as SE=73%, SP=68%, PPV=38%, NPV=90%, that corresponds to f1=49% and f2=62%.

We would like to stress that HypoMon® was designed not to predict NH, but to alarm when NH has been already occurred, which seems to be easier than NH prediction ("it is difficult to make predictions especially about the future"). Nevertheless, the comparison of all the above mentioned values of the performance metrics is in favor of our prediction approach.

Note that in contrast to *DIAdvisor* the other two testing datasets contain BG measurements at discrete time moments only. Therefore, a validation of hypoglycemia cases on these datasets has been performed similar to [4, 5] by examining BG measurements collected during the night period. Of course, in this way some asymptomatic nocturnal hypos may be missed. Therefore,

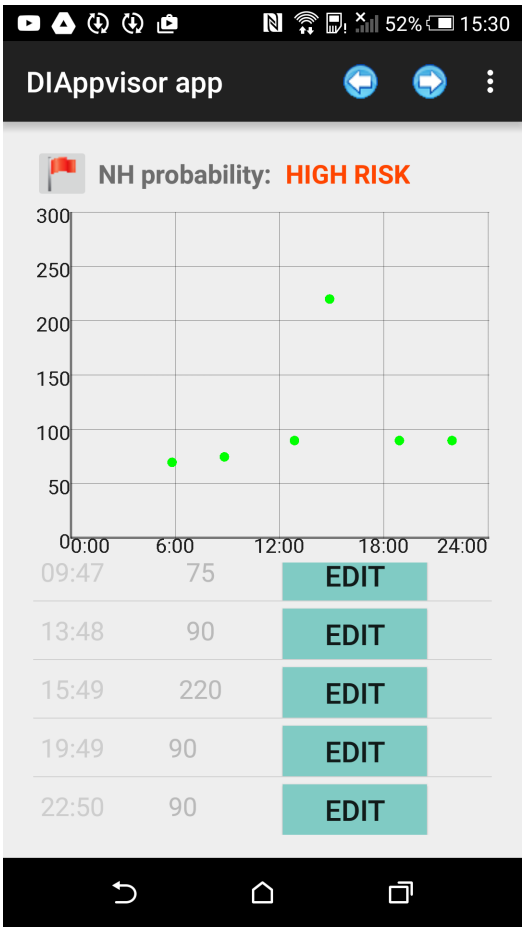

Figure 1: Screenshot of a diabetic smartphone app

the results reported above should be considered as a proof of concept only.

Acknowledgment

This research was supported by AMMODIT project (Approximation Methods for Molecular Modelling and Diagnosis Tools) in the frame of Horizon2020 programme. The authors affiliated with Johann Radon Institute gratefully acknowledge the support of the Austrian Science Fund (FWF): project P25424.

We thank Artem Symchuk, MD, Research Scientist and Junior Physician in PL Shupyk National Medical Academy of Postgraduate Education (Kyiv, Ukraine) for his activity related with *ChildrenData* dataset. We are also grateful to Dirk Nuyens, PhD, KU Leuven, Belgium, for valuable discussion and volunteering.

The interface of smartphone app prototype shown in Figure 1 has been designed by Lucian Nita (RomSoft, Iasi, Romania).

Conflict of interest statement

No competing financial interests exist.

References

[1] E. R. Seaquist, J. Anderson, B. Childs, P. Cryer, S. Dagogo-Jack, L. Fish, S. R. Heller, H. Rodriguez, J. Rosenzweig, R. Vigersky, Hypoglycemia and diabetes: A report of a workgroup of the american diabetes association and the endocrine society, Diabetes Care 36 (5) (2013) 1384–1395. arXiv:http://care.diabetesjournals.org/content/36/5/1384.full.pdf+html, doi:10.2337/dc12-2480.
URL http://care.diabetesjournals.org/content/36/5/1384.abstract

[2] G. Vervoort, H. Goldschmidt, L. van Doorn, Nocturnal blood glucose profiles in patients with type 1 diabetes mellitus on multiple (4) daily insulin injection regimens, Diabetic Medicine 13 (9) (1996) 794–799. doi:10.1002/(SICI)1096-9136(199609)13:9<794::AID-DIA185>3.0.CO;2-G.
URL http://dx.doi.org/10.1002/(SICI)1096-9136(199609)13:9<794::AID-DIA185>3.0.CO;2-G

[3] M. Kalergis, K. Aljaberi, A. Schiffrin, S. Meltzer, R. Gougeon, J.-F. Yale, Frequency and duration of nocturnal hypoglycemia in adults with type 1 diabetes undergoing intensive management as determined by continuous glucose monitoring (abstract), Can J Diabetes Care 25 (156A).

[4] G. Whincup, R. Milner, Prediction and management of nocturnal hypoglycaemia in diabetes, Archives of Disease in Childhood 62 (4) (1987) 333–337.

[5] A. Davies, Prediction and management of nocturnal hypoglycaemia in diabetes, Archives of Disease in Childhood 62 (10) (1987) 1085.

[6] D. Cavan, R. Hovorka, O. Hejlesen, S. Andreassen, P. Snksen, Use of the {DIAS} model to predict unrecognised hypoglycaemia in patients with insulin-dependent diabetes, Computer Methods and Programs in Biomedicine 50 (3) (1996) 241 – 246, computers in Diabetes. doi:http://dx.doi.org/10.1016/0169-2607(96)01753-1.
URL http://www.sciencedirect.com/science/article/pii/0169260796017531

[7] B. P. Kovatchev, D. J. Cox, L. A. Gonder-Frederick, D. Young-Hyman, D. Schlundt, W. Clarke, Assessment of risk for severe hypoglycemia among adults with iddm: validation of the low blood glucose index., Diabetes Care 21 (11) (1998) 1870–1875. arXiv:http://care.diabetesjournals.org/content/21/11/1870.full.pdf+html, doi:10.2337/diacare.21.11.1870.
URL http://care.diabetesjournals.org/content/21/11/1870.abstract

[8] D. J. Cox, L. Gonder-Frederick, L. Ritterband, W. Clarke, B. P. Kovatchev, Prediction of severe hypoglycemia, Diabetes Care 30 (6) (2007) 1370–1373. arXiv:http://care.diabetesjournals.org/content/30/6/1370.full.pdf+html, doi:10.2337/dc06-1386.
URL http://care.diabetesjournals.org/content/30/6/1370.abstract

[9] G. Kriukova, O. Panasiuk, S. V. Pereverzyev, P. Tkachenko, A linear functional strategy for regularized ranking, Neural Networks 73 (2016) 26 – 35. doi:http://dx.doi.org/10.1016/j.neunet.2015.08.012.

URL http://www.sciencedirect.com/science/article/pii/S0893608015001756

[10] Blood glucose index (bgi), http://hcp.accu-chek.co.uk/gbconnect/blood-glucose-index.html, accessed: 2016-15-02.

[11] M. Yamaguchi, C. Kaseda, K. Yamazaki, M. Kobayashi, Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining, Medical and Biological Engineering and Computing 44 (6) (2006) 451–457. doi:10.1007/s11517-006-0049-x.
URL http://dx.doi.org/10.1007/s11517-006-0049-x

[12] B. Sudharsan, M. Peeples, M. Shomali, Hypoglycemia prediction using machine learning models for patients with type 2 diabetes, Journal of Diabetes Science and Technology 9 (1) (2015) 86–90. arXiv:http://dst.sagepub.com/content/9/1/86.full.pdf+html, doi:10.1177/1932296814554260.
URL http://dst.sagepub.com/content/9/1/86.abstract

[13] E. I. Georga, V. C. Protopappas, D. Polyzos, D. I. Fotiadis, Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models, Medical & Biological Engineering & Computing 53 (12) (2015) 1305–1318. doi:10.1007/s11517-015-1263-1.
URL http://dx.doi.org/10.1007/s11517-015-1263-1

[14] V. Naumova, L. Nita, J. U. Poulsen, S. V. Pereverzyev, Meta-learning based blood glucose predictor for diabetic smartphone app, in: H. Kirchsteiger, B. J. Jørgensen, E. Renard, L. del Re (Eds.), Prediction Methods for Blood Glucose Concentration: Design, Use and Evaluation, Springer International Publishing, 2016, pp. 93–105. doi:10.1007/978-3-319-25913-0_6.
URL http://dx.doi.org/10.1007/978-3-319-25913-0_6

[15] V. Skladnev, N. Ghevondian, S. Tarnavskii, N. Paramalingam, T. Jones, Clinical evaluation of a noninvasive alarm system for nocturnal hypoglycemia, Journal of Diabetes Science and Technology 4 (1) (2010) 67–74.

[16] S. Agarwal, P. Niyogi, Generalization bounds for ranking algorithms via algorithmic stability, J. of Mach. Learn. Res. 10 (2009) 441–474.

[17] Y. Ying, D.-X. Zhou, Online Pairwise Learning Algorithms with Kernels, ArXiv e-printsarXiv:1502.07229v1.

[18] M. Beregszászi, N. Tubiana-Rufi, K. Benali, M. Noël, J. Bloch, P. Czernichow, Nocturnal hypoglycemia in children and adolescents with insulin-dependent diabetes mellitus: Prevalence and risk factors, The Journal of Pediatrics 131 (1) (1997) 27–33. doi:http://dx.doi.org/10.1016/S0022-3476(97)70121-5.
URL http://www.sciencedirect.com/science/article/pii/S0022347697701215

[19] HypoMon hypoglycaemia monitor, http://www.gooddesignaustralia.com/awards/past/entry/hypomon-hypoglycaemia-monitor/?year=2011, accessed: 2016-15-02.