

Received 18 October 2023, accepted 22 December 2023, date of publication 5 January 2024,
date of current version 18 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3350201

RESEARCH ARTICLE

Next-Day Prediction of Hypoglycaemic Episodes Based on the Use of a Mobile App for Diabetes Self-Management

ANASTASIOS ALEXIADIS¹, ATHANASIOS TSANAS^{2,3,4}, (Senior Member, IEEE),
LEONARD SHTIKA⁵, VASSILIS EFOPOULOS⁵, KONSTANTINOS VOTIS¹,
DIMITRIOS TZOVARAS¹, (Senior Member, IEEE), AND ANDREAS TRIANTAFYLLIDIS¹

¹Centre for Research and Technology Hellas, Information Technologies Institute (CERTH/ITI), 570 01 Thessaloniki, Greece

²Usher Institute, Edinburgh Medical School, The University of Edinburgh, EH16 4UX Edinburgh, U.K.

³School of Mathematics, The University of Edinburgh, EH8 9YL Edinburgh, U.K.

⁴The Alan Turing Institute, NW1 2DB London, U.K.

⁵Tessera Multimedia, Pefka, 57010 Thessaloniki, Greece

Corresponding author: Anastasios Alexiadis (talex@iti.gr)

ABSTRACT Hypoglycaemia is one of the most common complications in diabetes, which can be life threatening if not managed appropriately. So far, research on hypoglycaemia prediction has been scarce, focusing on small cohorts linked to specific geographical regions, thus limiting the generalizability of the findings. In this paper, we developed and validated different machine learning models for next-day hypoglycaemia prediction in type 2 diabetes. We used a large international cohort comprising 669 participants, who had been regular users (for over a couple of years) of a mobile app for diabetes self-management and used common portable commercial devices for measuring their blood glucose and blood pressure levels, collecting in total 96121 observations (from which we extracted a balanced dataset of 2998 observations). Random Forests (RF), Support Vector Machines, Adaptive Boosting and Feed-Forward Artificial Neural Networks were employed to train predictive models based on 10-day temporal sequences with blood glucose and blood pressure measurements towards estimating next day hypoglycaemic episodes. We used a leave-one-subject-out (LOSO) approach for model validation, and found that RF achieved the best accuracy (0.814) and F1-score (0.812) with sensitivity (0.805) and specificity (0.824) for next-day hypoglycaemia prediction. The results of this study provide an expedient and reliable app-based approach to accurately predict hypoglycaemia in day-to-day life, thereby facilitating patient and care provider awareness and potentially preventing other serious complications.

INDEX TERMS Hypoglycaemia, machine learning, mobile health, diabetes, self-management.

I. INTRODUCTION

Diabetes is one of the leading causes of mortality and disability in the world [1]. The global diabetes prevalence in 20-79 year olds in 2021 was estimated to be 10.5% (536.6 million people), and it is expected to rise to 12.2% (783.2 million) by 2045 [2]. The aging population along with other major risk factors such as obesity,

The associate editor coordinating the review of this manuscript and approving it for publication was Gyorgy Eigner¹.

have been drivers of diabetes higher prevalence, leading to reduced life expectancy and costly complications [3], [4], [5].

Assisting people to make changes towards a healthier lifestyle and self-manage diabetes is an important strategy to maintain a good quality of life and avoid possible complications [6]. In particular, low blood sugar level, i.e., hypoglycaemia, is one of the most common barriers in Type 1 Diabetes Mellitus (T1DM) and Type 2 Diabetes Mellitus (T2DM) to achieve tight glycaemic control, which

can be life-threatening if not treated quickly [7]. Research has shown that people with diabetes with an episode of severe hypoglycaemia during the last 5 years, had a 3.4-fold increased risk of death than those with mild or no hypoglycaemia [8]. The economic impacts of hypoglycaemia due to increased blood glucose monitoring, hospitalisations, medical contact, and absence from work, are enormous [9]. In this context, appropriate management of hypoglycaemia is of vital importance.

Mobile health (mHealth) has recently shown benefits [10] of efficient medical care for chronic diseases anytime-anywhere [11], [12]. Mobile devices such as smartphones and portable medical devices with sensing and communication capabilities could be utilized by participants and their care providers, in order to monitor patient health status continually, and thus reduce the probability of potential complications, by providing adjusted therapeutic plans and improving patient self-management or remote medical management [13], [14].

Machine learning methods, which harness data generated through mobile and sensing devices, have already shown their capability to predict disease exacerbations and health status deterioration, e.g. exacerbation in chronic obstructive pulmonary disease [15], anxiety [16] and cardiovascular risk [17]. Therefore, machine learning could be a vehicle to facilitate hypoglycaemia management, by acquiring knowledge derived from patient's data, and predicting the occurrence of hypoglycaemic episodes. Prediction of hypoglycaemic episodes could be useful for both patients and their care providers, because it may improve glycaemic control, reduce the possible fear or anxiety over facing hypoglycaemic episodes, and facilitate adherence to treatment [18], [19], [20], [21].

Related research works in hypoglycaemia prediction based on mHealth data, have been rather limited. Bertachi et al. [22], used the OhioT1DM dataset to predict hypoglycaemic episodes, based on Continuous Glucose Monitoring (CGM) data received from 6 individuals with T1DM for a limited period of 8 weeks. The same dataset (with the addition of 6 participants from the 2020 version of the OhioT1DM dataset) was also explored by Deng et al. [23], in order to test different neural network architectures. Along similar lines, Marcus et al. [24] used CGM data from 11 participants with T1DM for 50 days, to test kernel methods for prediction of hypoglycaemia. Sudharsan et al. [25], used prediction models based on Random Forests (RF), Support Vector Machines (SVM), k-nearest neighbor, and naïve Bayes, by harnessing 56K self-monitored blood glucose samples from a clinical trial with 163 T2DM participants over one year in Maryland, US. Other research studies have focused on hypoglycaemia prediction without considering mHealth data, but acquiring data from resources such as electronic health records and health insurer databases [26], [27]. Furthermore, a recent review [28] has called for the need to continue research work in the development of accurate machine learning models to predict hypoglycaemia, considering also the lack of focus



FIGURE 1. Screens to view recordings (blood glucose, blood pressure, meal, medication, physical activity) in forDiabetes mobile app.

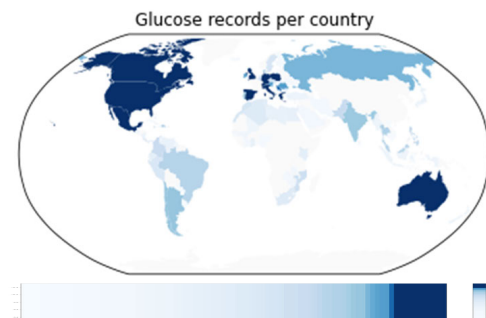


FIGURE 2. Mean of glucose records per country: 2%–20% dark blue, 1%–2% light blue, <1% white.

on T2DM and the inaccuracy of CGM in the hypoglycaemic range.

In this paper, our main objective is to develop and validate different machine learning models for next-day hypoglycaemia prediction in T2DM based on the use of a mobile app for diabetes self-management and regular consumer portable devices for measuring blood glucose and blood pressure. The ultimate aim of our work is to provide the means for accurate prediction of hypoglycaemia and the prevention of its complications, through the use of machine learning models within mHealth services provided for participants and their care providers in the real-world.

II. DATA

A. MOBILE APP FOR DIABETES SELF-MANAGEMENT

The mobile app ‘forDiabetes’¹ was developed in order to improve self-management and remote medical management of diabetes. The mobile app included several functions such as a diary for recording measurements (blood glucose, blood pressure, meals, physical activity, medication, HbA1c, etc.), goal setting and editing, measurement graphs, and exchange of data with care providers. The mobile app also allows the automated recording of blood glucose measurements measured using popular consumer glucometers such as the Contour Next ONE, Contour Plus ONE, GlucoMen areo, and Beurer GL50. The mobile app is GDPR-compliant and it has been available in Android and iOS (both free and commercial versions available) since May 2018. In addition to English it

¹<https://fordiabetes.app/>

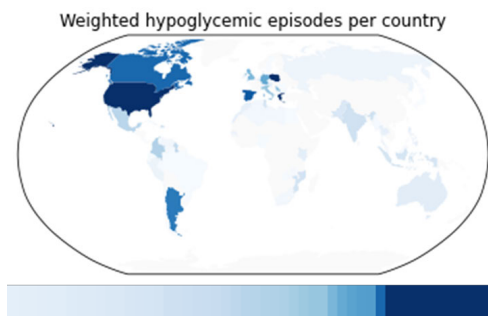


FIGURE 3. Mean of hypoglycemic incidents ($< 70\text{mg/dL}$) weighted by number of glucose measurements per country 5%–16% dark blue, 2%–5% light blue, $< 2\%$ white.

has been translated in 10 languages, and had received more than 40K downloads from users around the world until May 2022. Figure 1 presents the user interface of the forDiabetes mobile app.

B. DATA DESCRIPTION

The dataset was obtained from the database of forDiabetes and includes data from 2013 (the majority of data is from 2018) up to November of 2020. The dataset comprises:

- 808276 glucose records from 11165 different users
- 25025 hypoglycemic incidents ($< 70\text{mg/dL}$) from 1438 different users
- 180373 manual medication logs
- 5506 records of Glycated haemoglobin (HbA1c)
- 65708 records of blood pressure (systolic and diastolic)
- 27160 records of weight
- 29735 meal records (hydrocarbons and glycemic index)
- 113509 records of physical activity (type and duration, calorie burns, distance)

From these users only 998 provided data over ten-day consecutive intervals of glucose records. The mean number of ten-day consecutive intervals per patient with T2DM was 318 ± 371 , with a range from 22 (patient with minimum data points) to 3958 (patient with maximum data points). The data was extracted from the database of the forDiabetes app as a set of text files in the Comma Separated Values (CSV) format, whereas each file corresponded to a database table.

In Figure 2 we present the mean of glucose records geographically distributed, and in Figure 3 we present an overview of the weighed hypoglycemic episodes per country.

III. METHODS

A. DATA PRE-PROCESSING

We retained only the participant data which contained at least a ten-day interval of consecutive glucose data, which correspond to 998 participants that contain 317549 glucose records and 33287 blood pressure records. The selection of the ten-day interval was based on the work by Sudharsan et al. [25]. The devices supported for recording glucose measurements were the following:

“Accu-Chek Instant,” “Amazon Alexa,” “Apple Health,” “Beurer GL50 evo (Bluetooth),” “Contour Next ONE,” “Contour Plus ONE,” “ControlBios TD 4277,” “GlucoMen areo,” “GlucoMen areo 2K,” “forDiabetes.” The source “forDiabetes” corresponds to the user manually entering the measurement to the mobile app, whereas the other options correspond to the automatic insertion of the records by a specific device. No records were found in the data of the following 3 supported sources: Google Fit, Fitbit and the Beurer GL50 evo (NFC). Regarding blood pressure measurements 98% were entered manually, while 2% was automatically integrated in the participants’ record using “Apple Health.”

B. FEATURE EXTRACTION

Subsequently, we created temporal sequences of ten-day data for diabetes 2 participants using a sliding window technique, starting from the last glucose measurement of each user to the first in steps of -1, and taking the previous 10 glucose measurements of the same meal type (that is, we match by selecting only the observations that were measured at the same period of each day as the current observation for the window, that is either all before lunch, or all after lunch etc.) as the current glucose measurement, for the previous 10 days. The choice of using only 10 measurements for each prediction was inspired by other work [25] and we extend that approach by matching measurements of the same meal type. The 10 measurements of each window are not combined in any way, instead each one is used “raw” i.e. as a feature that will be used as input into the subsequent statistical learners. So for each glucose measurement selected as the current we fill its corresponding window by querying the past 10 days to fill in the features. The current glucose measurement, in each iteration, is not added to the data but is instead used to calculate the hypoglycemic episode status for each observation. As this process is iterative each participant will contain many time windows of ten-day data. The timestamps of the measurements are also added to the dataset. For example: suppose the current glucose measurement is 60 mg/dL and is taken before a meal. In this iteration the previous 10 glucose measurements that were taken before a meal will be added to the dataset and the value of 1 will be added as episode status (as $60 < 70\text{mg/dL}$). Moreover, for each glucose record its `source_id` ($1 \leq \text{source_id} \leq 10$) will also be added to the dataset. `Source_id` records the device used to take the measurement, where 1 defines the user manually inputting the measurement to the mobile app, 2 for Apple Health, 3 for Google Fit, 4 for Fitbit, 5 for Contour Next ONE etc (see Figure 5). In addition, in the same (ten-day) time windows for each iteration, the last 10 measurements of systolic blood pressure, as well as the last 10 measurements of diastolic blood pressure will be added, along with their respective timestamps. If there are missing measurements a value of -1 will be added for each missing blood pressure measurement (but not for glucose ones, if the respective time window contains less

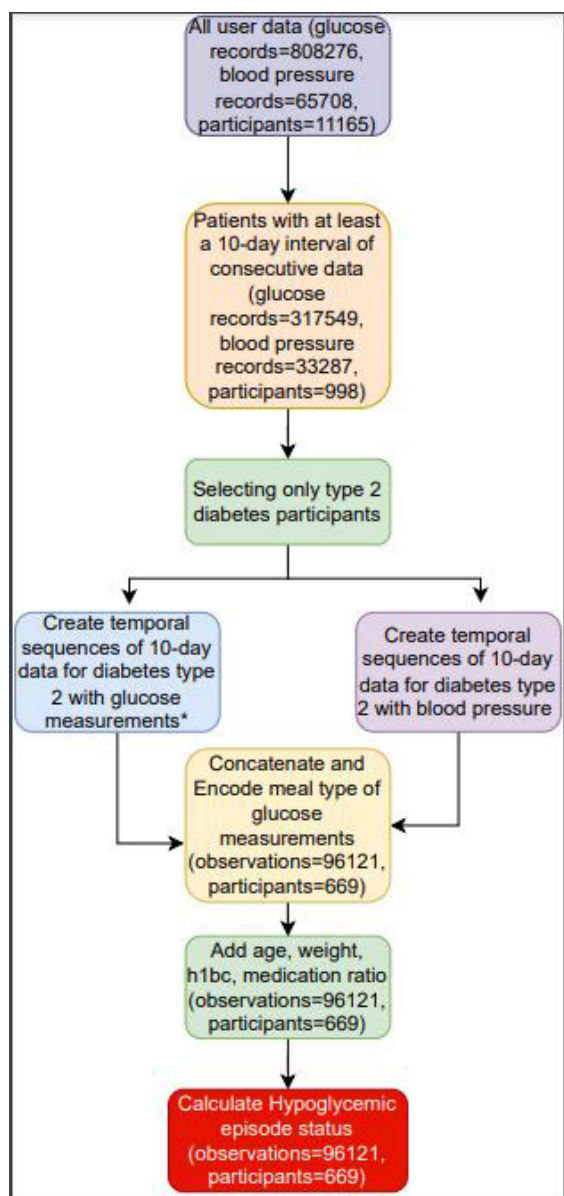


FIGURE 4. Data pre-processing pipeline.

than 10 previous glucose measurements this window will be dropped, thus the method requires 10 glucose measurements for each 10-day window). For each observation a one-hot encoding vector, of length 5, will be added encoding the meal type of the glucose measurements kept for each iteration. The five classes are: (fasting, before meal, after meal, bedtime, other). Finally, for each observation we add the participant's age, weight and a glycated hemoglobin (h1bc) measurement (nearest previous measurements in the time window) along with the timestamp of the measurement (or 0 and -1 respectively if no such record exists), as well as we calculate the ratio of $\frac{\text{sum_of_after_medication_glucose_measurements}}{\text{number_of_glucose_measurements}}$ for each time-window. The full pipeline for the final inclusion of 96121 observations from 669 participants is presented in Figure 4.

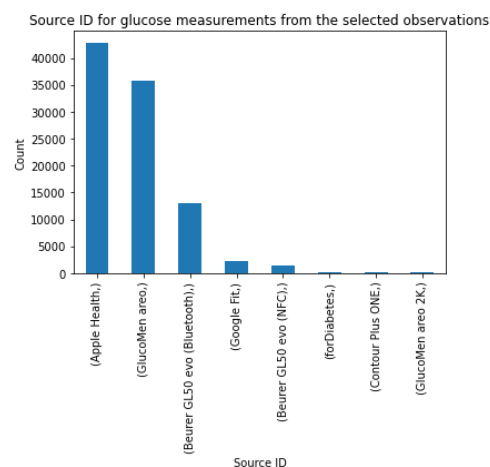


FIGURE 5. Devices used for recording glucose measurements for the selected data.

Before any further processing, we need to decide on a strategy to handle missing data. Only the 'age' variable was missing in 15 participants (5157 samples); otherwise this is a design matrix with complete entries. Given that preliminary analysis (see the section 'Statistical analysis') did not reveal age to be statistically significantly associated with the outcome, we decided not to include it in the model and retain the data from the 15 participants.

Figure 5 presents the sources for the measurements that are retained in the final dataset, after the application of the pre-processing pipeline. The final number of the participants is 669 and their average age is 57 ± 11 years. No information was stored about the gender of the participants. Table 1 presents the demographics of the 669 participants whose countries had at least 10 participants included in the study. They are from 86 countries in total. The countries in Table 1 are sorted according to the number of participants in descending order. However 24 participants had not defined their country in the app. We note these participants as *Undefined* in the table.

C. STATISTICAL ANALYSIS

We started data exploration using standard data visualization plots to assess probability densities, scatter plots, and computed statistical correlations between each of the features and the binary outcome to determine both whether these statistical relationships are statistically significant (at the $p = 0.05$ level) and also to assess the extent of the statistical strength. We used the empirical rule of thumb in medical applications that correlation coefficients that exhibit a magnitude above 0.3 are deemed to be statistically strong [29].

D. FEATURE SELECTION

Extracting a large number of features (70 in this study, once age has been excluded) may be detrimental for the performance of statistical models and challenging to interpret

TABLE 1. Participant demographics.

Country	Number of participants	Mean Age	Mean number of hypoglycemic episodes
US	156	53.92 ± 11.32	4.27 ± 15.41
DE	56	61.33 ± 9.98	0.84 ± 3.22
GR	55	55.04 ± 7.92	7.18 ± 33.47
PL	41	54.9 ± 10.77	2.46 ± 5.9
MX	30	51.6 ± 7.91	1.8 ± 4.0
Undefined	24	57.79 ± 13.29	14.0 ± 51.76
CA	22	56.45 ± 9.69	12.09 ± 25.88
IT	22	58.73 ± 13.83	12.18 ± 37.32
ES	22	55.36 ± 11.47	2.73 ± 11.02
GB	21	53.84 ± 9.17	1.71 ± 3.37
FR	15	62.07 ± 8.88	1.67 ± 4.92
CH	11	65.0 ± 6.93	1.73 ± 3.98
AT	10	58.3 ± 11.51	0.2 ± 0.6
AU	10	60.2 ± 11.3	10.6 ± 21.0
...

findings. According to Hastie et al. [30] *advanced statistical learning algorithms, in practice, are typically fairly robust to the inclusion of potentially noisy or irrelevant features*. However, identifying a smaller feature set always facilitates insight into the application by focusing on the key features contributing towards estimating the outcome [31]. Therefore, although in this study we do not have a very high-dimensional dataset, we nevertheless aimed to develop a parsimonious generalizable model with a succinct feature set. We used the new feature selection algorithm called relevance, redundancy and complementarity trade-off (RRCT), which was recently demonstrated to be extremely competitive across domains in 12 datasets when benchmarked against 20 state-of-art feature selection algorithms [32]. In brief, RRCT inherently accounts for the key elements towards identifying a robust information-rich compact feature subset, i.e. *relevance* (quantifying the statistical relationship of features with the outcome), *redundancy* (quantifying the statistical relationship between pairs of features in the selected subset), and *complementarity* (or conditional relevance, quantifying the conditional added value of joint feature sets over and above their univariate statistical association with the outcome). The features were selected using the strategy we have developed and explained in detail in previous work. The underlying concept is using a voting strategy to aggregate the feature sets selected when presented with perturbed versions of the dataset to ensure this robustly generalizes [32], [33], [34].

E. STATISTICAL MAPPING

We used state-of-the-art statistical mapping algorithms to develop a functional supervised learning model using the selected feature set from the preceding step to map onto the binary outcome. Specifically, we used: (1) RF [35], (2) Support Vector Machines (SVM), (3) Adaptive Boosting (AdaBoost), (4) XGBoost, (5) Feed-Forward Artificial Neural Network (ANN). We chose these methods as they are commonly used off-the-shelf classifiers that have been shown to be accurate in diverse supervised learning problems. Similarly to our previous studies, we explored different

approaches towards optimizing the statistical learners' hyperparameters [36]. For the RF we explored optimizing performance using Breiman's recommendation with half and twice the default recommended number of features over which to select features for the trees and explored the use of 500 and 1000 trees. We used the 'Statistics and Machine Learning Toolbox' for MATLAB for RF and the scikit-learn implementation (for Python 3) for SVMs and AdaBoost, xgboost (for Python 3) for XGBoost and tensorflow 2 for ANNs. We applied Z-Score Standardization for the features for all models except RF and the SVM. For the SVM we linearly rescaled the features to the [0..1] range and used a Gaussian radial basis function kernel. We clarify that for the scaling of the features in both the training and the testing subsets only the information from the training subset was used and subsequently applied to the testing subset. The regularization parameter C and the kernel coefficient γ were determined using a grid search where $C = [10^{-2}, 10^{-1}, \dots, 10^2]$ and $\gamma = [10^{-1}, \dots, 10^1]$. For AdaBoost, the learning rate hyper-parameter was explored in the range 0.01 to 0.5 in steps of 0.05 using 1000 trees with a maximum depth of 2. For ANNs we used a grid-search using a parameter grid of various parameters of 2-layer to 4-layer networks with first layer node count of $[2^5, \dots, 2^9]$ and subsequently halved the next layer node count. We utilized batch-sizes of 5, 20, 50 and tested 3 dropout configurations (0.2, 0.3, decrementing dropout ending with 0.2 on the final layer and increasing by 0.1 in each previous layer). We trained for 300 epochs with early-stopping (on validation loss with $patience = 7$) and use of the Adam Optimizer.

F. MODEL VALIDATION AND GENERALIZATION

The dataset in this study is highly unbalanced (94622/1499, total: 96121 samples indicating that most observations did not involve a hypoglycemic episode, i.e. > 98% samples in the dominant class). Problems where a class is dominating at that level are known to be particularly challenging for statistical learners, and hence we need to decide on a strategy towards the development and evaluation of the model. Given there is a very large number of samples available, for computational

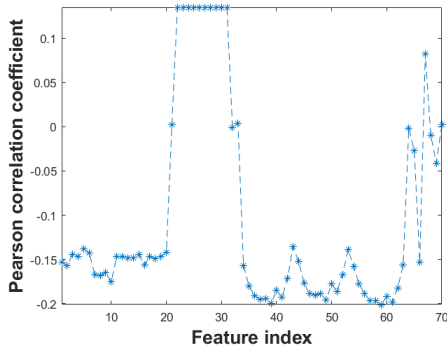


FIGURE 6. Correlation coefficients for all 70 features used in the study.

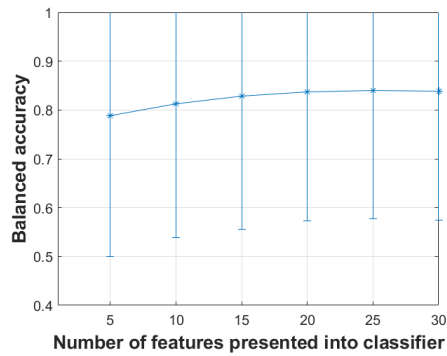


FIGURE 7. Errorbar depicting balanced accuracy as a function of the number of features presented into RF.

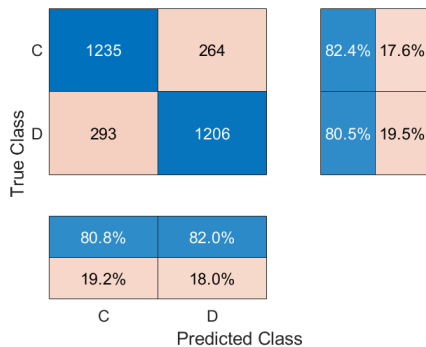


FIGURE 8. Confusion matrix along with probabilities of class estimates with the RF model comprising the 25 features selected using RRCT (C indicates control, D indicates diabetes).

efficiency and practicality, we created a balanced dataset comprising all 1499 samples from the non-dominant class and randomly selected 1499 samples from the dominant class.

We clarify that all analysis was carried on this balanced dataset, from the statistical analysis exploring associations to selecting a feature subset and statistical mapping. For the evaluation of the model performance we used the leave-one-subject-out (LOSO) approach because this is how we envisage this tool would likely be used in practice: we wanted to evaluate how well we might expect the model to generalize on new unseen people. We report different performance

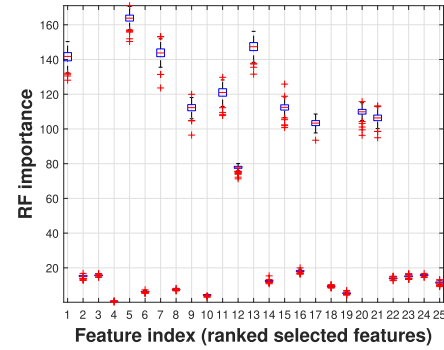


FIGURE 9. Random forest feature importance using the selected feature subset of 25 features. Higher value indicates that the corresponding feature contributes more towards the estimation of the binary response (detecting a hypoglycemic episode).

measures to assess model generalization, including confusion matrices, balanced accuracy, sensitivity, specificity, and F1-measure.

IV. RESULTS

We start our exploration by assessing statistical strength using correlation coefficients. We found that correlations (between the features used in the study) were generally relatively weak (univariately no feature is statistically strongly associated with the binary outcome). Nevertheless, some correlations were about $|0.2|$ (see Figure 6) which inspires confidence that when considered jointly in a statistical learning model may lead to good predictions.

Figure 6 presents the correlation coefficients for all features, whereas Table 2 presents the selected feature subset in descending order of importance. The five most important features were glucose₋₄, systolic₋₃, meal_{relatedHot_vec_1}, meal_{relatedHot_vec_3}, glucose₋₁. Figure 7 presents the balanced accuracy using LOSO as a function of the number of features presented into RF. We note that performance is very stable (in terms of using 5-30 features explored herein) and is optimized with 25 features, reaching 0.814 balanced accuracy. Figure 8 presents the confusion matrix of the RF model comprising the 25 features. As can be observed from the confusion matrix the misclassification rate between the two classes is similar.

Table 3 presents the balanced accuracy, sensitivity, specificity and F1-score for each model. The highest scoring model (in terms of balanced accuracy and F1-measure) is the one based on RF comprising the 25 features selected using RRCT with a balanced accuracy of 0.814, a sensitivity score of 0.805 and specificity of 0.824. Finally, Figure 9 presents the RF importance scores to obtain an overall impression of the actual contribution of each of the selected features towards estimating the two classes. As can be observed from the figure, some features were more important than others in estimating the binary response. The feature importance for Random forest could be grouped in 3 clusters, with cluster #1 containing all features with RF importance > 100 , cluster

TABLE 2. Selected features using RRCT in descending order of importance.

Feature name	Description
glucose_-4	Seventh glucose record of the time window
systolic_-3	Eighth systolic blood pressure record of the time window
meal_relatedHot_vec_1	Second value of meal type one-hot encoding vector
meal_relatedHot_vec_3	Fourth value of meal type one-hot encoding vector
glucose_-1	Tenth glucose record of the time window
meal_relatedHot_vec_0	First value of meal type one-hot encoding vector
glucose_-8	Third glucose record of the time window
glucose_-3	Eighth glucose record of the time window
medicationsRatio_onGlucoseRecords_	$\frac{\text{sum_of_after_medication_glucose_measurements}}{\text{number_of_glucose_measurements}}$ of the time window
hymoglobin	Nearest (previous) h1bc measurement in the time window
glucose_-10	First glucose record of the time window
glucose_-2	Ninth glucose record of the time window
glucoseDates_-10	Timestamp of the first glucose record of the time window
meal_relatedHot_vec_2	Third value of meal type one-hot encoding vector
weight	participant weight (nearest previous measurement)
glucose_-9	Second glucose record of the time window
glucose_-7	Fourth glucose record of the time window
glucose_-6	Fifth glucose record of the time window
systolic_-10	First systolic blood pressure record of the time window
glucose_-5	Sixth glucose record of the time window
source_id_-10	Source ID of the First glucose record of the time window
diastolic_-1	Tenth diastolic blood pressure record of the time window
diastolic_-7	Fourth diastolic blood pressure record of the time window
source_id_-9	Source ID of the Second glucose record of the time window
source_id_-3	Source ID of the Eighth glucose record of the time window

TABLE 3. Evaluation metrics.

	Balanced accuracy	Sensitivity	Specificity	F1-measure
All features RF	0.805	0.831	0.779	0.810
RF, 25 features (RRCT)	0.814	0.805	0.824	0.812
SVM, 25 features (RRCT)	0.65	0.57	0.74	0.66
AdaBoost, 25 features (RRCT)	0.78	0.72	0.84	0.78
XGBoost, 25 features (RRCT)	0.78	0.75	0.82	0.78
ANN, 25 features (RRCT)	0.80	0.79	0.80	0.80

#2 with feature #12 having an RF importance of around 80 and cluster #3 containing all features with RF importance scores of ≤ 20 . For convenience we present these results using boxplots to also present the variability in the internal RF importance weights arising from the different LOSO repetitions.

V. DISCUSSION

We presented a machine learning approach for next-day prediction of hypoglycaemia in daily life. Our primary finding is that prediction of hypoglycaemic episodes for T2DM based on mHealth data, such as blood glucose records and blood pressure measurements captured by widely available mobile devices for self-monitoring, is accurate.

The predictive outcomes relied on an international cohort which used a mobile app for diabetes self-management in the real world, in contrast with previous single-centre studies restricted to small geographic areas [25], [37], [38]. Furthermore, other works have focused mostly on CGM or data derived from electronic health records [28], and not data captured from ordinary mHealth devices used in everyday life. To the authors' knowledge this is the first study to assess next-day prediction of hypoglycaemia in an international cohort.

We compared the performance of different machine learning models in this study. RF proved to be empirically superior to SVMs, Adaptive Boosting and Feed-Forward Artificial Neural Networks. We do not have theoretical proof for the justification of this finding, however we note that in our experience RF has often worked well in complicated practical settings, thus providing further evidence to support the notion of being best *of-the-shelf* classifier as indicated by Hastie et al. [30].

The highest scoring model, based on RF with the 25 features selected with RRCT, reached a balanced accuracy of 0.814 and F1-score of 0.812 (with 0.805 sensitivity and 0.824 specificity). From the original features we selected features that had a more direct impact on hypoglycaemic incidents. Detailed information of medication, meals and physical activity (see the section 'Data description') was not utilised, because we found that to be of variable quality and depending on how adherent participants were to enter that information. The glucose records' table from the database of 'forDiabetes' contained columns which recorded the measurement's proximity to medication, physical activity and meal, i.e., if the measurement was recorded after taking medication/having a meal and exercising. Two of these features were used in the original 71, that is the ratio of

glucose measurements that were taken after medication to the glucose measurements of each time window, and the one-hot vector encoding the meal-type of the glucose measurements of each time window (i.e., if the measurements were taken after a meal/before a meal). The column containing information correlating the measurement with medication use was not included in the original 71 features as it was highly imbalanced (after medication samples were around 10% of the total samples).

In contrast with previous studies, our study relies on a large international cohort of individuals who have downloaded and used the mobile app in their daily lives. The dataset we used was imbalanced in regards to hypoglycemic incidents, as there were only 1499 cases of hypoglycemic incidents out of 96121 observations (indicating that most observations did not involve a hypoglycemic episode, i.e. > 98% samples in the dominant class). This provided the need to create a balanced dataset. We have repeated analysis twice using a different randomly selected subset from samples in the dominant class and repeated the methodology described. We found that the out of sample reported performance was very similar, which inspires confidence that the developed model will likely generalize well in new unseen data. However, additional studies in the real world are required to confirm our findings and accumulate robust evidence.

VI. CONCLUSION AND FUTURE WORK

We demonstrated that accurate and practical next-day hypoglycaemia prediction is feasible using real-world data with a custom-built diabetes-specific smartphone application. Therefore, the current work has enormous potential to enable day-to-day glycaemic control by participants with diabetes in the community, empower individuals to monitor potential problems, and facilitate the optimization of diabetes type 2 therapeutic management by care providers.

For the future we plan to expand the work presented by including a more diverse dataset covering various ethnicities, age groups, and co-morbid conditions to improve the robustness and universality of the findings, addressing potential biases in machine learning models in healthcare. Moreover we plan to contrast using other methods for feature selection and compare them to RRCT, as well as conduct a more thorough analysis of the model inaccuracies to understand the limitations.

REFERENCES

- [1] A. H. Heald, M. Stedman, M. Davies, M. Livingston, R. Alshames, M. Lunt, G. Rayman, and R. Gadsby, "Estimating life years lost to diabetes: Outcomes from analysis of national diabetes audit and office of national statistics data," *Cardiovascular Endocrinol. Metabolism*, vol. 9, no. 4, pp. 183–185, 2020.
- [2] H. Sun, P. Saedi, S. Karuranga, M. Pinkepank, K. Ogurtsova, B. B. Duncan, C. Stein, A. Basit, J. C. N. Chan, J. C. Mbanya, M. E. Pavkov, A. Ramachandaran, S. H. Wild, S. James, W. H. Herman, P. Zhang, C. Bommer, S. Kuo, E. J. Boyko, and D. J. Magliano, "IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Res. Clin. Pract.*, vol. 183, Jan. 2022, Art. no. 109119.
- [3] E. Standl, K. Khunti, T. B. Hansen, and O. Schnell, "The global epidemics of diabetes in the 21st century: Current situation and perspectives," *Eur. J. Preventive Cardiol.*, vol. 26, no. 2, pp. 7–14, Nov. 2019, doi: 10.1177/2047487319881021.
- [4] A. G. Ampofo and E. B. Boateng, "Beyond 2020: Modelling obesity and diabetes prevalence," *Diabetes Res. Clin. Pract.*, vol. 167, Sep. 2020, Art. no. 108362, doi: 10.1016/j.diabres.2020.108362.
- [5] R. Williams, S. Karuranga, B. Malanda, P. Saedi, A. Basit, S. Besançon, C. Bommer, A. Esteghamati, K. Ogurtsova, P. Zhang, and S. Colagiuri, "Global and regional estimates and projections of diabetes-related health expenditure: Results from the international diabetes federation diabetes atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 162, Apr. 2020, Art. no. 108072, doi: 10.1016/j.diabres.2020.108072.
- [6] Y. C. Kueh, T. Morris, and A.-A.-S. Ismail, "The effect of diabetes knowledge and attitudes on self-management and quality of life among people with type 2 diabetes," *Psychol., Health Med.*, vol. 22, no. 2, pp. 138–144, Feb. 2017.
- [7] R. Silbert, A. Salcido-Montenegro, R. Rodriguez-Gutierrez, A. Katabi, and R. G. McCoy, "Hypoglycemia among patients with type 2 diabetes: Epidemiology, risk factors, and prevention strategies," *Current Diabetes Rep.*, vol. 18, no. 8, pp. 1–16, Aug. 2018.
- [8] R. G. McCoy, H. K. Van Houten, J. Y. Ziegenfuss, N. D. Shah, R. A. Wermers, and S. A. Smith, "Increased mortality of patients with diabetes reporting severe hypoglycemia," *Diabetes Care*, vol. 35, no. 9, pp. 1897–1901, Sep. 2012.
- [9] R. Aronson, G. Galstyan, M. Goldfracht, S. Al Sifri, L. Elliott, and K. Khunti, "Direct and indirect health economic impact of hypoglycaemia in a global population of patients with insulin-treated diabetes," *Diabetes Res. Clin. Pract.*, vol. 138, pp. 35–43, Apr. 2018.
- [10] C. Rodriguez-León, C. Villalonga, M. Munoz-Torres, J. R. Ruiz, and O. Banos, "Mobile and wearable technology for the monitoring of diabetes-related parameters: Systematic review," *JMIR mHealth uHealth*, vol. 9, no. 6, Jun. 2021, Art. no. e25138, doi: 10.2196/25138.
- [11] A. K. Triantafyllidis, C. Velardo, D. Salvi, S. A. Shah, V. G. Koutkias, and L. Tarassenko, "A survey of mobile phone sensing, self-reporting, and social sharing for pervasive healthcare," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 218–227, Jan. 2017.
- [12] K. Woodward, E. Kanjo, D. J. Brown, T. M. McGinnity, B. Inkster, D. J. Macintyre, and A. Tsanas, "Beyond mobile apps: A survey of technologies for mental well-being," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1216–1235, Jul. 2022.
- [13] A. Triantafyllidis, H. Kondylakis, K. Votis, D. Tzovaras, N. Maglaveras, and K. Rahimi, "Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews," *Int. J. Med. Informat.*, vol. 132, Dec. 2019, Art. no. 103984.
- [14] M. Cui, X. Wu, J. Mao, X. Wang, and M. Nie, "T2DM self-management via smartphone applications: A systematic review and meta-analysis," *PLoS ONE*, vol. 11, no. 11, Nov. 2016, Art. no. e0166718.
- [15] S. A. Shah, C. Velardo, A. Farmer, and L. Tarassenko, "Exacerbations in chronic obstructive pulmonary disease: Identification and prediction using a digital health system," *J. Med. Internet Res.*, vol. 19, no. 3, p. e69, Mar. 2017.
- [16] Y. Fukazawa, T. Ito, T. Okimura, Y. Yamashita, T. Maeda, and J. Ota, "Predicting anxiety state using smartphone-based passive sensing," *J. Biomed. Informat.*, vol. 93, May 2019, Art. no. 103151.
- [17] B. Ballinger, J. Hsieh, A. Singh, N. Sohoni, J. Wang, G. H. Tison, G. M. Marcus, J. M. Sanchez, C. Maguire, J. E. Olgin, and M. J. Pletcher, "DeepHeart: Semi-supervised sequence learning for cardiovascular risk prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 2079–2086.
- [18] P. Martyn-Nemeth, S. Schwarz Farabi, D. Mihailescu, J. Nemeth, and L. Quinn, "Fear of hypoglycemia in adults with type 1 diabetes: Impact of therapeutic advances and strategies for prevention—A review," *J. Diabetes Complications*, vol. 30, no. 1, pp. 167–177, Jan. 2016.
- [19] Y. Wang, Z. Zeng, J. Ding, R. Yuan, R. Wang, Y. Zhang, L. Bai, H. Yu, and J. Tang, "Fear of hypoglycaemia among patients with type 2 diabetes mellitus: A cross-sectional study," *Sci. Rep.*, vol. 11, no. 1, p. 7971, Apr. 2021, doi: 10.1038/s41598-021-86954-0.

- [20] P. E. Cryer and B. P. Childs, "Negotiating the barrier of hypoglycemia in diabetes," *Diabetes Spectr.*, vol. 15, no. 1, pp. 20–27, Jan. 2002, doi: [10.2337/diaspect.15.1.20](https://doi.org/10.2337/diaspect.15.1.20).
- [21] T. Murata et al., "Prevention of hypoglycemia by intermittent-scanning continuous glucose monitoring device combined with structured education in patients with type 1 diabetes mellitus: A randomized, crossover trial," *Diabetes Res. Clin. Pract.*, vol. 195, Jan. 2023, Art. no. 110147, doi: [10.1016/j.diabres.2022.110147](https://doi.org/10.1016/j.diabres.2022.110147).
- [22] A. Bertachi, L. Biagi, I. Contreras, N. Luo, and J. Vehí, "Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks," in *Proc. 3rd Int. Workshop Knowl. Discovery Healthcare Data, 27th Int. Joint Conf. Artif. Intell., 23rd Eur. Conf. Artif. Intell. (IJCAI-ECAI)*, vol. 2148, K. Bach, R. C. Bunescu, O. Farri, A. Guo, S. A. Hasan, Z. M. Ibrahim, C. Marling, J. Raffa, J. Rubin, and H. Wu, Eds., Stockholm, Sweden, Jul. 2018, pp. 85–90.
- [23] Y. Deng, L. Lu, L. Aponte, A. M. Angelidi, V. Novak, G. E. Karniadakis, and C. S. Mantzoros, "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–13, Jul. 2021.
- [24] Y. Marcus, R. Eldor, M. Yaron, S. Shaklai, M. Ish-Shalom, G. Shefer, N. Stern, N. Golan, A. Z. Dvir, O. Pele, and M. Gonen, "Improving blood glucose level predictability using machine learning," *Diabetes/Metabolism Res. Rev.*, vol. 36, no. 8, Nov. 2020, Art. no. e3348.
- [25] B. Sudharsan, M. Peeples, and M. Shomali, "Hypoglycemia prediction using machine learning models for patients with type 2 diabetes," *J. Diabetes Sci. Technol.*, vol. 9, no. 1, pp. 86–90, Jan. 2015.
- [26] L. Mueller, P. Berhanu, J. Bouchard, V. Alas, K. Elder, N. Thai, C. Hitchcock, T. Hadzi, I. Khalil, and L.-A. Miller-Wilson, "Application of machine learning models to evaluate hypoglycemia risk in type 2 diabetes," *Diabetes Therapy*, vol. 11, no. 3, pp. 681–699, Mar. 2020.
- [27] Y. Ruan, A. Bellot, Z. Moysova, G. D. Tan, A. Lumb, J. Davies, M. van der Schaar, and R. Rea, "Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records," *Diabetes Care*, vol. 43, no. 7, pp. 1504–1511, Jul. 2020.
- [28] S. Kodama, K. Fujihara, H. Shiozaki, C. Horikawa, M. H. Yamada, T. Sato, Y. Yaguchi, M. Yamamoto, M. Kitazawa, M. Iwanaga, Y. Matsubayashi, and H. Sone, "Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: Meta-analysis," *JMIR Diabetes*, vol. 6, no. 1, Jan. 2021, Art. no. e22458.
- [29] A. Tsanas, M. A. Little, and P. E. McSharry, *A Methodology for the Analysis of Medical Data*. New York, NY, USA: Springer, 2013, pp. 113–125.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer Series in Statistics). New York, NY, USA: Springer, 2001.
- [31] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, Eds., *Feature Extraction, Foundations and Applications* (Series Studies in Fuzziness and Soft Computing). Cham, Switzerland: Springer, 2006.
- [32] A. Tsanas, "Relevance, redundancy, and complementarity trade-off (RRCT): A principled, generic, robust feature-selection tool," *Patterns*, vol. 3, no. 5, May 2022, Art. no. 100471.
- [33] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 181–190, Jan. 2014.
- [34] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, Apr. 2010, doi: [10.1109/TBME.2009.2036000](https://doi.org/10.1109/TBME.2009.2036000).
- [35] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [36] S. Arora and A. Tsanas, "Assessing Parkinson's disease at scale using telephone-recorded speech: Insights from the Parkinson's voice initiative," *Diagnostics*, vol. 11, no. 10, p. 1892, Oct. 2021.
- [37] N. N. Mathioudakis, E. Everett, S. Routh, P. J. Pronovost, H.-C. Yeh, S. H. Golden, and S. Saria, "Development and validation of a prediction model for insulin-associated hypoglycemia in non-critically ill hospitalized adults," *BMJ Open Diabetes Res. Care*, vol. 6, no. 1, Mar. 2018, Art. no. e000499, doi: [10.1136/bmjdc-2017-000499](https://doi.org/10.1136/bmjdc-2017-000499).
- [38] J. Yang, L. Li, Y. Shi, and X. Xie, "An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1251–1260, May 2019, doi: [10.1109/JBHI.2018.2840690](https://doi.org/10.1109/JBHI.2018.2840690).



ANASTASIOS ALEXIADIS received the bachelor's degree in computer science from Brunel University London, the first master's degree in intelligent systems from the University of Sunderland, the second master's degree from the University of Macedonia, Thessaloniki, Greece, and the Ph.D. degree in artificial intelligence from the University of Macedonia. He is a Computer Scientist and a Research Fellow with the Centre of Research and Technology Hellas, Information

Technologies Institute. He has participated in a number of national and European projects and has authored or coauthored more than 25 publications. His research interests include artificial intelligence and machine learning, optimisation, cryptography, and computer languages.



ATHANASIOS TSANAS (Senior Member, IEEE) received the bachelor's and master's degrees in engineering and the Ph.D. degree in applied mathematics from the University of Oxford, in September 2012. He was with the University of Oxford, as a Research Fellow of biomedical engineering and applied mathematics (2012–2016), a Stipendiary Lecturer of engineering science (2014–2016), and a Lecturer with the Said Business School (2016). Since January 2017, he has been a Chancellor's Fellow of data science with the Usher Institute of Population Health and Informatics.

He was a recipient of the National Greek State Scholarship, in 2003 and 2004, for graduated as the top student in biomedical engineering in Athens; the BNFL Prize for Best Undergraduate Thesis at the University of Liverpool, in 2007; the CTA Scholarship for the M.Sc. study with The University of Newcastle, in 2007; the Intel/EPSC Scholarship for the Ph.D. study with the University of Oxford, from 2008 to 2012; the Student Paper Award at the NOLTA International Meeting, in 2010; the Andrew Goudie Award from St. Cross College, University of Oxford, in 2011; the EPSC Doctoral Prize Award, in 2012; the Young Scientist Award at the International Workshop MAVEBA, in 2013; and the EPSC Statistics and Machine Learning Award, in 2015. He was shortlisted in the final six candidates for the Papanikolaou Prize, in 2011, and was part of the Oxford Biomedical Engineering Team that won the annual Physionet/Computing in Cardiology Competition, in 2012, for "Predicting Mortality of ICU Patients." He was an Outstanding Reviewer of the journal *Computers in Biology and Medicine*, in 2015; and won the Best Reviewer Award from the IEEE JOURNAL OF BIOMEDICAL HEALTH INFORMATICS, in 2015.



LEONARD SHTIKA received the B.Sc. degree in applied informatics from the University of Macedonia (UOM) and the M.Sc. degree in information and communication technology systems from International Hellenic University (IHU). He is a Software Engineer with over 17 years of professional experience in a wide range of technologies. He is an open source enthusiast and loves to explore, create, and contribute to open source tools. He is the Chief Technology Officer (CTO) at Tessera Multimedia and also holds a position as a Lecturer at the University of York Europe Campus.



VASSILIS EFOPOULOS received the first B.Sc. degree in physics from the Aristotle University of Thessaloniki, the second B.Sc. degree in computer science from the University of Crete, and the Ph.D. degree in computer education from the University of Macedonia. He is the General Manager of Tessera Multimedia.



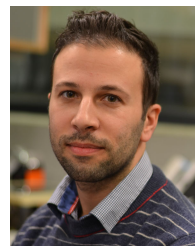
he is also the Director of the Visual Analytics Laboratory. He is also a Visiting Professor with De Montfort University, U.K., in the field of human–computer interaction, virtual and augmented reality (VR), and augmented reality (AR). His Ph.D. study was in the domain of service oriented architecture and information management systems in the e-inclusion and healthcare. Since 2001, he has published more than 100 international scientific articles on e-Health, m-Health, human-technology interaction visualisation (AR, VR, and information visualisation with visual analytics), mobile and ambient technologies, and human-centred design. Since 2001, he has participated in 29 research and development projects in the fields of HCI, the IoT, information visualisation and data analytics, personal health systems, cybersecurity, blockchain, ambient assisted living, e-inclusion, and big data. His research interests include human–computer interaction (HCI), information visualisation and management of big data, knowledge engineering and decision support systems, the Internet of Things, cybersecurity, and pervasive computing, with major application areas, such as m-Health, e-Health, and personalized healthcare.

KONSTANTINOS VOTIS received the M.Sc. and Ph.D. degrees in computer science and service oriented architectures from the Computer Engineering and Informatics Department, University of Patras, Greece, and the M.B.A. degree from the Business School Department, University of Patras. He is a Computer Engineer and a Senior Researcher (Researcher Grade B') with the Centre for Research and Technologies Hellas, Information Technologies Institute (CERTH/ITI), where



of CERTH (CERTH/ITI), in October 2013. His main research interests include visual analytics, 3-D object recognition, search and retrieval, behavioral biometrics, assistive technologies, information and knowledge management, multimodal interfaces, computer graphics, and virtual reality. He has been a researcher, since September 1999. He has involved in more than 60 projects, funded by the EC and the Greek Ministry of Research and Technology. His involvement with those research areas has led to the coauthoring of over 80 articles in refereed journals and more than 150 papers in international conferences. He is a member of EURASIP. He has served as a regular reviewer for a number of international journals and conferences.

DIMITRIOS TZOVARAS (Senior Member, IEEE) received the M.Eng. and Ph.D. degrees in electrical and computer engineering from the Aristotle University of Thessaloniki. He is a Senior Researcher (Researcher A') with the Centre for Research and Technology Hellas (CERTH). He is the Director of the Central Directorate and the Chairperson of the Board of Directors of CERTH, since December 2020. Prior to his current position, he was the Director of the Information Technologies Institute



of CERTH (CERTH/ITI), in October 2013. His main research interests include visual analytics, 3-D object recognition, search and retrieval, behavioral biometrics, assistive technologies, information and knowledge management, multimodal interfaces, computer graphics, and virtual reality. He has been a researcher, since September 1999. He has involved in more than 60 projects, funded by the EC and the Greek Ministry of Research and Technology. His involvement with those research areas has led to the coauthoring of over 80 articles in refereed journals and more than 150 papers in international conferences. He is a member of EURASIP. He has served as a regular reviewer for a number of international journals and conferences.

ANDREAS TRIANTAFYLLIDIS is a Research Associate with the Centre for Research and Technology Hellas, Information Technologies Institute (CERTH-ITI). He trained as a Postdoctoral Researcher in mobile health with the Department of Engineering Science, University of Oxford, from 2013 to 2015. During the last 11 years, he has involved in several EU and U.K. NHS-funded projects (e.g., PATHway, REMOTE, and SUPPORT-HF), leading in the design, development, and evaluation of novel computerised systems toward achieving patient independent living, better health self-management, and improved clinical outcomes. His research has attracted national and international grants. He has published in international peer-reviewed journals, such as the *International Journal of Medical Informatics* and the *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*. His research has covered in Tech Talks Central, in 2014; OxfordMail, in 2015; and Deutsche Welle, in 2016. His research interests include medical informatics, pervasive computing, personalised healthcare, physical activity, and computerised decision support. He received the Best Paper Awards in international conferences, such as International Symposium on Ambient Intelligence 2011 and MobiHealth 2010.

...