IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Machine Learning Tools for Long-term Type 2 Diabetes Risk Prediction

NIKOS FAZAKIS[1], OTILIA KOCSIS[1], ELIAS DRITSAS[1], SOTIRIS ALEXIOU[1], NIKOS FAKOTAKIS[1], (MEMBER, IEEE), AND KONSTANTINOS MOUSTAKAS[1], (SENIOR MEMBER, IEEE)

[1]Department of Electrical and Computer Engineering, University of Patras, 26504 Rion, Greece

Corresponding author: Nikos Fazakis (e-mail: fazakis@ece.upatras.gr).

This work has been partially supported by the SmartWork project (GA 826343), EU H2020, SC1-DTH-03-2018 - Adaptive smart working and living environments supporting active and healthy ageing.

ABSTRACT A steady rise has been observed in the percentage of elderly people who want and are still able to contribute to society. Therefore, early retirement or exit from the labour market, due to health-related issues, poses a significant problem. Nowadays, thanks to technological advances and various data from different populations, the risk factors investigation and health issues screening are moving towards automation. In the context of this work, a worker-centric, IoT enabled unobtrusive users health, well-being and functional ability monitoring framework, empowered with AI tools, is proposed. Diabetes is a high-prevalence chronic condition with harmful consequences for the quality of life and high mortality rate for people worldwide, in both developed and developing countries. Hence, its severe impact on humans' life, e.g., personal, social, working, can be considerably reduced if early detection is possible, but most research works in this field fail to provide a more personalized approach both in the modeling and prediction process. In this direction, our designed system concerns diabetes risk prediction in which specific components of the Knowledge Discovery in Database (KDD) process are applied, evaluated and incorporated. Specifically, dataset creation, features selection and classification, using different Supervised Machine Learning (ML) models are considered. The ensemble WeightedVotingLRRFs ML model is proposed to improve the prediction of diabetes, scoring an Area Under the ROC Curve (AUC) of 0.884. Concerning the weighted voting, the optimal weights are estimated by their corresponding Sensitivity and AUC of the ML model based on a bi-objective genetic algorithm. Also, a comparative study is presented among the Finnish Diabetes Risk Score (FINDRISC) and Leicester risk score systems and several ML models, using inductive and transductive learning. The experiments were conducted using data extracted from the English Longitudinal Study of Ageing (ELSA) database.

INDEX TERMS T2DM, long-term health risk prediction, machine learning, ensemble learning

I. INTRODUCTION

Diabetes, also known as diabetes mellitus (DM), is a chronic disorder characterized by high blood glucose levels, due to the inability of the pancreas to generate a sufficient quantity of insulin (Diabetes Mellitus Type-1 (T1DM)) or the failure of cells and tissues to utilize it (Diabetes Mellitus Type-2 (T2DM)) [1]. Apart from T1DM and T2DM, another type is Gestational diabetes, which affects women and develops during pregnancy. Since the prevalence of T2DM in ageing population (i.e., elderly people) is rising [2], [3], the analysis in the following sections focuses on such age group which constitutes the participants in SmartWork. Some characteristic signs and symptoms of high glucose include itching, frequent fatigue, unexplained weight loss, excessive urination, dry mouth and increased hunger [4]. The prevention and/or early diagnosis of diabetes is of high importance in order to avoid or mitigate the serious lifetime complications including cardiovascular ailment, stroke, kidney failure, ulcers in the foot, and eye complications etc [5], [6]. In conventional healthcare, the patient demographic data, case history, diagnostics and medication are manually managed and maintained, which may lead to human errors and affect patients suffering from chronic diseases. It is known that, diabetes patients need to check their glucose level regularly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

or even continuously to make sure that their lifestyle (i.e., diet and physical activity) is the appropriate one to keep glucose levels under control. There are many such medical devices that facilitate the measuring of glucose levels from the patients themselves.

Yet, the recent technological advances in networking, namely mobile communications (e.g., 5G and beyond networking), Cloud Computing, Internet-of-Things (IoT), Artificial Intelligence (AI) and Machine Learning have increased the number of internet-connected smart devices, such as wearable sensors, and revolutionized the way the medical industry operates. In fact, they paved the way to robust, fast and smart systems, known as Internet of Medical Things (IoMT), able to handle massive users data rapidly. IoMT with smart sensors, smart devices and smart communication protocols facilitated the development of various smart systems in the field of healthcare [7], [8], [9]. Such systems have become essential as they are expected to eliminate human intervention, thus significantly reducing human errors and assisting medical experts in diagnosing the diseases easily, remotely and accurately, by combining various data collected from the monitoring devices over a sensor network with a decision support system. In [10], authors conducted an extended literature review in different domains, such as clinical decision support systems, wireless body area networks, cloud computing and big data analytics, in which they identified a positive impact in mobile healthcare for diabetes mellitus. Recently, in [11], a smart healthcare framework for ambient assisted living using IoMT and big data analytics techniques was suggested.

In the special case of diabetes, smart devices measure the glucose level of the patients and make it available in real-time to the doctors through mobile or web applications. Authors in [12] suggest a personalized recommendation system to support diabetes management by the American Indians patients themselves. Some other remote monitoring systems for diabetic patients are mentioned in [13]. T2DM and other chronic diseases monitoring can be enhanced with the implementation of appropriate machine learning algorithms. Machine learning and data mining methods constitute key approach in T2DM research for extracting knowledge. The severe social impact of T2DM renders it one of the main priorities in medical research, which unavoidably generates huge amounts of data. Hence, predictive analytics, machine learning and data mining approaches in T2DM are of major concern when it comes to diagnosis, management and other related clinical aspects.

Machine learning approaches can be categorized as supervised, semi-supervised and unsupervised learning. In the context of this work, our focus is on supervised machine learning methods with the aim to predict the risk of T2DM. Supervised ML algorithms, and especially classification algorithms, use a two-stage methodology for the pattern recognition task. The first stage is dedicated to the development/construction of the model using existing labeled training datasets, while the second stage involves the prediction for new or unseen input datasets. During the training phase, the annotated dataset, for which both the inputs (features) and the outputs (classes) are known, is partitioned into two sets (training and test), with the model being trained on the training set and tested on the test set, and the performance of the model being evaluated based on the correct predictions made.

Predictive analytics [14], [15] is the process of learning from historical data in order to make predictions about future events. It is widely applicable to almost every domain, and enhanced by the increasing availability of large volumes of data. Statistical data analysis methods were the go-to choice in predictive analytics, but when it comes to pattern recognition in large data sets (e.g. dense time series), they are consistently outperformed by ML algorithms, both in terms of accuracy and scalability.

The individual risk of developing non-contiguous chronic conditions is linked to controllable lifestyle behaviour. The quantification of said risk is an important goal of prediction analysis in healthcare [16], since, not only is it linked to both the long-term well being of the individual, but is also beneficial to social care systems. Recent research [17], [18] has demonstrated that it is possible to use ML tools to predict individual risk of hospitalization by only using data related to socioeconomic features (age group, gender and race) and behavioural data, without requiring clinical risk factors [19]. An extremely large number of ML algorithms and variations exist, and there is no unique or widely applicable solution for a specific domain or problem. As such, each particular problem and prediction task requires performance evaluation of multiple algorithms in order to identify the best performing one [20].

Given that T2DM is a multifactorial chronic condition, it requires adjustments in multiple aspects of a person's daily life in order to prevent it. For instance, alterations in dietary habits and physical activity might be deemed necessary, depending on their personal data. A person's motivation is important for the engagement and success of a digital health personal intervention. It is highly unlikely that people, who are used to a sedentary lifestyle, will suddenly adhere to guidelines regarding physical activity and dietary restrictions, even if the digital health intervention systems dictates it. Also, people, who do not need or want to change real-life behaviour, will not use any application as intended. Therefore, the motivation of the individual to be healthy, during and outside working hours, is very relevant for SmartWork System implementation. Previous studies performed in the context of the SmartWork project were focused on assessing individual/group motivation to be healthy (e.g. in the physical activity domain) and various factors impacting on office worker's performance (e.g. sleep quality) [21], which are out of the scope of the current work.

Motivated by the aforementioned challenges, the main contributions of this work are summarized as follows:

- We describe the data-driven AI component of the SmartWork system, comprised of Personalized Predictive

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

Models and Decision Support Tools. These sub-systems implement long-term predictive models and data mining techniques to provide probabilistic prediction of specific risk indicators aiming at supporting decision making and intervention for T2DM, among other chronic conditions. A detailed description of the functional ability modelling components and rules manager is elaborated in Section 3.

- Although a multitude of potential prediction tasks for several chronic diseases have been elaborated in the system, the analysis here only concerns the long-term T2DM risk prediction. For this case, various ML algorithms are investigated for the selection of the best performing model to be integrated in the SmartWork system. In the scope of training the SmartWork prediction models about T2DM (and other chronic diseases), a subset of the ELSA longitudinal dataset is employed to train the supervised algorithms for the assessment of T2DM long-term risks. It is worth to mention that, the generated dataset may contribute to the prognosis of T2DM as we choose to monitor the features' values of users who, in reference waves, have not been diagnosed with diabetes. Note that, the diabetic or non-diabetic class label is indicated by the follow-up assessment after 2-years, as it is explained in Section 3.1.2.
- A comparative analysis of the trained models is performed in relation to different performance metrics such as AUC, Sensitivity (or Recall) and Specificity, to name a few. Remark that, the sensitivity of the model is quite important when comparing classification models, as in T2DM case indicates the percentage of correctly identified instances of diabetic class.

The remaining of this paper is structured as follows. In Section II, we overview previous related studies. In Section III, we introduce the proposed system architecture. In Section IV, the design of the T2DM risk assessment system is described in detail. In Section V, the system performance is evaluated. Finally, concluding remarks and plans for our future work are provided in Sections VI and VII.

II. RELATED WORK

As regards the T2DM risk prediction, there are several representative works about the application of ML techniques and moreover suggestions of derived risk scoring systems that can be adopted on the early prognosis of diabetes. Furthermore, a number of intelligent systems have been developed that enable the remote (continuous) monitoring for diabetic patients, risk prediction and personalized health services, based on the data collected from smart body sensors which are given as input to ML models.

A. RISK SCORING AND MACHINE LEARNING IN T2DM

Up to date, an extensive research has been conducted from the scientific community for diabetes detection. To this end, several non-invasive risk score systems have been proposed, such as FINDRISC, Latin America FINDRISC (LA-FINDRISC) [22], Australian Type 2 Diabetes Risk Assessment Tool (AUSDRISK) [23], Risk Test from American Diabetes Association (ADA) [24], Leicester Practice Risk Score [25], Test2Prevent , which proved to be an effective screening tool to assess the risk of undiagnosed T2DM, especially in cases where confirmation tests data are not available. However, a significant constraint is that most of them were developed for particular populations and their performance was not satisfactory when applied to other ones. Assuming that fasting plasma glucose (FPG) or hemoglobin A1C (hbA1c) testing or an oral glucose tolerance test (OGTT) data is available, the diagnostic accuracy of the aforementioned risk score systems can be verified [26]. Liu et al. in [18] showed that the risk scoring systems can be combined with other ML models, constructing ensemble learners, to improve prediction performance.

Machine learning methods have gained popularity in the research community for automating the risk prediction process of T2DM, more accurately and with reduced medical cost. Artificial neural networks (ANNs), Logistic Regression (LR), Naive Bayes (NB), k-Nearest Neighbours (k-NN), Random Forests (RFs), Decision Trees (DT), and Support Vector Machines (SVMs), [27], [28] are the most popular algorithms which can be utilized. Naz and Ahuja, in their work [29], explore several of these models on the PIMA Indians diabetes database, proposing a deep neural network (DNN) able to achieve an accuracy of 98.07%. The classifiers can be used either individually or as base classifiers for ensemble (namely, stacking, voting, bagging etc.) algorithms [30], [31]. Ensemble learning aims to reduce bias and variance, and thus, enhance the prediction performance.

The aforementioned models have been used in several decision support systems for medical applications demonstrating satisfactory predictive performance. The researchers, in order to automate in an intelligent and effective way the process of diabetes monitoring, resorted to solutions combining Information and Communication Technology (ICT) with biomedicine. Such solutions are presented in the following paragraphs.

B. SMART SYSTEMS IN DIABETES HEALTHCARE

In [32], an intelligent system consisting of smart devices and sensors, and smartphones for monitoring diabetic patients, by means of machine learning algorithms, is elaborated. The smart system collects data from body sensors and makes diabetes diagnosis using several classification models from supervised machine learning. As the experimental results show, the suggested algorithm, namely the sequential minimal optimization (SMO), behaves better in terms of classification accuracy, sensitivity and precision than other well-known algorithms, i.e., Naive Bayes, J48 [33], ZeroR, OneR, Logistic, Random Forests). Another intelligent system is suggested in [34] for the remote monitoring of diabetic patients health through smartphones and other smart portable

[0]https://www.idf.org/type-2-diabetes-risk-assessment/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

devices. They designed a small portable device capable of measuring the blood glucose level for diabetics and body temperature which could be connected with a smartphone through a secure wireless mechanism.

Also, in [35] a smart health monitoring architecture is recommended for diabetic patients to monitor symptoms/signs regarding blood sugar level, heart pulse, food intake, sleep time and exercise. A sensors network is feeding continuously the input of the system with data which are then utilized as input to a neural network. The health risk levels range from low, medium and high to extreme, depending on patient's profile and health historical data. Moreover, if a patients health status is at high or extreme risk, an automatic notification (such as, phone call and/or SMS) is being sent to his/her relative with information about his/her location. Besides, in case of very high risk, the system communicates with the nearest to patient hospital.

The scientific work in [36] suggests several new wearable devices, such as smart neck band, smart wrist band and a pair of smart socks - to continuously monitor the health status of diabetic patients. The sensors of these devices report patient's food intake, heart rate, skin moisture, ambient temperature, walking patterns and weight gain/loss. With the help of controllers, these devices transmit sensors data via Bluetooth to the Mobile App. Machine Learning is employed to predict the variations in patient health status and alert them.

Moreover, there are many proposals for remote health monitoring of older persons [37], [38], [39]. Understanding and improving age-friendly living and working environments is an enormous challenge that today's societies have only just begun to approach. As the number of older people who are active members of society and want to live independently continues to rise, the importance of this research area constantly increases. The overall objective of the SmartWork system [40] is to support office workers remain professionally active as they grow old, in a holistic way, by designing, implementing and validating the system in real-world settings.

III. THE SMARTWORK

In the core development of the system, a worker-centric AI module [41] supports the sustainability of work skills, combining unpretentious and ubiquitous sensing and flexible worker-status aware job support. In addition, the careful and systematic monitoring of personal health, lifestyle, cognitive and emotional state of the worker makes it possible to determine the likelihood of functional and cognitive decline. By combining all aspects of the older workers' profile, a decision support system will enable triggering personalized interventions in order to maintain the work ability of the user. More specifically, the automatic creation and maintenance of the personalized virtual user model considers adaptation levels that consist of two layers: initialization of the user profiles based on generic group modelling derived through the observation of common patterns and characteristics of populations (e.g. gender, age group chronic conditions), and personalized models based on the monitored characteristics of a specific user (e.g. stress, emotions, activity, nutrition etc.). Based on the synchronous and asynchronous analysis of the data collected by the Smartwork sensing system, the initial user profiles are evolving to personalized user models.

It should be pointed out, here, that the problem of interest in this work is to emphasize on Long-Term Health Risk Assessment related to diabetes that statistically affects people older than fifty, which may suffer from hypertension, high cholesterol or heart disease as well.

A. SYSTEM ARCHITECTURE

Considering that the whole system is dynamically capturing the evolving state of the worker and the context of work and working environment (e.g. work task resources requirements), the office worker profile aspects are constantly monitored and analyzed using various services and agents. In more detail, the AI software tools package consists of a set of modules (Figure 1) dedicated to initialize the first user profiles, match them to lifestyle and behavioral patterns, continuously monitor, self-adapt and trigger interventions relevant for the work and health self-management of the office worker. In the following paragraphs, we will briefly elaborate on the different modules whose results are fed on to the module that performs Long-Term Health Risk Assessment in order to derive a predictive score reflecting the overall risk of the individual to experience the T2D chronic disease, which may result in early exit from the market labor.

The User Profile Initialization process takes place at the user's first contact with the SmartWork system, and it concerns collection of data regarding socio-demographic characteristics and lifestyle attitudes of the user, such as age, gender, marital status, education level, physical activity frequency, drinking and smoking status, etc. The user's history of diagnosed chronic conditions, including diabetes, asthma, high blood pressure, cholesterol and cardiovascular diseases is also assessed. Once the profile is completed, based on this initial data provided by the user, the prediction models are used to initialize the Long-term Cognitive Capacity Assessment and the Long-Term Health Risks Assessment modules.

Another important module is the Rules Manager Service (RMS), which is the software package implementing the different sub-modules needed in order to systematically monitor and activate the triggering of the SmartWork interventions in respect to the primitive or derived virtual user model data. The SmartWork continuously monitors a wide range of variables regarding the users' lifestyle, functional, cognitive and work ability status, which represent input for the RMS, either in the form of original raw data or as processed information generated by the SmartWork pre-processing algorithms, statistical analysis tools and ML-based prediction models. Although a series of physiological parameters are monitored, which are related to user's health status, it is important to mention that the SmartWork does not aim to provide any diagnostics, treatment or cure, but rather aims to provide the user with advice, guidance and suggestions

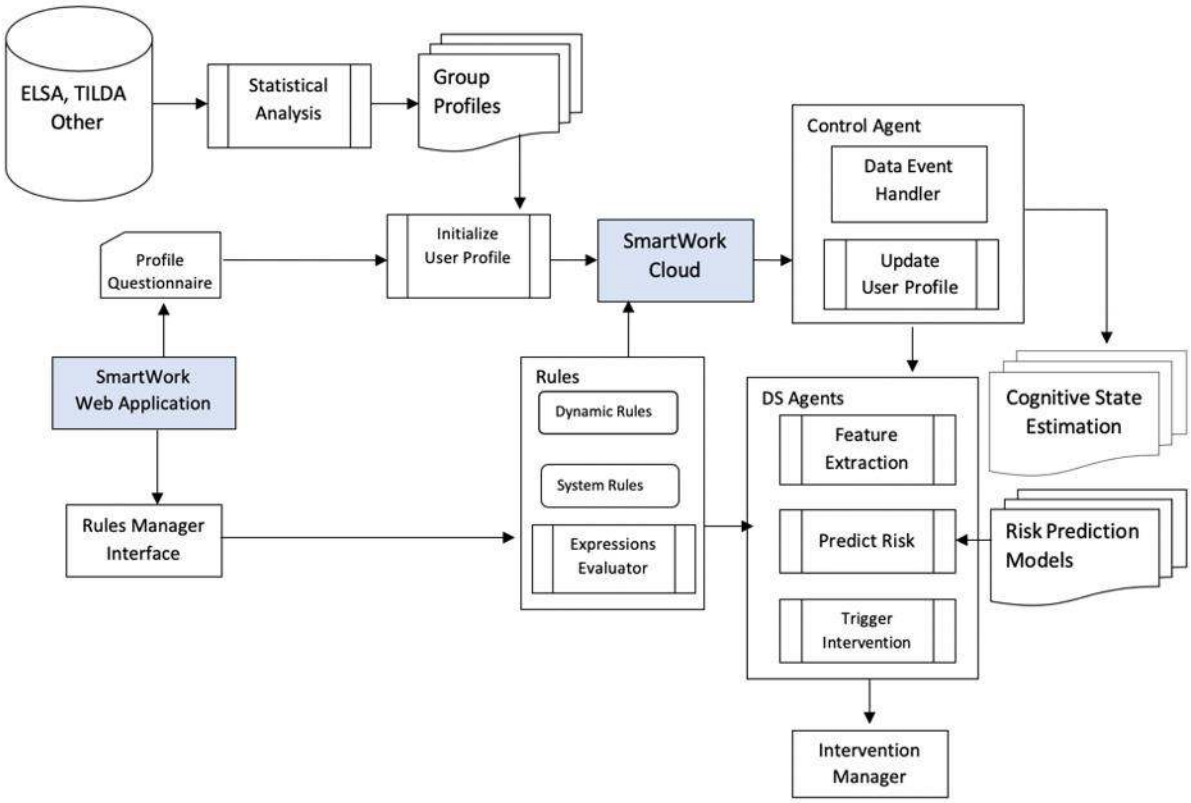

FIGURE 1: AI Tools Software Modules and Interconnections.

that can lead behavioural changes aiming to improve his/her overall health and work ability in alignment with the principles of professionally active and healthy ageing. The basic sub-modules of the RMS are the Rules Manager Daemon, the Run-Time Expression Evaluator and the Rules Manager Control Interface. The Rules Manager Daemon (RMD) is the main micro-service around which the RMS is designed. In practice, the RMD acts as an integrated server that orchestrates the real-time monitoring and evaluation of the user model variables against specific rules in order to identify the accomplishment of conditions that may trigger associated interventions. At the core of the RMD micro-service algorithm, the run-time Expression Evaluator performs logical and arithmetic operations dynamically based on the virtual user profile variables, thus evaluating the accomplishment of triggering conditions in the defined rules, and providing the RMS with a higher level of abstraction and the ability to evaluate complex expressions based on the available input variables.

The Rules Manager Control Interface (RMCI) is a web application designed to provide a convenient solution for the generation and management of intervention triggering rule sets which are then passed to RMD micro-service to populate the Rules Table. It is a multi-user environment, able to administer different user privilege levels that can have specific access on each virtual user profile dynamic rule set settings. The RMS has a client-server architecture and the RMCI was built as a stand-alone client application which can be used by the end users through a web browser or as a desktop application.

The next sections provide the necessary background knowledge for the remainder of the paper. In following, useful definitions and notations will be recorded under the problem definition and formulation, with the most characteristic being the dataset preparation and Machine Learning components under the investigating issue.

IV. LONG-TERM DIABETES RISK ASSESSMENT

A. PROBLEM DEFINITION

Chronic diseases are diseases that cannot be cured but can be controlled and thus they require continuous monitoring and acute care to avoid critical conditions. Diabetes is a chronic disease that occurs when the pancreas is no longer able to produce insulin, or when the body cannot make good use of the insulin it produces. Insulin is a hormone that lets glucose from the consumed food pass from the blood stream into the cells to produce energy. Not being able to produce insulin or use it effectively leads to raised glucose levels in the blood, also known as hyperglycemia. Over the long-term, high glucose levels are associated with damage to the body and failure of various organs and tissues. Although there is more than one type of diabetes (e.g. type 1 diabetes, type 2 diabetes, gestational diabetes), prevalence of type 2 diabetes amongst the older people is particularly high overall and in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

comparison with prevalence of other types of diabetes [42]. T2DM usually affects adults, but it can begin at any time in people life. The main risk factors [43], [44], [45] that are correlated to the occurrence of T2DM include:

- Age: is one of the most important risk factors for diabetes, as older people have a higher risk to get type 2 diabetes.
- Obesity/ High Body Mass Index (BMI): increased BMI, and consequently obesity, is a top risk factor for type 2 diabetes.
- Impaired glucose tolerance, also known as prediabetes, is a milder form of type 2 diabetes, which is usually diagnosed with a simple blood test, and represents a high risk for the individual to develop T2DM.
- Ethnicity/Race: prevalence of diabetes is overall higher in the case of Hispanic/Latino Americans, African Americans, Native Americans, Asian-Americans, Pacific Islanders, and Alaska natives.
- Gender: male/female
- Gestational diabetes: this short-term condition that may occur during pregnancy, raises a women's chances of getting type 2 diabetes later in life.
- Polycystic ovary syndrome (PCOS): women with polycystic ovary syndrome have a higher risk to develop T2DM.
- Family history: if a parent/sibling has diabetes, then risk of getting type 2 diabetes is increased.
- Physical Activity: sedentary persons are at higher risk of developing T2DM.
- Smoking: smoking is associated with a higher risk of T2DM.
- High Blood Pressure (HBP): it is a high-risk factor for developing T2DM.
- Alcohol: although moderate drinking is associated with a lower risk of, excessive alcohol intake is associated with an increased risk of type 2 diabetes.

Many studies aimed at long-term risk prediction for diabetes, including also different regression models for predicting glucose regulation for those already diagnosed with prediabetes or type 2 diabetes. However, the main goal of long-term diabetes risk prediction tools is to develop and validate a diabetes risk assessment score for healthy/undiagnosed participants based on main risk factors, including socio/demographic data, lifestyle, and simple anthropometric measures.

In SmartWork, a long-term risk prediction model for T2DM based on ML approaches is implemented, which takes into account a large number of risk factors which are usually employed by the screening tools used in medical practice, but also some factors which have shown high correlation based on our study with the ELSA dataset as shown in Tables 1 and 9. In order to test our model, we selected the FINDRISC [46], Leicester [25] Diabetes Risk Scores to apply it in parallel to the training and test dataset. The Leicester Practice Risk Score was developed by researchers within the Diabetes Research Centre at the University of Leicester and the score identifies people who may be at high risk of developing diabetes in the future (e.g. next 10 years) or currently having undiagnosed T2DM or prediabetes, taking into account the following risk factors: age, gender, BMI, ethnicity, family history of diabetes and diagnosis of high blood pressure or anti-hypertensive drugs use. In order to compare the results of the FINDRISC and Leicester risk classification to the ML prediction models, we fit Logistic Regression models to our data and estimate the probability an instance to be classified as "Diabetics" or "Yes" and "Non Diabetics" or "No".

The English Longitudinal Study of Ageing [47], which is a rich resource of information on the dynamics of health, social well being and economic circumstances in the English population aged 50 and older, has currently reached wave 9 of longitudinal data collection (e.g. covering a period of 18 years) and it is designed to be used for the investigation of a broad set of topics relevant to understanding the ageing process. The database contains both objective and subjective data related to health, disability, and healthy life expectancy, with specific data being assessed by a nurse every four years. In the scope of training the Smartwork prediction models, the waves at which nurse collected data are available are of particular interest, as these include physical examination and performance data and blood tests (e.g. height and weight, waist and hip circumference, blood pressure, lung function, total and DHL-cholesterol, etc.). Note that, these waves are considered reference waves in Smartwork.

B. METHODS

We assume a training set TR of size M, a test set TS of size N and a categorical variable c which captures the class label of an instance i in ELSA Database. Under the investigating problem, it has two possible states, e.g., $c = "Diabetic" or "1"$ or $c = "NonDiabetic" or "0"$. The features vector of an instance i is denoted as $\boldsymbol{f}_i = (f_{i1}, f_{i2}, \ldots, f_{iF})$.

Our aim is to achieve high sensitivity and Area Under Curve through the supervised machine learning, meaning that the Diabetic class can be predicted correctly. The proposed methodology for T2DM prediction consists of the following steps which are explained in detail below.

1) Data preprocessing

The raw data quality may be degraded either due to missing values and/or noisy and inconsistent data, so the final results-predictions quality may be low as well. Therefore, is necessary, processing, including redundant values reduction, feature selection and discretization of data to make it more appropriate for data mining and analysis.

In the proposed framework, missing or null values were dropped, rather than imputed by the mean values of the attributes as in [48], only for the specific features that are used for the fitting of FINDRISC [46] and Leicester [25] risk inspired models (see Section V), since it is impossible for the logistic regression to reasonably deal with missing values.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

However, in case of ML all of the rest of the selected features were considered as is, given that the missing values can be handled by them.

Also, data is not always in appropriate form to be fed into a machine learning algorithm, e.g. plain text feature values may cause problems during the learning process, or data may be represented in different scales. Hence, feature transformation from one format to another is necessary. Some relevant techniques include the standardization or Z-score normalization which re-scales the attributes for achieving standard normal distribution with zero mean and unit variance. Also, in this research work, several categorical and ordinal features are considered, further details concerning the ordering of the categories and the discretized values for each one are shown in Table 9. Also, another reason for applying features transformation is to reduce the dimension of the features to boost the training stage or improve the accuracy of a specific ML model.

2) Feature Selection

It is common knowledge that, the accuracy of the classifiers improves with the increase of the attributes dimension until the optimal number of features is reached. Adding more features on the same sized training dataset can often lead to classifier performance degradation, which is known as the curse of dimensionality. Ultimately, this indicates that the number of samples an ML model needs to achieve a given level of accuracy should grow exponentially with respect to the number of input features (i.e., dimensionality) to avoid overfitting (inability to generalize). Feature selection constitutes a core component in building accurate and reliable prediction models in machine learning, as it can highly impact the training of the selected model and thus, its performance. Feature selection is defined as the process of identifying the most relevant features in a dataset. This way the most significant or relevant ones are considered, namely, these ones which contribute much to the target variable, with the aim to improve or boost the model accuracy. Such methods can be classified as Filter, Wrapper and Embedded [49].

The Filter category includes information gain, chi-square test, fisher score, correlation coefficient and variance threshold. Among the traditional state-of-art filter methods, Pearson coefficient was selected. Its values vary between -1 (higher negative correlation) and 1 (higher positive correlation) that indicate the linear dependency between two features. Hence, if coefficient value is closer to 0 implies weaker correlation, while zero coefficient value implies no correlation. Pearson coefficient [50], denoted as p_c, is defined as:

$$p_c = \frac{\sum_{i=1}^{M}(f_{im} - \bar{f}_m)(f_{in} - \bar{f}_n)}{\sqrt{\sum_{i=1}^{M}(f_{im} - \bar{f}_m)^2 \sum (f_{in} - \bar{f}_n)^2}} \qquad (1)$$

where $f_{im}, f_{in}\bar{f}_{im}, \bar{f}_{in}$ denote features m, n and mean values of them on dataset, respectively.

The feature selection depends on user defined threshold value about p_c. For example, in diabetes case, haemoglobin help clinicians to estimate the average blood sugar levels over a period of weeks or months thus, p_c is expected to be close to 1, implying that it is highly correlated with blood glucose.

From the Wrapper feature selection methods, a simple and often used is the forward/backward stepwise selection [51]. The former refers to a search that begins with an empty set of features and which are added one by one, while the latter works conversely, i.e., it begins with all features which are removed gradually, one by one. From the Wrapper methods, stepwise backward with Naive Bayes, Logistic Regression and Decision Tree ML models were investigated. Although it is more accurate than the Filter methods, it is computationally expensive, since it applies an iterative greedy search process.

Moreover, the Embedded methods include regularization based techniques with L1 regularization or LASSO (Least Absolute Shrinkage and Selection Operator) and L2 regularization or Ridge be the most representative. These methods have built-in penalization functions to reduce overfitting contrary to Ordinary Least Squares (OLS), which would overfit the data [52]. From the Embedded methods, in the experiments, LASSO method will be applied, due to its simplicity (lower complexity) and better interpretability than Ridge. Consider that, the aim of feature selection is not only to improve the accuracy, but also to increase the interpretability and reduce the complexity and training time of the ML model.

The LASSO or penalized least squares regression with L1-penalty function has the form of

$$Loss = \sum_{j=1}^{M}(y_i - a_0 - \sum_{i=1}^{F} a_i f_{ji})^2 + \lambda \sum_{i=1}^{n} |a_i| \qquad (2)$$

where y is the output (target) variable for the prediction, $f_1, f_2, \ldots, f_F$ are the features that decides the value of y, a_0 is the bias, $a_1, a_2, \ldots, a_F$ are the weights attached to $f_1, f_2, \ldots, f_F$, respectively and λ is the regularization parameter that controls the significance of the regularization term.

The initial features, considered for the training of the ML-based models, included over 100 variables collected from those at the reference waves of ELSA dataset. Also, a group of variables related to the FINDRISC [53] and Leicester questionnaires were included such as, variables representing gender, age, race, physical activity (at least 30 min during the day), fruit and vegetable consumption as well as keeping a track of medical history including the history of antihypertensive drug treatment, history of high blood glucose levels. To evaluate the performance of ML models, feature importance was established using some of the feature selection techniques discussed in Section IV-B2. Moreover, Tables 1 and 9 describe the variables considered in the various feature selection methods.

The features in relation to LASSO, Correlation and Greedy Stepwise with Backward Selection under three different clas-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 1: Features Information

Feature Name	Feature Description	Feature Name	Feature Description
wstval	Valid Mean Waist (cm)	hlthlm	Health problem limits work
chol	Blood total cholesterol level (mmol/l)	adlwa	ADLs (bathe, dress, and eat)
fglu	Glucose level (mmol/l)	lbrfe	Labor force status
sys	Blood pressure systolic reading (mmHg)	finea	Fine motor index: picking up a 5p coin, eating, and dressing activities
hbA1c	Glycated haemoglobin level(%)	physActive	Is physically active
workl65	Self-reported probability of having a work limiting health problem before age 65	raeducl	Education level
dias	Blood pressure diastolic reading (mmHg)	AgeGroup	Belonging age group
workat	Self-reported probability of working full-time after a specific age	jphysa	Level of physical effort at current job
cfood1m	Amount spent weekly on food consumption outside house	jpress	Work stress - under pressure due to workload
liv10	Self-reported probability of living to a specific age	hipe	Ever had hip fracture
drinkde	Days/week drinks	cesd	Mental health: the respondent's feelings much of the time over the week prior to the interview
ldl	LDL level (mmol/l)	mstat	Marital Status
cfoodi	Amount spent weekly on food consumption inside house	arthre	Ever had arthritis
itot	Total family income	work	Currently working for pay
everHighGlu	Ever have high glucose	fcntf	Social activity - weekly contact with friends
bmicat	BMI category	iadlza	IADLs: using the phone, managing money, taking medications, shopping for groceries, preparing hot meals
weight	Weight in (Kg)	Gender	Belonging gender
estwt	Nurse measured weight (Kg)-final estimated	hchole	Ever had high cholesterol
shlt	Self-report of health	rcntf	Social activity - weekly contact with relatives
trig	Triglycerides level (mmol/l)	psyche	Ever had psychological problem
grossa	Gross motor index: walking 100 yards, walking across a room, climbing one flight of stairs, getting in or out of bed, and bathing activities	smoken	Smokes now
hdl	HDL level (mmol/l)	hearte	Ever had heart problems
mobilb	Mobility index: walking 100 yards, walking across a room, climbing one flight of stairs, and climbing several flights of stairs activities	stroke	Ever had stroke
HBP	Ever had high blood pressure	smokev	Smoke ever
hemda	Taking high blood pressure medication	cancre	Ever had cancer
lgmusa	Large muscle index: sitting for 2 hrs, getting up from a chair, stooping, kneeling or crouching, and pushing or pulling large objects activities	parkine	Ever had Parkinson disease
adla	ADLs: bathe, dress, eat, getting in/out of bed, walking across a room	asthma	Ever had asthma disease
relhite	Reliability of standing height according to nurse	catrctf	Ever had cataracts
eatVegFru	Tablespoons ate yesterday	work2	Works at second job
lunge	Ever had lung disease	demene	Ever had dementia
memrye	Ever had memory problems		

sifiers are listed below:

- **LASSO**: wstval, chol, fglu, sys, hbA1c, workl65, dias, workat, cfood1m, liv10, drinkde, ldl, cfoodi, itot
- **Correlation**: hba1c, everHighGlu , wstval, bmi, bmicat, fglu, weight, estwt, shlt, trig , sys, grossa, hdl , mobilb, HBP, sys, drinkde, hemda, lgmusa, adla, hlthlm, adlwa, lbrfe, finea, physActive, raeduc, AgeGroup, ldl, chol, drink, jphysa, jpress, hipe, liv10, cesd, mstat, arthre, cfoodo1m, work, dias, fcntf, iadlza, Gender, hchole, rcntf, psyche, smoken, hearte, stroke, smokev, cancre, parkine, asthma, relhite, catrctf, eatVegFru, work2, lunge, demene, memrye
- **Greedy Stepwise with Logistic Regression (GSW-LR)**: cesd, HBP, AgeGroup, hchole, parkine, hipe, bmicat, weight, physActive, drinkde, smoken, itot, cfoodi, work, wstval, chol, trig, dias,sys, fglu, hba1c, everHighGlu
- **Greedy Stepwise with Naïve Bayes (GSW-NB)**: Race, raeducl, mstat, HBP, AgeGroup, bmicat, physActive, drinkde, smoken, fcntf, work, jphysa, wstval, chol, ldl, trig, sys, fglu, hba1c, hemda, everHighGlu
- **Greedy Stepwise with Decision Trees (GSW-DT)**: mstat, hlthlm, adla, adlwa, lgmusa, finea, cesd, HBP, cancre, lunge, hearte, stroke, psyche, arthre, asthma, hchole, catrctf, bmi, bmicat, physActive, drink, drinkde, smokev, smoken, cfoodi, cfoodo1m, rcntf, fcntf, work, work2, jpress, workl65, estwt, wstval, hdl, ldl, sys, dias, fglu, hba1c, hemda, everHighGlu

All selected ML models were trained with the same features (i.e., risk factors) derived from the GSW-NB feature selection method, excluding the irrelevant by the literature features fcntf and work. As the feature selection process is a highly empirical one, GSW-NB was selected as it shares the most common variables with the rest selection methods, which are also inline with the literature. In addition to these, we also included the variables shlt, hlthlm, mobilb, lgmusa, grossa, finea, hearte, psyche, itot, cfoodo1m, estwt, hdl, dias, eatVegFru and Gender as these capture risk factors or signs that are actually considered in diabetes detection by the literature. The resulting feature set was constructed by the above 34 features plus the ELSA derived class feature rYdiabe which indicates if a subject is actually diabetic.

3) Machine Learning Models

Let recall that, in the context of this work, we investigate the problem of T2DM prediction on ELSA database with various machine learning models. As a first approach, the problem is managed using single classifiers as independent entities. Then, ensemble learning based on majority voting, either weighted or not, and stacking is employed. All of them are compared in order to evaluate the appropriate one for diabetes prediction.

Some well-known classification methods, considered in this work, are Naïve Bayes, Decision Trees [33], Random Forests [54] and Logistic Regression [17], [55]. Finally two

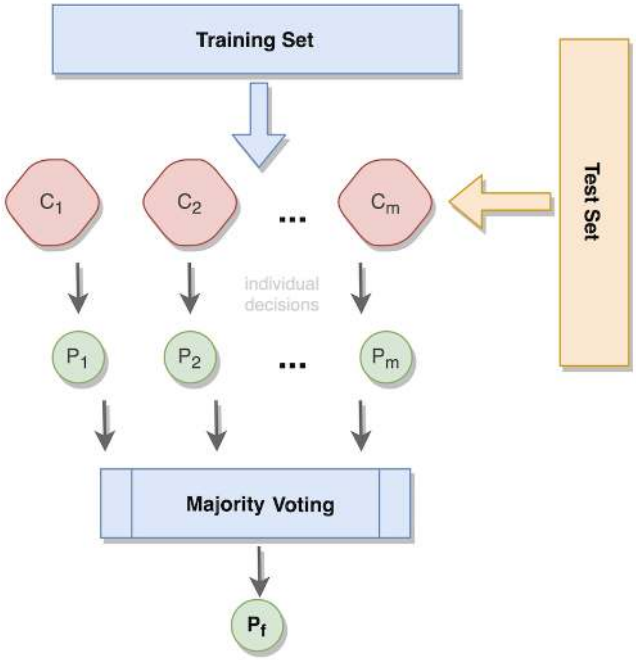

FIGURE 2: Ensemble Learning with Voting.

of them, with similar success according to AUC and Sensitivity metrics, are utilized as base-learners and their outputs are combined to define the final prediction score, adopting different ensemble learning approaches, namely majority voting, weighted voting and stacking. Here, it should be noted that, the key difference between voting and stacking lies in the final aggregation. Although in voting, appropriate weights are utilized to combine the classifiers predictions, in stacking the aggregation is performed by using a meta classifier. In following, useful information about the adopted models will be described.

a: Single Learning

1) **Naive Bayes**: It is a simple but powerful algorithm for classification, since it is based on conditional probability. It is an appropriate solution for unbalanced data and missing values. It uses Bayes theorem to calculate the posterior probability [56] as:

$$P(c|\boldsymbol{f}) = \frac{P(c)P(\boldsymbol{f}|c_j)}{P(\boldsymbol{f})}, \tag{3}$$

where $c = 0, 1$, $P(c|\boldsymbol{f})$ is the Posterior Probability, $P(c)$ is the class Prior Probability, $P(\boldsymbol{f}|c)$ is the Likelihood and $P(\boldsymbol{f})$ is the Predictor Prior Probability.

2) **Decision Trees**: They build classification model in the form of tree structure by breaking dataset into smaller subsets and simultaneously developing the associated decision tree. The decision tree is a top-down structure with one root node, and it splits the branches which have parent–child relationship. The tree includes a root node, some leaf nodes that represent any classes and internal nodes representing test condition.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

3) **Random Forests**: It constitutes a classification method that creates many decision trees on different instances to perform prediction and regression. Each decision tree in RFs will export its own classification result and vote, and then the final output of the RFs will be the one that most trees agree. Moreover, it has a significant role in ensemble machine learning and is commonly applied in various research areas, such as bio-medicine. The final output is computed as

$$\hat{C} = \frac{1}{R}\sum_{r=1}^{R}\hat{C}_r(\boldsymbol{f}), \tag{4}$$

where $\hat{C}$ stands for the final tree prediction; R is the total number of trees, r represents the index of the current decision tree and $\boldsymbol{f}$ is the training instance.

4) **Logistic Regression**: It is a classification algorithm, used for categorical variables in nature and especially when the output of the data is binary. The diabetes model has one binary output variable, in which $p = P(Y = 1)$ denotes the probability an instance to belong in "Diabetics" class, so $1 - p = P(Y = 0)$ stands for the probability an instance to belong in "Non Diabetics" class. The linear relationship between log-odds with base b and model parameters β_i is as follows:

$$log_b(\frac{p}{1-p}) = \beta_0 + \beta_1 f_1 + \ldots + \beta_p f_F \tag{5}$$

b: Ensemble Learning

1) **Majority Voting**: Assuming a set of K ensemble models the output of the ensemble, in simple majority voting (Figure 2), can be outlined with the following equation:

$$\max \sum_{k=1}^{K} P_{k,c}, \tag{6}$$

where $c = 0, 1$. The classification, based on majority voting, can be approached as either hard or soft voting. The former (hard voting) sums the predictions for each class label and predicts the class label with the most votes. The latter (soft voting) sums the predicted probabilities for each class label and predicts the class label with the largest probability. Here, soft voting is adopted. Nonetheless, since the base classifiers in an ensemble may not perform equally well, it would be more efficient to weight each classifier soft vote.As it will be seen next, weighted majority voting is compared with majority voting in terms T2DM long-term risk prediction.

2) **Weighted Majority Voting**: Given $w_1, w_2, , \ldots, w_K$, where $w_k \geq 0$ and $w_k \leq 1$ for $i = 1, 2, \ldots, K$ that represent the weight with which the corresponding classification model contributes to the final output, the final prediction class for each test instance is done based on the highest weighted soft votes.

$$\max \sum_{k=1}^{K} w_k P_{k,c}, \tag{7}$$

where $c = 0, 1$ denotes the label of the corresponding class. The main issue in weighting schemes is how to appropriately determine the optimal weights of the classifiers, which can strongly influence the performance of the ensemble. In this study, the genetic algorithm NSGA-II for multi-objective optimization [57] is considered in order to determine the optimal weights and construct a prediction model with high both AUC and Sensitivity.

3) **Stacking**: It is an ensemble learning technique that employs multiple classification ML models and combines them in a meta-classifier. The base models are trained based on a complete training set, then, the meta-model is trained on the outputs of the base models as features. In the base level, different learning algorithms can be applied and, therefore, stacking ensembles are often heterogeneous. Such an approach is considered in this work. Specifically, the stacking ensemble will consist of Random Forests and Logistic Regression as base classifiers, whose predictions are combined by Random Forests as a meta-classifier.

V. EXPERIMENTATION

A. TRAINING AND TEST DATASET

The training and test dataset for the T2DM risk prediction models is a subset of the ELSA database, which consists of reference waves 2, 4 and 6 as baseline and the respective waves 3, 5, and 7 for the 2-years follow-up assessment. Although the number of participants in ELSA waves selected as reference one (namely waves 2, 4, and 6) is very large, initially we drop out participants that already have diagnosed diabetes at reference waves and participants that did not take the interview at both, the reference and the corresponding follow-up wave. In Tables 2,3, the distributions of selected participants that satisfied the above criteria, per age group are presented.

As shown in Table 2, the distributions of selected participants, however, correspond to an unbalanced dataset, as they do not relate to prevalence of diabetes for these age groups, as they have been reported at country level and at European level. The proportion of older people who have diabetes increases with age: 9% of people aged 45 to 54 have diabetes, but for over 75s the percentage increases to aproximately 24%. Taking into account these findings, we balanced the dataset using random undersampling [58] in order to reach a 9%, 12%, 15%, 18%, 21% and respectively 24% of participants with diabetes at the 2-years follow-up for the selected age groups.

The demographics and some health-related characteristics of the participants per age group and gender in the balanced

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 2: Distribution per age group of newly diagnosed diabetes at 2-years follow-up in the original dataset.

	T2DM	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74	75+	Total
Ref Wave 2	No	473	1,181	916	852	659	894	4,975
F-up 3	Yes	3	17	23	16	19	14	92
Ref wave 4	No	695	1,236	1,362	953	879	931	6,056
F-up wave 5	Yes	9	23	27	23	17	21	120
Ref wave 6	No	389	927	1,209	1,107	806	1,140	5,578
F-up wave 7	Yes	6	13	34	19	14	24	110
All waves	No	1,557	3,344	3,487	2,912	2,344	2,965	16,609
	Yes	18	53	84	58	50	59	322

TABLE 3: Distribution per age group of newly diagnosed diabetes at 2-years follow-up in the balanced dataset.

	T2DM	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74	75+	Total
Ref Wave 2	No	33	142	153	89	90	58	565
F-up 3	Yes	3	17	23	16	19	14	92
Ref wave 4	No	100	192	180	128	81	88	769
F-up wave 5	Yes	9	23	27	23	17	21	120
Ref wave 6	No	67	108	227	106	67	100	675
F-up wave 7	Yes	6	13	34	19	14	24	110
All waves	No	200	442	560	323	238	246	2,009
	Yes	18	53	84	58	50	59	322

dataset are summarized in Table 4. In addition, independent group t-tests were run wherever applicable, comparing the mean scores between the different groups. Of the 2009 participants, 53.4% were women of whom 13.8% identified as diabetic in the follow-up, the same indicator in males was 18.6%. Note that, 14.3% of participants had high education and just 11.8% had physical effort at work. Focusing on those who were diagnosed with diabetes in a follow-up wave, 29.2% are employed, 11.2% had physical effort at work, 79.8% stated that they were physically active and 64.0% were diagnosed with high blood glucose at least once. Moreover, concerning diabetics and irrespective of gender, they had average BMI of 31.7 kg/m^2 and waist size of 106.46 cm. P-values showed that the difference between men and women was statistically significant at the level of 0.93 for age and 0.69 for BMI. Also, the statistical significance in terms of variables cholesterol, drinker and waist was at level of zero, 0.0022 and 0.001 for food consumption outside home and income variables, respectively. In comparison with non diabetics, diabetics had higher overall means for age, BMI, waist and income characteristics, and the differences were significant at the 0 level for variables age, BMI, food outside home, cholesterol, drinker and waist except income.

B. T2DM MODELING AND RESULTS

The different single and ensemble classification models that were presented on the previous sections were compared in a series of exhaustive experiments in order to identify the most effective models regarding the classification of T2DM on the constructed dataset of 2009 instances as depicted in Table 3. Moreover, the comparisons included the four benchmark models, the logistic regression models based on the corresponding works of Leicester and FINDRISC score systems and two neural network models utilizing the architectures discussed in [29]. Furthermore, an optimized voting ensemble (WeightedVotingLRRFs) was also considered in the comparisons and is discussed on the last paragraphs of the section.

The experimentation methodology can be summarized by the following steps:

- Data preprocessing as elaborated in Section IV-B.
- Divide the constructed dataset based on ELSA database using the standard technique of stratified train-test split procedure with *10*-times random repeat, thus preserving the class proportions of the original dataset and ensuring that the sub-datasets are representative (random samples) by the use of different seeds in the repeating

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 4: Overview of Demographic and Health-related Features.

Features	Total	Male	Female	P-value	Non-Diabetic (follow-up wave)	Diabetic (follow-up wave)	P-value
Age (years)	64.04082 ±.1793756	64.02241 ±.2571762	64.0569 ±.2500601	0.9236	63.60166 ±.1937851	66.3416 ±.4502968	0
Gender (N)	2009	937	1072	-	763(M) 924(F)	174(M) 148(F)	-
BMI ($\frac{kg}{m^2}$)	28.43476 ±.1151526	28.48427 ±.1486709	28.39149 ±.172347	0.6878	27.80501 ±.1173959	31.7341 ±.312884	0
Food Outside House (money/week)	53.90977 ±1.638046	59.65021 ±2.700247	48.87676 ±1.949294	0.001	55.8782 ±1.822772	43.60625 ±3.592217	0
Cholesterol ($\frac{mmol}{L}$)	5.806188 ±.0299581	5.55391 ±.0437444	6.026304 ±.0396412	0	5.859203 ±.0316898	5.496186 ±.0854177	0
Ever had high blood glucose (N)	512	252	260	-	306	206	-
Drinking (days/week)	2.62197 ±.0566743	3.055006 ±.083562	2.239806 ±.0751508	0	2.785408 ±.0619196	1.756494 ±.1301927	0
Education Level High (N)	287	166	121	-	262	25	-
Waist (cm)	97.01553 ±.3045124	102.5559 ±.3928172	92.1729 ±.4012403	0	95.21257 ±.3141611	106.4615 ±.7563727	0
Married (N)	1473	715	659	-	1161	213	-
Physical Effort at Work (N)	238	160	78	-	202	36	-
Employed (N)	775	420	355	-	681	94	-
Income (Couple Level)	27141.72 ±705.8337	29449.91 ±1127.65	25129.23 ±878.4182	0.0022	26827.44 ±522.9635	28793.44 ±3458.923	0.3075
Physically Active (N)	1790	811	979	-	1533	257	-

process. The 70% and 30% of the data are chosen as training dataset and testing dataset each time.

- Application of the selected ML models, single and ensemble by either voting or stacking methods. These models use the selected features as independent variables and the diabetes risk status as output variable.
- Performance measures estimation.

As regards the software tools that were employed for the implementation of the compared models, the Java Weka [59] library and the Python Statsmodels [60] were considered, as they are both open-source, making it possible to integrate the implemented models in the deployable solution in the context of the SmartWork project.

A number of measures are recorded for evaluating the performance of ML models. The most commonly used in literature [61], [62], [63] which will be considered as well in our analysis, are the following:

Sensitivity (True Positive Rate) corresponds to the proportion of participants that have T2DM (e.g., positive data instances) that are correctly considered as positive, with respect to all positive participants.

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

Specificity (True Negative Rate) corresponds to the proportion of participants that don't have T2DM (e.g.,negative data instances) that are correctly considered as negative, with respect to all negative participants.

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

Positive Predictive Value (+PV) corresponds to the proportion of participants that have T2DM (e.g., true data instances) that are correctly considered as positive, with respect to all positively predicted participants.

$$+PV = \frac{TP}{TP + FP} \tag{10}$$

Negative Predictive Value (-PV) corresponds to the proportion of participants that don't have T2DM (e.g., negative data instances) that are correctly considered as negative, with respect to all negatively predicted participants.

$$-PV = \frac{TN}{TN + FN} \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

Positive Likelihood Ratio (+LR) is defined as the ratio of the true positive rate (sensitivity) to the false positive rate (1–specificity).

$$+LV = \frac{Sensitivity}{1 - Specificity} \tag{12}$$

Negative Likelihood Ratio (-LR) is defined as the ratio of the false negative rate (1-sensitivity) to the true negative rate (specificity).

$$-LV = \frac{1 - Sensitivity}{Specificity} \tag{13}$$

Likelihood ratios measure the certainty of the test about a positive and negative diagnosis, correspondingly. Indicate that, in previous equations TP:True Positive, TN:True Negative, FP:False Positive and FN:False Negative.

Another useful metric is Area Under Curve, which takes values in the range $[0, 1]$. The higher its value, the better is the ML model performance in distinguishing positive (Diabetics) from negative (Non Diabetics) class instances. In best (ideal) case where AUC equals 1, the ML model can perfectly distinguish all positive (Diabetic) from negative (Non Diabetic) class instances. In worst case where AUC equals 0, the classifier will predict all negatives as positives and vice versa. Also, the Younden Index was considered in combination with Receiver Operating Characteristic (ROC) analysis. This metric summarises the performance of a diagnostic test, it is defined for all points of a ROC curve, and its maximum value may be used for the selection of the optimum cut-off point.

$$J = Sensitivity + Specificity - 1, J \in [0, 1]. \tag{14}$$

The quantitative analysis of the two selected risk score systems showed that the best performing, according to AUC metric, is FindLogist with AUC equals 0.821 which proves that it performs better than LeicLogist with AUC 0.788 by 3.3% in the constructed dataset. Although the sensitivity and specificity of the selected risk score systems were not considerable better than others (Table 5), if combined with other existing ones, may improve the performance of the ensemble methods.

Moreover, the use of single classifiers and ensemble methods, such as voting and stacking, could overcome the limitations of risk score systems in order to build a single or combined reliable T2DM risk assessment system. Figure 3 and Table 5 summarize the performance metrics values for the diabetes prediction according to the adopted ML models described in Section IV-B3. Also, in the same figure and table, respectively, for the same metrics, the results of FindLogist and LeicLogist models have been recorded. Note again that, these systems apply Logistic Regression with specific features less than those considered in the ML models.

To further investigate the performance of the ML models, we compare the Youden indices and AUCs. The results unveiled that the selected voting methods performed not only the best but also considerably better than all the ML models and the two selected score systems. Among

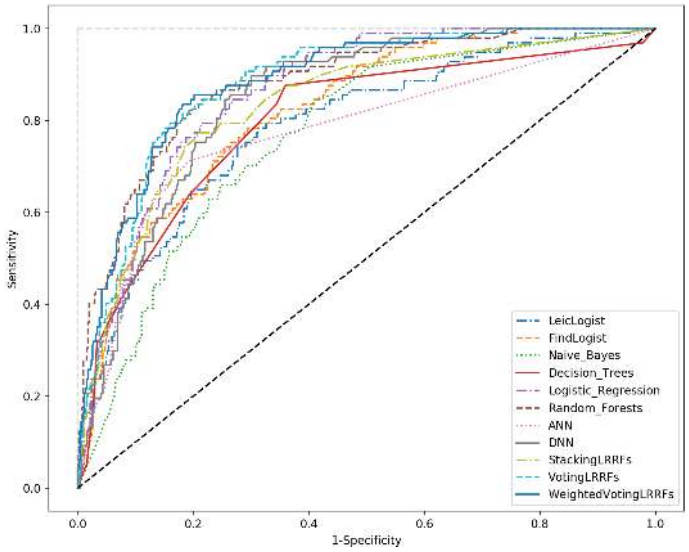

FIGURE 3: AUC-ROC behavior: Inductive Learning.

the different combination methods, the superiority of the two voting methods against stacking was revealed. Voting typically works well if the base classifiers perform the same task and have comparable success, although stacking works well for different types of first-level classifiers. A comparison of sensitivities and specificities for different ML models can be found in Table 5, while the exact hyperparameters of the models can be seen in 6.

The significance of the classifiers' AUCs was tested using the Wald test statistic [64]. In detail, the discrimination ability of each classifier is tested compared to a classifier with random chance discrimination ability ($TPR = FPR$ i.e. $AUC = 0.5$). The utilized null hypothesis states that $AUC = 0.5$ and the alternative hypothesis that $AUC \neq 0.5$. The calculated p-values for all the models were equal to 0 (<0.05), thus clearly indicating that the calculated AUCs are significant using a level $a = 0.05$, with the lower AUC recorded being equal to 0.727.

Additionally, the receiver operating characteristic (ROC) curves for the ML models and the score systems are summarized in Figure 3. Focusing on the combination methods, we again conclude that the voting algorithms with the selected single models produced again the best performance (prediction result) against stacking method. Here, it should be pointed out that, the ROC curves produced by the voting algorithms are similar and are also positioned above the rest model curves.

As the results witness, Random Forests classifier is the best performing among the rest single classifiers with Logistic Regression's performance being closer, than the rest models. This lies in the fact that the Random Forests can learn a non-linear decision boundary and thus can achieve higher scores in all metrics. In other words, Logistic Regression poorly segments the Diabetes and No Diabetes classes while the Random Forests model learns a more flexible decision

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 5: Performance comparison of different prediction models (inductive results).

	AUC	Sensitivity	Specificity	+PV	-PV	+LR	-LR	Cut-off	J	P-value
LeicLogist	0.788 (0.738,0.838)	0.784 (0.688,0.861)	0.688 (0.645,0.728)	0.325 (0.284,0.451)	0.943 (0.91,0.953)	2.509 (2.125,2.963)	0.315 (0.215,0.462)	0.151	0.471	0
FindLogist	0.821 (0.780,0.863)	0.742 (0.643,0.826)	0.747 (0.707,0.784)	0.360 (0.315,0.481)	0.938 (0.905,0.949)	2.934 (2.426,3.549)	0.345 (0.245,0.485)	0.176	0.489	0
Naive Bayes	0.766 (0.719,0.814)	0.845 (0.758,0.911)	0.591 (0.547,0.634)	0.284 (0.249,0.425)	0.952 (0.919,0.96)	2.066 (1.806,2.365)	0.262 (0.163,0.419)	0.003	0.436	0
Decision Trees	0.797 (0.747,0.847)	0.876 (0.794,0.934)	0.640 (0.597,0.682)	0.318 (0.280, 0.484)	0.964 (0.936, 0.97)	2.436 (2.122,2.797)	0.193 (0.113,0.329)	0.071	0.517	0
Logistic Regression	0.863 (0.830,0.896)	0.794 (0.700,0.869)	0.787 (0.748,0.821)	0.416 (0.365,0.552)	0.952 (0.923,0.961)	3.719 (3.058,4.523)	0.262 (0.177,0.388)	0.175	0.580	0
Random Forests	0.880 (0.844,0.916)	0.845 (0.758,0.911)	0.785 (0.746,0.820)	0.429 (0.378,0.584)	0.964 (0.938,0.971)	3.924 (3.256,4.730)	0.197 (0.123,0.315)	0.180	0.629	0
ANN	0.776 (0.725,0.827)	0.711 (0.610,0.799)	0.808 (0.771,0.842)	0.416 (0.363,0.534)	0.936 (0.903,0.949)	3.711 (2.980,4.620)	0.357 (0.261,0.489)	0.001	0.519	0
DNN	0.847 (0.811,0.882)	0.897 (0.819,0.949)	0.700 (0.658,0.739)	0.364 (0.321,0.553)	0.973 (0.948,0.977)	2.986 (2.572,3.466)	0.147 (0.082,0.266)	0.111	0.596	0
Stacking: LR,RFs	0.833 (0.789,0.877)	0.773 (0.677,0.852)	0.792 (0.755,0.827)	0.417 (0.365,0.547)	0.948 (0.918,0.958)	3.726 (3.046,4.558)	0.286 (0.198,0.414)	0.190	0.566	0
Voting: LR,RFs	0.881 (0.849,0.913)	0.794 (0.700,0.869)	0.840 (0.805,0.871)	0.487 (0.428,0.621)	0.955 (0.928,0.965)	4.959 (3.964,6.203)	0.245 (0.166,0.363)	0.242	0.634	0
Weighted Voting: LR,RFs	0.884 (0.850,0.918)	0.856 (0.770,0.919)	0.798 (0.761,0.833)	0.449 (0.395,0.608)	0.967 (0.942,0.973)	4.245 (3.504,5.142)	0.181 (0.111,0.294)	0.193	0.654	0

boundary for the discrimination of instances of the two classes [65].

Among the three different ensembling approaches, the weighted voting scheme boosts the performance of diabetes prediction. The optimal weights are calculated by running the NSGA-II algorithm on the constructed dataset. The optimization procedure aims to maximize both AUC and Sensitivity. The relevant Pareto Front behavior is depicted in Table 7. Note that the sensitivities reported in the first column of Table 7 were significantly lower than the final reported in the inductive results table due to the fact that the Youden criterion was not utilized during the optimization process, and the default cut-off point of 0.5 probability was set. All the weight sets were applied in the inductive experimentation setup of WeightedVotingLRRFs using the Youden optimal cut-off criterion (displayed in the last two columns of Table 7) and the weight set of $[0.2733, 0.7266]$ was found to yield the best performance results in terms of AUC and Sensitivity, thus its performance was recorded in Table 5.

A more focused graphic analysis of the different evaluation metrics for WeightedVotingLRRFs is found in Figure 4, where its ROC curve, Sensitivity-Specificity and Distribution graphs are presented. In the first graph, the specific Youden optimal cut-off point is located on the ROC curve. In the second graph, the sensitivity and specificity curves are depicted showing the trade off for the different selections of cut-off points. The next two graphs, give a good overview of how well the Youden optimal cut-off point of 0.193 separates the two classes.

In addition to the inductive experiments, transductive learning [66] experiments were also employed. The aim of this learning approach is to exploit patterns that are hidden in the test samples by utilizing them as unlabeled data in the training phase, thus taking advantage of the information embedded in the test set by augmenting the training set [67], [68]. During the transductive experimentation, the partitioning of the dataset was kept the same as in the inductive experimentation, while the unlabeled set was used under a common self-training wrapper algorithm using the different prediction models that were compared in the inductive experiments. The performance results are summarized in Table 8 and Figure 5, while the exact parameters utilized in the self-training scheme can be found in Table 6. Similarly with the work of Triguero et al. [69], the transductive self-training wrapper uses as base classifiers the compared models which are initially trained using the labeled set and are then used to predict the labels of the unlabeled set in order to repeatedly increase the labeled set, while in each iteration the base model is being retrained. A confidence probability threshold of 0.90 for the predicted labels is set to ensure that less

TABLE 6: Models Hyperparameters.

Algorithm	Parameters
Naive Bayes	Kernel Estimator = None
Decision Trees	Tree pruning = True Confidence factor used for pruning = 0.25 Min. number of instances per leaf = 2 Min. description length correction = True
Random Forests	Size of each bag = 100% Maximum tree depth = Unlimited Number of iterations = 100
ANN	Input layer units = 66 Hidden layer units = 8 (x2) Hidden and output layers type = Sigmoid Hidden layers momentum = 0.2 Optimizer = SGD Learning rate = 0.1 Epochs = 500
DNN	Input layer units = 66 Hidden layer units = 50 (x2) Hidden layers type = Rectifier Hidden layers dropout = 10% Hidden layers L1 = 0.00001 Hidden layers L2 = 0 Output layer type = Softmax Optimizer = SGD Learning rate = 0.1 Epochs = 500
Stacking	Stacking models = LR, RFs Meta-classifier = RFs Number of execution slots = 1
Voting	Voting models = LR, RFs Combination rule = Avg. of Probabilities Number of execution slots = 1
Weighted Voting	Voting models = LR, RFs Weights = 0.2733, 0.7266 Combination rule = Avg. of Probabilities Number of execution slots = 1
Transductive Self-training Wrapper	Selection metric = Prediction probabilities Confidence threshold = 0.90 Maximum iterations = 10

TABLE 7: Weighted Voting with NSGA-II algorithm.

Sensitivity	AUC	Weight LR	Weight RFs	Youden Sensitivity	Cut-off
0.2989	0.88381	0.2733	0.7266	0.856	0.193
0.3402	0.88344	0.1276	0.8723	0.825	0.214
0.3298	0.88364	0.1325	0.8674	0.835	0.197
0.3195	0.88375	0.1373	0.8626	0.845	0.188
0.3195	0.88375	0.1387	0.8612	0.845	0.188

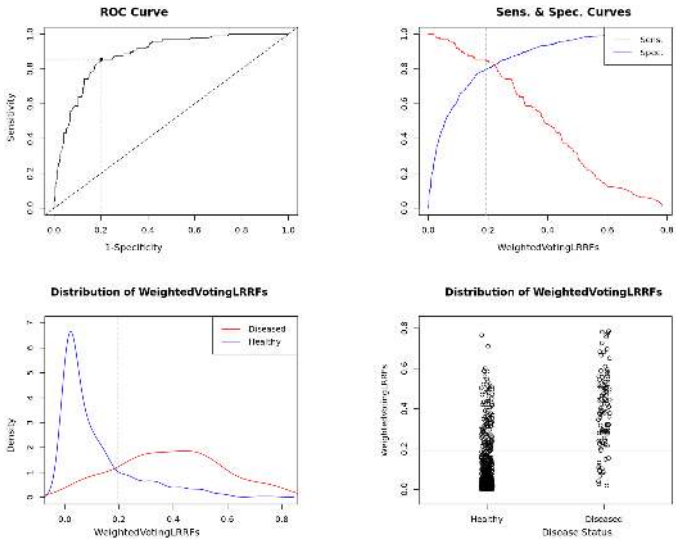

FIGURE 4: Analysis of the ROC curve, optimal cut-off and distribution graphs for the WeightedVotingLRRFs model.

confident predictions are not integrated in the retraining of the base model, and moreover the maximum iterations of the self-training scheme are limited to 10. By comparing the transductive AUCs against their inductive equivalents, it is concluded that the logistic models, while do not significantly decrease their performance, they gain no benefit from the exploitation of the unlabeled data. The same stands true for the single classifiers i.e. NB, DT and ANN. In contrast, the more complex models such as the RFs, DNN and the rest ensembles marginally improve their classification performance. Specifically, the proposed WeightedVotingLRRFs model scores an $AUC_{\text{transductive}} = 0.888$ which is the highest that was recorded, suggesting that strict selection of unlabeled data (due to voting) can lead to possible performance increase of the model.

VI. DISCUSSION

In this research, several strengths and limitations are highlighted. In terms of the former, to our knowledge, it is the first to assess various ML models and provide participants with personalized long-term risk prediction of T2DM occurrence and appropriate guidance regarding lifestyle interventions. Also, the research findings were derived from a cross-sectional study on a representative English cohort (e.g., elderly office workers) with follow-up data; thus, we may identify causal and temporal associations between elderly lifestyle and T2DM.

Another positive aspect of this work is that, during the balanced dataset creation, we drew instances of the initially "Non-Diabetics" class from the reference waves, whose class label was finally defined in the follow-up waves. This approach may give us a view of features behaviour for participants diagnosed with T2DM in the follow-up examination, contributing to T2DM prognosis. Moreover, our study

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 8: Performance comparison of different prediction models (transductive self-training results).

	AUC	Sensitivity	Specificity	+PV	-PV	+LR	-LR	Cut-off	J	P-value
LeicLogist	0.788 (0.739,0.838)	0.784 (0.688,0.861)	0.692 (0.649,0.732)	0.328 (0.287,0.454)	0.943 (0.91,0.953)	2.541 (2.15,3.004)	0.313 (0.213,0.459)	0.145	0.475	0
FindLogist	0.821 (0.780,0.863)	0.753 (0.655,0.835)	0.743 (0.703,0.781)	0.360 (0.315,0.482)	0.940 (0.907,0.951)	2.929 (2.43,3.532)	0.333 (0.234,0.473)	0.167	0.495	0
Naive Bayes	0.727 (0.675,0.779)	0.825 (0.734,0.894)	0.561 (0.517,0.605)	0.265 (0.232,0.394)	0.944 (0.907,0.952)	1.88 (1.643,2.151)	0.312 (0.201,0.484)	0.001	0.386	0
Decision Trees	0.788 (0.734,0.843)	0.856 (0.770,0.919)	0.644 (0.601,0.686)	0.316 (0.277,0.468)	0.959 (0.929,0.966)	2.405 (2.085,2.775)	0.224 (0.137,0.365)	0.059	0.499	0
Logistic Regression	0.863 (0.830,0.896)	0.794 (0.7,0.869)	0.787 (0.748,0.821)	0.416 (0.365,0.552)	0.952 (0.923,0.961)	3.719 (3.058,4.523)	0.262 (0.177,0.388)	0.173	0.580	0
Random Forests	0.886 (0.850,0.922)	0.866 (0.782,0.927)	0.796 (0.759,0.831)	0.449 (0.396,0.615)	0.969 (0.945,0.975)	4.254 (3.521,5.141)	0.168 (0.101,0.28)	0.175	0.662	0
ANN	0.763 (0.712,0.815)	0.701 (0.6,0.79)	0.8 (0.763,0.834)	0.402 (0.351,0.519)	0.933 (0.899,0.946)	3.512 (2.825,4.366)	0.374 (0.275,0.508)	0.001	0.501	0
DNN	0.852 (0.818,0.887)	0.887 (0.806,0.942)	0.719 (0.678,0.758)	0.377 (0.332,0.557)	0.971 (0.946,0.976)	3.159 (2.701,3.695)	0.158 (0.09,0.276)	0.089	0.606	0
Stacking: LR,RFs	0.857 (0.817,0.898)	0.763 (0.666,0.843)	0.844 (0.809,0.874)	0.484 (0.424,0.611)	0.949 (0.92,0.96)	4.886 (3.879,6.156)	0.281 (0.196,0.402)	0.220	0.607	0
Voting: LR,RFs	0.885 (0.853,0.916)	0.876 (0.794,0.934)	0.773 (0.734,0.809)	0.425 (0.375,0.598)	0.970 (0.947,0.976)	3.856 (3.230,4.603)	0.160 (0.094,0.272)	0.162	0.649	0
Weighted Voting: LR,RFs	0.888 (0.856,0.92)	0.825 (0.734,0.894)	0.846 (0.811,0.876)	0.506 (0.446,0.649)	0.962 (0.937,0.97)	5.35 (4.278,6.692)	0.207 (0.134,0.32)	0.212	0.670	0

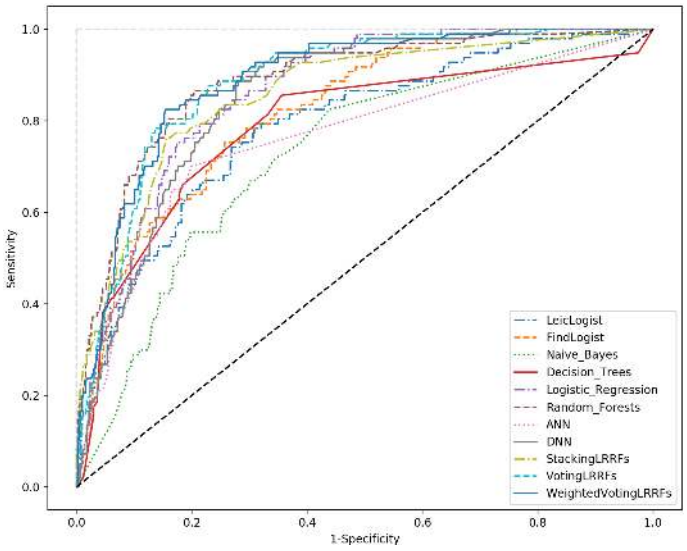

FIGURE 5: AUC-ROC behavior:Transductive Learning.

revealed the importance of different risk factors in T2DM prediction for elder persons. The results of feature selection techniques coincided with the corresponding literature about T2DM risk factors. The selected features for the ML models training and testing are among the symptoms/factors that doctors consider for quantifying long-term risk prediction or identifying its occurrence.

Featurewise, all models were trained using the selected 35 features as described in section IV-B2 except the LeicLogist and FindLogist models. Those two models were fitted using the constructed dataset based on the feature sets according to the original Leicester and FINDRISC score systems, excluding the feature that considers the family history of diabetes as it was not available in the ELSA database. Both logistic models where significant at a level of 0.05 and their analysis (supplementary Figures S1 and S2) confirmed that almost all the features from the original research works were still significant on the constructed dataset. Unlike existing researches [70], [71], for the training of the ML models, family history of diabetes and women with gestational diabetes were excluded from the features set. This may be a limitation of this study since these factors are among the important ones for T2DM risk prediction. Nevertheless, they were not available in the current dataset.

Moreover, contrary to previous works of [48], [72], [73], which use the Pima Indian Diabetes Dataset (PIDD) as benchmark dataset for their experiments, in this study the ELSA dataset is utilized, consisting of elder office workers' data. Furthermore, Perveen et al. [33] examined the Canadian

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

Primary Care Sentinel Surveillance Network (CPCSSN), while Dalakleidi et al. [73] evaluated the suggested models on Hippokration dataset, which was granted from the General Hippokrateion Hospital of Athens.

As far as classification is concerned, k-NN, Decision Trees, Random Forests, Naive Bayes [74], ANN and DNN [75] are the most frequently applied for long-term risk prediction of T2DM. The ANN and DNN topologies presented in [29] were kept identical in order to draw useful comparison results regarding the performance of neural networks on the constructed dataset, with the exception of the insertion of dropout [76] in the DNN topology to reduce overfitting. The results were promising for the DNN model in both the experimentation setups, but were still lacking an approximate 3.7% in terms of $AUC_{\text{inductive}}$ due to significant underperformance in terms of specificity. Considering the performance results of the LeicLogist and FindLogist, the compared metrics suggest similar predictive ability with the rest single classifiers (i.e. NB, DT, ANN). Although, LeicLogist and FindLogist are based on logistic regression, they present far lower AUCs than the LR model trained using the 35 features, thus strengthening the argument that a more personalized approach on the T2DM modeling and prediction can be significantly better.

More to the point, in contrast with [48], Adaptive Boosting (AdaBoost) and Extreme Gradient Boosting (XGBoost) are left for future experimentation on the constructed dataset. Also, in [48], the weighted ensembling of different ML models is proposed where AUC is maximized during hyperparameter tuning using the grid search technique. However, in our analysis, a bi-objective genetic algorithm is applied; the optimal weights are estimated to maximize AUC and Sensitivity of the ML based models simultaneously, under the weigthed voting ensemble. To identify the best performing model, different performance metrics such as sensitivity, specificity and the receiver operating curves were analysed.

The proposed WeightedVotingLRRFs model provides a mechanism of more confident prediction probabilities due to the ensembling of its base models. It is known that an ensemble, such as the proposed, can produce steadily better predictive results than its counterparts under the condition that its base classifiers are accurate and diverse [77]. Both conditions hold true for the proposed model, while the experimentation results validate the assumption of increased predictive ability for the WeightedVotingLRRFs.

To our knowledge, it is the first paper to assess T2DM risk prediction on English cohort (namely, elder office workers and T2DM) from ELSA database. There is a lack of studies to fairly compare it with the previous research, in terms of ML models performance. Previous works in the same dataset mainly focus on diabetes risk factors analysis. Specifically, in [78], the authors found that T2DM diagnosis in older adults did not motivate them changing their health behaviour, other than smoking. Moreover, Hacket et al. in [79] demonstrated associations between sleep problems and daily cortisol levels in response to stress in a part of people with T2DM from ELSA. Moreover, the study in [80], aimed to build a predictive model using RFs, Deep learning and Linear models to accurately estimate health status based on sociodemographic characteristics, in aging populations using data from the ELSA database.

At this point, a limitation of this study is that the experiments have been conducted with a fixed size dataset consisted of a limited number of subjects amount to 2009, as shown in Table 3. It is worth noting that the performance of a ML model improves as the number of training samples increases, as was also observed by the transductive experimentation on the current dataset. To tackle this limitation, we aim to conduct similar research from a big data viewpoint focusing on more and different ML models, evaluating the impact of data volume on their performance in terms of T2DM risk prediction.

VII. CONCLUSIONS

In this study, we applied different ensemble algorithms to a dataset constructed based on the ELSA database, combining different families of ML models to predict the risk of T2DM, taking into account lifestyle variables of elder office workers. Moreover, an IoT enabled framework [81] was developed that integrates the long-term T2DM risk prediction model. It aims to provide personalized interventions according to the users needs. Our empirical study showed that all investigated ML algorithms could produce satisfactory prediction results that are at significantly better than the existing simple score systems. In particular, the voting method could significantly increase the predictability in relation to any conventional risk score system.

It is worth to note that, we chose a multi-objective optimization based technique since it is more robust compared to the single objective one and constructs more efficiently the classifier ensemble (WeightedVotingLRRFs), as it optimizes more than one classification quality measures i.e. AUC and Sensitivity simultaneously, resulting in the highest compared $AUC_{\text{inductive}} = 0.884$.

To sum up, according to our experimental analysis and results, ensemble methods constitute a useful tool for predicting type 2 diabetes. Overall performance attained by the investigated techniques shows the effectiveness and superiority of the multi-objective optimization based, weighted voting ensemble method in relation to single classifiers and risk score systems, while the better learning ability of WeightedVotingLRRFs against its rivals was observed using inductive and transductive learning setups. Hence, embedding it in the recommended system, lifestyle or medication interventions can be implemented to participants at high risk in order to prevent and/or delay diabetes occurrence.

As future work, at first, it would be beneficial to apply different techniques for handling of missing values such as [82] and experiment with even more feature selection techniques. Moreover, it would be interesting to evaluate the impact of dimentionality reduction with techniques such as principal component analysis [83] in T2DM prediction

performance under the ELSA-based constructed dataset. In addition, the comparison of state of the art techniques such as XGBoost, AdaBoost or high layer DNNs would probably provide better insights regarding the predictive limitations of the constructed dataset. Finally, the exploitation of semi-supervised and unsupserivsed methodologies in the training process could also be proven beneficial, as was also suggested by the AUC improvements observed during the transductive experimentation. The previous argument is strengthened by taking into account that there are plenty of unlabeled instances in ELSA that could be incorporated in the constructed dataset.

REFERENCES

[1] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," Artificial intelligence in medicine, vol. 98, pp. 109–134, 2019.

[2] Y.-F. Du, H.-Y. Ou, E. A. Beverly, and C.-J. Chiu, "Achieving glycemic control in elderly patients with type 2 diabetes: a critical comparison of current options," Clinical interventions in aging, vol. 9, p. 1963, 2014.

[3] Y. Gao, Y. Xiao, R. Miao, J. Zhao, M. Cui, G. Huang, and M. Fei, "The prevalence of mild cognitive impairment with type 2 diabetes mellitus among elderly people in china: a cross-sectional study," Archives of Gerontology and Geriatrics, vol. 62, pp. 138–142, 2016.

[4] A. Ramachandran, "Know the signs and symptoms of diabetes," The Indian journal of medical research, vol. 140, no. 5, p. 579, 2014.

[5] D. Mellitus, "Diagnosis and classification of diabetes mellitus," Diabetes care, vol. 28, no. S37, pp. S5–S10, 2005.

[6] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," Computational and structural biotechnology journal, vol. 15, pp. 104–116, 2017.

[7] G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, "Internet of medical things (iomt): applications, benefits and future challenges in healthcare domain," J Commun, vol. 12, no. 4, pp. 240–247, 2017.

[8] S. Nousias, A. S. Lalos, G. Arvanitis, K. Moustakas, T. Tsirelis, D. Kikidis, K. Votis, and D. Tzovaras, "An mhealth system for monitoring medication adherence in obstructive respiratory diseases using content based audio classification," IEEE Access, vol. 6, pp. 11 871–11 882, 2018.

[9] O. Kocsis, A. Lalos, G. Arvanitis, and K. Moustakas, "Multi-model short-term prediction schema for mhealth empowering asthma self-management," Electronic Notes in Theoretical Computer Science, vol. 343, pp. 3–17, 2019.

[10] S. El-Sappagh, F. Ali, S. El-Masri, K. Kim, A. Ali, and K.-S. Kwak, "Mobile health technologies for diabetes mellitus: current state and future challenges," IEEE Access, vol. 7, pp. 21 917–21 947, 2018.

[11] L. Syed, S. Jabeen, S. Manimala, and A. Alsaeedi, "Smart healthcare framework for ambient assisted living using iomt and big data analytics techniques," Future Generation Computer Systems, vol. 101, pp. 136–151, 2019.

[12] S. Alian, J. Li, and V. Pandey, "A personalized recommendation system to support diabetes self-management for american indians," IEEE Access, vol. 6, pp. 73 041–73 051, 2018.

[13] S. Vishnu, S. J. Ramson, and R. Jegan, "Internet of medical things (iomt)-an overview," in 2020 5th International Conference on Devices, Circuits and Systems (ICDCS). IEEE, 2020, pp. 101–104.

[14] G. Shmueli and O. R. Koppius, "Predictive analytics in information systems research," MIS quarterly, pp. 553–572, 2011.

[15] N. Mishra and S. Silakari, "Predictive analytics: A survey, trends, applications, oppurtunities & challenges," International Journal of Computer Science and Information Technologies, vol. 3, no. 3, pp. 4434–4438, 2012.

[16] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," in 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2017, pp. 492–499.

[17] R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, "Prediction and diagnosis of future diabetes risk: a machine learning approach," SN Applied Sciences, vol. 1, no. 9, p. 1112, 2019.

[18] Y. Liu, S. Ye, X. Xiao, C. Sun, G. Wang, G. Wang, and B. Zhang, "Machine learning for tuning, selection, and ensemble of multiple risk scores for predicting type 2 diabetes," Risk Management and Healthcare Policy, vol. 12, p. 189, 2019.

[19] S. Chen, D. Bergman, K. Miller, A. Kavanagh, J. Frownfelter, and J. Showalter, "Using applied machine learning to predict healthcare utilization based on socioeconomic determinants of care," The American Journal of Managed Care, vol. 26, no. 1, pp. 26–31, 2020.

[20] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in 2016 Management and Innovation Technology International Conference (MITicon). IEEE, 2016, pp. MIT–80.

[21] I. Konstantoulas, O. Kocsis, N. Fakotakis, and K. Moustakas, "An approach for continuous sleep quality monitoring integrated in the smartwork system," in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2020, pp. 1968–1971.

[22] A. Bernabe-Ortiz, P. Perel, J. J. Miranda, and L. Smeeth, "Diagnostic accuracy of the finnish diabetes risk score (findrisc) for undiagnosed t2dm in peruvian population," Primary care diabetes, vol. 12, no. 6, pp. 517–525, 2018.

[23] L. Chen, D. J. Magliano, B. Balkau, S. Colagiuri, P. Z. Zimmet, A. M. Tonkin, P. Mitchell, P. J. Phillips, and J. E. Shaw, "Ausdrisk: an australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures," Medical Journal of Australia, vol. 192, no. 4, pp. 197–202, 2010.

[24] A. D. Association et al., "2. classification and diagnosis of diabetes: Standards of medical care in diabetes—2020," Diabetes care, vol. 43, no. Supplement 1, pp. S14–S31, 2020.

[25] L. Gray, N. Taub, K. Khunti, E. Gardiner, S. Hiles, D. Webb, B. Srinivasan, and M. Davies, "The leicester risk assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic uk setting," Diabetic medicine, vol. 27, no. 8, pp. 887–895, 2010.

[26] J. K.-O. Chung, H. Xue, E. W.-H. Pang, and D. C.-C. Tam, "Accuracy of fasting plasma glucose and hemoglobin a1c testing for the early detection of diabetes: A pilot study," Frontiers in Laboratory Medicine, vol. 1, no. 2, pp. 76–81, 2017.

[27] T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, G. Yang, and Y. Chen, "A machine learning-based framework to identify type 2 diabetes through electronic health records," International journal of medical informatics, vol. 97, pp. 120–127, 2017.

[28] N. P. Tigga and S. Garg, "Prediction of type 2 diabetes using machine learning classification methods," Procedia Computer Science, vol. 167, pp. 706–716, 2020.

[29] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using pima indian dataset," Journal of Diabetes & Metabolic Disorders, vol. 19, no. 1, pp. 391–403, 2020.

[30] Z. Xu and Z. Wang, "A risk prediction model for type 2 diabetes based on weighted feature selection of random forest and xgboost ensemble classifier," in 2019 eleventh international conference on advanced computational intelligence (ICACI). IEEE, 2019, pp. 278–283.

[31] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "Development of disease prediction model based on ensemble learning approach for diabetes and hypertension," IEEE Access, vol. 7, pp. 144 777–144 789, 2019.

[32] A. Rghioui, J. Lloret, S. Sendra, and A. Oumnad, "A smart architecture for diabetic patient monitoring using machine learning algorithms," in Healthcare, vol. 8, no. 3. Multidisciplinary Digital Publishing Institute, 2020, p. 348.

[33] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, "Metabolic syndrome and development of diabetes mellitus: Predictive modeling based on machine learning techniques," IEEE Access, vol. 7, pp. 1365–1375, 2018.

[34] A. Rghioui, J. Lloret, M. Harane, and A. Oumnad, "A smart glucose monitoring system for diabetic patient," Electronics, vol. 9, no. 4, p. 678, 2020.

[35] M. I. A. Efat, S. Rahman, and T. Rahman, "Iot based smart health monitoring system for diabetes patients using neural network," in International Conference on Cyber Security and Computer Science. Springer, 2020, pp. 593–606.

[36] M. Saravanan and R. Shubha, "Non-invasive analytics based smart system for diabetes monitoring," in International Conference on IoT Technologies for HealthCare. Springer, 2017, pp. 88–98.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3098691, IEEE Access

TABLE 9: Features Information

Features Description	ELSA Values	Dataset Values
BMI Categories	1.underweight less than 18.5 2.normal weight from 18.5 to 25 3.pre-obesity from 25 to 29 4.obesity class 1 from 30 to 35 5.obesity class 2 from 35 to 40 6.obesity class 3 greater than 40	1-6
Self-report of health	1.Excellent 2.Very good 3.Good 4.Fair 5.Poor	1-5,NaN
Gross monitor index	0-none, 5-all	0,5,NaN
Mobility index	0-none, 4-all	0,4,NaN
Fine motor index	0-none, 3-all	0,3, NaN
Large muscle index	0-none, 4-all	0,4,NaN
ADLs	0-none, 5-all	0,5,NaN
HLTHLM	0:No, 1:Yes	0,1,NaN
ADLs	0-none, 3-all	0,3,NaN
Education level	1.less than secondary 2.upper secondary and vocat 3.tertiary	1-3,NaN
Level of physical effort at current job	1.Sedentary occupation 2.Standing occupation 3.Physical work 4.Heavy manual work	1-2,NaN
Work stress-under pressure due to workload	1.strongly agree 2.agree 3.disagree 4.strongly disagree	1-4,NaN
Ever had hip fracture	0:No, 1:Yes	0,1,NaN
Mental health-the respondent's feelings much of the time over the week prior to the interview	0:Negative, 8:Positive	0,8,NaN
Marital Status	1.married 2.partnered 3.separated 4.divorced 5.widowed 6.never married	1-6,NaN
Currently working for pay	0:No, 1:Yes	0,1,NaN
Social activity-weekly contact with friends/relatives	0:No, 1:Yes	0,1,NaN
IADLs: using the phone, managing money, taking medications, shopping for groceries, preparing hot meals	0-none, 5-all	0,5,NaN
Ever had high cholesterol	0:No, 1:Yes	0,1,NaN
Ever had psychological problem	0:No, 1:Yes	0,1,NaN
Ever had heart problems	0:No, 1:Yes	0,1,NaN
Ever had asthma/cataracts/lung disease/dementia/memory problems	0:No, 1:Yes	0,1,NaN
Ever had hip fracture/stroke/arthritis /cancer/Parkinson	0:No, 1:Yes	0,1,NaN
Smoke ever or now	0:No, 1:Yes	0,1,NaN

[37] M. Hussain, M. Afzal, W. A. Khan, and S. Lee, "Clinical decision support service for elderly people in smart home environment," in 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV). IEEE, 2012, pp. 678–683.

[38] M. M. Bujnowska-Fedak and U. Grata-Borkowska, "Use of telemedicine-based care for the aging and elderly: promises and pitfalls," Smart Homecare Technology and TeleHealth, vol. 3, pp. 91–105, 2015.

[39] D. Seo, B. Yoo, and H. Ko, "Data-driven smart home system for elderly people based on web technologies," in International Conference on Distributed, Ambient, and Pervasive Interactions. Springer, 2016, pp. 122–131.

[40] O. Kocsis, K. Moustakas, N. Fakotakis, C. Vassiliou, A. Toska, G. C. Vanderheiden, A. Stergiou, D. Amaxilatis, A. Pardal, J. Quintas et al., "Smartwork: designing a smart age-friendly living and working environment for office workers," in Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2019, pp. 435–441.

[41] O. Kocsis, K. Moustakas, N. Fakotakis, H. J. Hermens, M. Cabrita, T. Ziemke, and R. Kovordanyi, "Conceptual architecture of a multi-dimensional modeling framework for older office workers," in Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2019, pp. 448–452.

[42] O. Geman, R. Toderean, M. M. Lungu, I. Chiuchisan, and M. Covasa, "Challenges in nutrition education using smart sensors and personalized tools for prevention and control of type 2 diabetes," in 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, 2017, pp. 444–447.

[43] S. M. Haffner, "Epidemiology of type 2 diabetes: risk factors," Diabetes care, vol. 21, no. Supplement 3, pp. C3–C6, 1998.

[44] L. J. Corbin, R. C. Richmond, K. H. Wade, S. Burgess, J. Bowden, G. D. Smith, and N. J. Timpson, "Bmi as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using mendelian randomization," Diabetes, vol. 65, no. 10, pp. 3002–3007, 2016.

[45] M. I. Harris, R. C. Eastman, C. C. Cowie, K. M. Flegal, and M. S. Eberhardt, "Racial and ethnic differences in glycemic control of adults with type 2 diabetes." Diabetes care, vol. 22, no. 3, pp. 403–408, 1999.

[46] M. Salinero-Fort, C. Burgos-Lunar, C. Lahoz, J. Mostaza, J. Abánades-Herranz, F. Laguna-Cuesta, E. Estirado-de Cabo, F. García-Iglesias, T. González-Alegre, B. Fernández-Puntero et al., "Performance of the finnish diabetes risk score and a simplified finnish diabetes risk score in a community-based, cross-sectional programme for screening of undiagnosed type 2 diabetes mellitus and dysglycaemia in madrid, spain: the spredia-2 study," PLoS One, vol. 11, no. 7, p. e0158489, 2016.

[47] M. Marmot, Z. Oldfield, S. Clemens, M. Blake, A. Phelps, J. Nazroo et al., "English longitudinal study of ageing: Waves 0–8, 1998–2017," 2018.

[48] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," IEEE Access, vol. 8, pp. 76 516–76 531, 2020.

[49] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," Egyptian Informatics Journal, vol. 19, no. 3, pp. 179–189, 2018.

[50] B. Kalaiselvi and M. Thangamani, "An efficient pearson correlation based improved random forest classification for protein structure prediction techniques," Measurement, vol. 162, p. 107885, 2020.

[51] K. Tanaka, T. Kurita, F. Meyer, L. Berthouze, and T. Kawabe, "Stepwise feature selection by cross validation for eeg-based brain computer interface," in The 2006 IEEE International Joint Conference on Neural Network Proceedings. IEEE, 2006, pp. 4672–4677.

[52] R. Muthukrishnan and R. Rohini, "Lasso: A feature selection technique in predictive modeling for machine learning," in 2016 IEEE international conference on advances in computer applications (ICACA). IEEE, 2016, pp. 18–20.

[53] J. Lindström and J. Tuomilehto, "The diabetes risk score: a practical tool to predict type 2 diabetes risk," Diabetes care, vol. 26, no. 3, pp. 725–731, 2003.

[54] W. Xu, J. Zhang, Q. Zhang, and X. Wei, "Risk prediction of type ii diabetes based on random forest model," in 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). IEEE, 2017, pp. 382–386.

[55] L. Muhammad, E. A. Algehyne, and S. S. Usman, "Predictive supervised machine learning models for diabetes mellitus," SN Computer Science, vol. 1, no. 5, pp. 1–10, 2020.

[56] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," Procedia computer science, vol. 132, pp. 1578–1585, 2018.

[57] J. D. Canary, L. Blizzard, R. P. Barry, D. W. Hosmer, and S. J. Quinn, "A comparison of the hosmer–lemeshow, pigeon–heyse, and tsiatis goodness-of-fit tests for binary logistic regression under two grouping methods," Communications in Statistics-Simulation and Computation, vol. 46, no. 3, pp. 1871–1894, 2017.

[58] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, "Metabolic syndrome and development of diabetes mellitus: Predictive modeling based on machine learning techniques," IEEE Access, vol. 7, pp. 1365–1375, 2019.

[59] K. P. S. Attwal and A. S. Dhiman, "Exploring data mining tool-weka and using weka to build and evaluate predictive models," Advances and Applications in Mathematical Sciences, vol. 19, no. 6, pp. 451–469, 2020.

[60] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in 9th Python in Science Conference, 2010.

[61] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," Expert Systems with Applications, vol. 138, p. 112821, 2019.

[62] N. Gogtay and U. Thatte, "Statistical evaluation of diagnostic tests (part 1): sensitivity, specificity, positive and negative predictive values," J Assoc Physicians India, vol. 65, no. 6, pp. 80–84, 2017.

[63] R. Trevethan, "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice," Frontiers in public health, vol. 5, p. 307, 2017.

[64] D. Goksuluk, S. Korkmaz, G. Zararsiz, and E. Karaagaoglu, "easyroc: An interactive web-tool for roc curve analysis using r language environment," The R Journal, vol. 8, pp. 213–230, 12 2016.

[65] K. Kirasich, T. Smith, and B. Sadler, "Random forest vs logistic regression: Binary classification for heterogeneous datasets," SMU Data Science Review, vol. 1, no. 3, p. 9, 2018.

[66] V. Vapnik, "Statistical learning theory wiley-interscience," New York, 1998.

[67] Z. He and W. Yu, "Stable feature selection for biomarker discovery," Computational biology and chemistry, vol. 34, no. 4, pp. 215–225, 2010.

[68] C. Thiyagarajan, K. A. Kumar, A. Bharathi, and I. T. Sathyamangalam, "Diabetes mellitus diagnosis based on transductive extreme learning machine," International Journal of Computer Science and Information Security (IJCSIS), vol. 15, no. 6, 2017.

[69] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," Knowledge and Information systems, vol. 42, no. 2, pp. 245–284, 2015.

[70] M. Moosazadeh, Z. Asemi, K. B. Lankarani, R. Tabrizi, N. Maharlouei, A. Naghibzadeh-Tahami, G. Yousefzadeh, R. Sadeghi, S. R. Khatibi, M. Afshari et al., "Family history of diabetes and the risk of gestational diabetes mellitus in iran: a systematic review and meta-analysis," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 11, pp. S99–S104, 2017.

[71] S. S. Casagrande, B. Linder, and C. C. Cowie, "Prevalence of gestational diabetes and subsequent type 2 diabetes among us women," Diabetes research and clinical practice, vol. 141, pp. 200–208, 2018.

[72] G. Battineni, G. G. Sagaro, C. Nalini, F. Amenta, and S. K. Tayebati, "Comparative machine-learning approach: A follow-up study on type 2 diabetes predictions by cross-validation methods," Machines, vol. 7, no. 4, p. 74, 2019.

[73] K. Dalakleidi, K. Zarkogianni, A. Thanopoulou, and K. Nikita, "Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications," Expert Systems, vol. 34, no. 6, p. e12214, 2017.

[74] M. Kowsher, F. S. Tithi, T. Rabeya, F. Afrin, and M. N. Huda, "Type 2 diabetics treatment and medication detection with machine learning classifier algorithm," in Proceedings of International Joint Conference on Computational Intelligence. Springer, 2020, pp. 519–531.

[75] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: A systematic review," IEEE Journal of Biomedical and Health Informatics, 2020.

[76] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014.

[77] L. Hasen and P. Salamon, "Neural networks ensembles," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993–1001, 1990.

[78] R. A. Hackett, C. Moore, A. Steptoe, and C. Lassale, "Health behaviour changes after type 2 diabetes diagnosis: Findings from the english longitudinal study of ageing," Scientific reports, vol. 8, no. 1, pp. 1–8, 2018.

[79] R. A. Hackett, Z. Dal, and A. Steptoe, "The relationship between sleep problems and cortisol in people with type 2 diabetes," Psychoneuroendocrinology, vol. 117, p. 104688, 2020.
[80] W. Engchuan, A. C. Dimopoulos, S. Tyrovolas, F. F. Caballero, A. Sanchez-Niubo, H. Arndt, J. L. Ayuso-Mateos, J. M. Haro, S. Chatterji, and D. B. Panagiotakos, "Sociodemographic indicators of health status using a machine learning approach and data from the english longitudinal study of aging (elsa)," Medical science monitor: international medical journal of experimental and clinical research, vol. 25, p. 1994, 2019.
[81] N. Fazakis, S. Alexiou, and O. Kocsis, "SmartWork Web Interface," https://profiles.smartworkproject.eu, 2021, [Online; accessed 10-May-2021].
[82] N. Fazakis, G. Kostopoulos, S. Kotsiantis, and I. Mporas, "Iterative robust semi-supervised missing data imputation," IEEE Access, vol. 8, pp. 90 555–90 569, 2020.
[83] C. Zhu, C. U. Idemudia, and W. Feng, "Improved logistic regression model for diabetes prediction by integrating pca and k-means techniques," Informatics in Medicine Unlocked, vol. 17, p. 100179, 2019.

NIKOS FAZAKIS received the Diploma degree from the Department of Electrical and Computer Engineering, and also the M.B.A. degree from the University of Patras in Greece. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Patras. He has participated in numerous European and National research programs. He has a variety of publications in the fields of machine learning and data mining, and through the years he has earned several best paper awards for numerous works in international conferences.

OTILIA KOCSIS is a Senior Researcher holding a BSc degree in Physics with a major in Biophysics from the University "Al.I. Cuza" Iasi (Romania), and MSc and PhD degrees in Medical Physics from the University of Patras (Greece). She has extensive experience from both the industry (Knowledge SA, SingularLogic AE, BOKtech SRL) and the academic/research environments (Technical Educational Institute of Patras, University of Patras), being actively involved in the coordination and implementation of a large number of national and European projects (e.g. GEMINI, AMIGO, POLIAS, MoveOn, PlayMancer, AmiBio, DRYMOS, TELECARE, Cloud4All, Prosperity4All, myAirCoach, OActive, SmartWork, GATEKEEPER). She is currently working as an Associate Senior Researcher at the Visualization and Virtual Reality Group, University of Patras. She is author and co-author of more than 50 publications in international journals, conferences and edited books, and she has acted as a reviewer for several conferences and journals. She is member of the Hellenic Artificial Intelligence Society (EETN).

ELIAS DRITSAS received his Diploma, M.Sc and Ph.D degrees in Computer Science and Informatics from the Department of Computer Engineering and Informatics, University of Patras. Also, he received his MBA from University of Derby (U.K.). He is author and co-author of publications in the area of Machine Learning and Data Analysis. As Ph.D. candidate, he served as research scholar under the project funded by the Hellenic Foundation for Research and Innovation (HFRI) and General Secretariat for Research and Technology (GSRT). Also, he has worked as Software Engineer in Research Committee of University of Patras. He is currently working as Data Analyst at the Visualization and Virtual Reality Group, University of Patras.

SOTIRIS ALEXIOU received his Diploma from Computer Engineering and Informatics Department in 2019. He is a postgraduate student in the Interdisciplinary Master Program entitled "Information Processing Systems And Engineering Information", University of Patras. He is currently working as a Research Associate and Data Analyst in the Visualization and Virtual Reality Group (VVR), at the Department of Electrical and Computer Engineering, University of Patras.

NIKOS FAKOTAKIS received the BSc degree in electronics from the University of London (United Kingdom), in 1978, the MSc degree in electronics from the University of Wales (United Kingdom), and the PhD degree in speech processing from the University of Patras, (Greece), in 1986. From 1986 to 1992 he was lecturer in the Electrical and Computer Engineering Department of the University of Patras, from 1992 to 1999 assistant professor, from 2000 to 2003 he has been associate professor and since 2003 he is professor in the area of Speech and Natural Language Processing. He is currently the director of the Communication and Information Technology Division of the Electrical and Computer Engineering Department (since 2005), director of the Wire Communications Laboratory (WCL) (since 2004), and Head of the Artificial Intelligence Group. The results of the scientific work conducted by he or under his supervision has resulted in more than 400 scientific publications in internationally recognized journals and conferences, which have been cited more than 3,000 times.

KONSTANTINOS MOUSTAKAS received the Diploma and Ph.D. degrees in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003 and 2007, respectively. From 2007 to 2011, he was a Post-Doctoral Research Fellow with the Information Technologies Institute, Centre for Research and Technology Hellas, Hellas, Greece. He is currently an Associate Professor with the Electrical and Computer Engineering Department, University of Patras, Patras, Greece, where he is also the Head of the Visualization and Virtual Reality Group. He has authored or coauthored over 150 papers in refereed journals, edited books, and international conferences. His main research interests include virtual, augmented, and mixed reality, 3-D geometry processing, haptics, virtual physiological human modeling, information visualization, physics-based simulations, computational geometry, computer vision, and stereoscopic image processing. He is a member of the Organizing Committee of three international conferences and a member of the Technical Program Committee for more than 15 international conferences. He is a member of the IEEE Computer Society. He serves as a regular reviewer for several technical journals. He participated in more than 20 research and development projects funded by the EC and the Greek Secretariat of Research and Technology. He serves as the coordinator or the scientific coordinator in four of them.