Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

# *A Random Forest a Day Keeps the Doctor Away*

Markus Eyting

December 07, 2020

Discussion paper number 2026

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Markus Eyting
Chair of Digital Economics
University of Mainz
Jakob-Welder-Weg 4
55128 Mainz
Goethe University Frankfurt
60323 Frankfurt am Main
Germany

meyting@uni-mainz.de

All discussion papers can be downloaded from http://wiwi.uni-mainz.de/DP

# A Random Forest a Day Keeps the Doctor Away

Markus Eyting*

Graduate School of Economics, Finance and Management

This version: December 7, 2020

**Abstract**

Using a unique dataset from a German health check-up provider including detailed individual questionnaire data as well as medical test data, I apply a random forest to predict several health risk factors. I evaluate the prediction performance using various metrics and find decent prediction qualities across all outcomes. By identifying the most relevant predictor variables, I compile concise and validated questionnaire tools to identify individuals' blood pressure, blood glucose, and cholesterol levels, their risk of a coronary heart disease, whether or not they suffer from plaque or a metabolic syndrome as well as their relative fitness levels. In a second step, I compare the prediction results to physician predictions of the same patient observations. I find that the random forest outperforms the physicians if predictions are based on the same information set. When additionally providing the physicians with the random forest predictions for a particular patient observation, the physicians align with the random forest predictions. Finally, while the random forest considers various psychological scales, the physicians focus on family health history information instead.

* *Address for correspondance:* Markus Eyting, JGU Mainz, Jakob-Welder-Weg 4, 55128 Mainz, Germany. E-mail: meyting@uni-mainz.de. Phone: +49-6131-39-22166.

# 1 Introduction

The recent progress in computer processing power and the enormous increase in available data allows algorithms to incorporate available information that an individual would consider too costly to screen or process. In addition, recent advances in machine learning methods render computerized methods less prone to distortions in the perceptions of certain information. Consequently, in recent years computerized methods gained popularity in many fields, one of which is the field of "personalized medicine" (or "precision medicine"). Personalized medicine refers to a medical model, that proposes healthcare designed for individual patient attributes (Pokorska-Bocci *et al.*, 2014). This implies, that the more patient attributes are considered, the more personalized predictions can be made. In light of the above-mentioned recent technological progress, it comes at no surprise that computerized risk estimators providing personal health information have become widely available.

Today, a large number of websites provide possible disease diagnoses for individuals at home and assist them to decide whether and how to treat a particular condition (Ryan and Wilson, 2008). However, a large amount of available information online has been published without a careful evaluation of its validity (Diviani *et al.*, 2015) and hence the quality and reliability of these diagnoses is often quite low. Especially for health information, the uncertainty in the reliability of online information can have harmful consequences, e.g. when the information is used to inform potentially hazardous treatment decisions (Cline and Haynes, 2001) or when the diagnoses lead to more harmful subsequent health behavior (Cline and Haynes, 2001; Mullainathan and Obermeyer, 2017). Contrary, reliable health predictors can benefit patients, physicians and the society as a whole. On the patients' side, reliable predictors could inform about certain conditions early on. Patients can gain first information about whether or not to consult a physician or start potential treatments and thereby limit potential pain later on in the treatment process. On the physicians' side, reliable health predictors can consolidate information and give first insights about a new patient's conditions. Finally, from a social welfare point of view, the early identification of potential health risk factors can drastically save treatment costs and capacities and, depending on the disease, contain the possible danger of a spreading disease. Thus, in order to benefit from the recent advances in personalized medicine, health predictors need to be both, validated and simple to use for individuals.

In this paper, I apply a flexible machine learning algorithm (random forest) to predict several individual health outcomes based on unique data from a German health check-up provider. I make use of a rich dataset that entails both questionnaire data as well as medical test data and use the former to predict the results of the latter. Ultimately, the goal is to not only accurately predict health outcomes but also to find out which patient information need to be collected in order to do so. Put differently, I aim to provide a concise and validated health prediction tool that can widely be applied. For applicability reasons, I drop data that is context specific or highly sensitive before building the random forest. Hence, I only focus on data that can easily be gathered through surveys. I evaluate the random forest predictions using various metrics from the computer science literature in order to provide a comprehensive picture of its applicability. Moreover, in order to identify the relevant questions for the prediction tool, I go beyond pure predictions and identify the most important attributes for the prediction of each health outcome.

Predicting a mix of different health risk factors and diseases (hypertension, diabetes, cholesterol levels, risk for a coronary heart disease, plaques, metabolic syndrome and fitness), I find that appropriate questionnaire data can provide meaningful insights into an individual's health condition. With ROC-AUC scores[1] ranging from 0.69 (for blood pressure) to 0.94 (for risk of coronary heart diseases) the random forest provides decent prediction results for all health outcomes. A more in-depth analysis of the information retrieval capacities of the random forest predictions confirms these results. With respect to the variables that the random forest considers important for predictions, I find that weight, height and the amount of alcohol an individual consumes provide meaningful insights for all outcomes. Other important variables include an individual's smoking and sports behavior as well as psychological measures of a person's stress coping and prevention capabilities.

In a final step, I compare the random forest predictions to physician predictions on the same patient in order to gain further insights into the applicability of the constructed prediction tool.[2] I vary the information set of physicians by partly providing them with the random forest predictions as additional information. As for the random forest predictions, I

[1] The ROC-AUC (Receiver Operating Characteristics - Area Under Curve) score shows the probability, that a randomly drawn positive case has a higher likelihood to be predicted positive than a randomly drawn negative case.

[2] Initially, the idea was to only identify a concise set of relevant questions with which one can reliably predict an individual's health status. Subsequently, I planned to use the created prediction tool in a field experiment in which an individual's health status served as a control measure. However, when the field experiment got cancelled by a third party, I decided to expand the scope of this study by further validating my prediction results through a comparison with physician predictions.

also collect data on the ten most important variables for the predictions of the physicians.[3]

Using the same metrics as before, I find that the random forest predictions are at least as good as the physician predictions in six of the seven outcomes (with diabetes being the only outcome where results are mixed). The physician predictions greatly improve up to a similar performance as the random forest predictions when physicians are provided with the random forest predictions as additional information. Looking at the attributes that were considered for making the predictions, I first find a few common attributes that were used for both types of predictions. However, I also find first evidence that random forest predictions place higher importance on psychological variables as well as working hours, whereas physicians consider gender and family health history variables much more important.

This paper contributes to five main literatures. First, this paper contributes to the existing literature on computerized methods in the field of "precision medicine" (Ahmed *et al.*, 2020; Brighton, 2011; Collins and Varmus, 2015; Rajkomar *et al.*, 2019; Shickel *et al.*, 2018). Most studies in this field aim to accurately predict hazardous health conditions like cancer[4] or cardiovascular diseases[5]. Other studies focus on the measurement of in vitro drug responses and genome-wide gene expressions to drugs (Lee *et al.*, 2018). I join the current literature on using machine learning methods to predict health conditions but in doing so I take a different angle and focus on the prediction of health risk factors from self-reported questionnaire data.

Second, there is an ongoing debate about the validity of health self-reports (Goldman *et al.*, 2003; Jürges, 2007; Kriegsman *et al.*, 1996; Margolis *et al.*, 2008; Midthjell *et al.*, 1992; Tretli *et al.*, 1982). Existing epidemiological and experimental studies rely on insufficiently validated self-reports (Hu *et al.*, 2018; Manson *et al.*, 1991) or proxies for health conditions such as physician visits (Bartling *et al.*, 2012) as data of medical tests are often not available. In order to derive meaningful implications from these studies, it is important that these health self-reports are credible and accurate. By providing a validated questionnaire tool that is based on self-reports as well as medical data, I bridge the gap between invalidated self-reports and medical test results. Validated computerized methods can decrease the risk of false diagnoses and thereby lower the social cost of delayed or even false treatments.

[3]For the random forest, importance is measured by the Gini impurity whereas for the physicians I rely on self-reports of the top ten most important variables.

[4]Kourou *et al.* (2015) review various ML approaches employed in the modelling of cancer progression.

[5]Seetharam *et al.* (2019) provide several examples of recent machine learning utilization in echocardiography, nuclear cardiology, computed tomography, and magnetic resonance imaging.

Third, this paper adds to the current literature on predictive machine learning methods and regularization tools in economics (Athey, 2018). This line of research has its foundation in the computer science literature, but the recent interest in machine learning methods among economists has created room for several interdisciplinary projects using these methods. Not long ago, economists have provided and extended the theoretical background of random forest estimations (Chernozhukov *et al.*, 2018; Wager, 2014; Wager and Athey, 2018), which renders the random forest one of the most prominent machine learning methods in economics. By applying a method that was invented in the computer science literature, but that got recently adopted in the economic literature, to a medical context, I combine several strands of literature in an interdisciplinary project.

Fourth, by comparing the predictions of the random forest to physician predictions, this paper contributes to the literature on "man vs. machine" (Acemoglu and Restrepo, 2018; Wong *et al.*, 2018). Thereby, I do not only carefully assess the prediction performance difference between individuals and an algorithm, but by providing physicians with random forest predictions, I also touch upon the very recent literature on individuals' level of trust towards machine learning algorithms (Spiegelhalter, 2020; Yeomans *et al.*, 2019).

Finally, I add to the discussion on individual information processing imperfections. By comparing the random forest predictions to physician predictions, I uncover physicians' difficulties to process a large amount of readily available data. Handel and Schwartzstein (2018) review possible reasons for why individuals do not use available information and broadly categorize them into two camps, "frictions" and "mental gaps". Frictions are responsible, when individuals form accurate beliefs from the information that they consider worth processing but thereby consider certain information simply too costly to process. This argument is closely related to the concept of "rational inattention" and described in several studies, such as Stigler (1961), McCall (1970), Caplin and Dean (2015), Gabaix (2014), Woodford (2012), and Sims (2003). In my setting, this might imply that physicians rationally disregard some of the patient information as they consider the added value of the information not worth the costs of processing it. Contrary, mental gaps describe a situation in which individuals suffer from psychological distortions or misperceptions of the available information and thereby create a gap between what they believe and what they should rationally believe given the information costs. This would imply that physicians falsely interpret (the importance of) some of the patient information and hence misjudge its effect on the patient's health risk. Models that describe the working and implications of mental gaps more closely include Bordalo *et al.* (2013); Enke and Zimmermann (2019); Koszegi and Szeidl (2013); Schwartzstein (2014). By providing physicians with a large amount of data and evaluating their predictions of patients' health risk, I provide an example of individual information gathering and processing that suffers from frictions/mental gaps and show how machine learning methods can provide helpful support for both camps.

Section 2 gives a detailed account of the data on which I train and evaluate the random forest before Section 3 describes the random forest specification and evaluation metrics that were used for the predictions. Section 4 shows the results of the random forest predictions for all health outcomes. In Section 5, I describe the survey among the physicians and discuss its results in Section 6. Section 7 discusses the findings and caveats of this study. Finally, Section 8 concludes.

# 2 Data

I use a unique dataset from a German health check-up provider that conducts health check-ups in several of the largest cities across Germany.[6] The initial dataset consisted of $11,426$ patient observations. One patient observation represents all data that is gathered during a patient's health check-up between 2008 and 2017. During a health check-up, a patient is first asked to fill in detailed questionnaires about their demographics, personal and family health history and behavior, as well as an extensive survey about psychological characteristics. Subsequently, several medical tests are performed. Hence, my dataset combines both, extensive self-reported survey data as well as objective medical test data. In 2014, the original questionnaires as well as the scope of the medical tests were revised and extended from previously 103 to then 416 different questionnaire and test items.

I conduct the following pre-processing steps in order to prepare the data for the main analysis. First, I only consider patient observations from the years 2014 to 2017. Second, I summarize detailed (psychological) items into broader variables. Third, I drop variables with a majority of missing values as well as highly sensitive data like personal health history data or information about current medications in order to exclude variables, where it seems difficult to collect self-reports from patients. Fourth, I compile a correlation matrix of all remaining variables and randomly drop one of two highly correlated (correlation $\geq 0.9$)

[6]My dataset comprises patient observations from eight of the ten largest cities in Germany, including Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart, Dusseldorf, and Dortmund. Additionally, it includes patient observations from Hanover, Mainz and Münster.

variables. Finally, I impute remaining missing values using Strawman imputation (Tang and Ishwaran, 2017).[7] The final dataset consists of 6344 patient observations and includes 72 explanatory variables as well as 7 outcome variables.

## 2.1 Explanatory Variables

The explanatory variables are exclusively based on self-reported questionnaire data and can hence be collected through standard surveys, so that the resulting prediction tool can easily be applied. I broadly classify them into the four categories general, health behavior, psychological, and family health history.

### 2.1.1 General Data

The general data includes demographic variables such as age, gender, the number of children the patient has and whether or not the patient lives alone or with a partner. Moreover, it includes a few work related variables, such as whether the patient had personnel responsibility at work, whether they are leading a team or how many hours the individual works per week. Finally, the number of times that the individual has taken part in the health check-up is recorded. See Table 1 for a brief summary of the general variables.

### 2.1.2 Health Behavior Data

The health behavior data provide a detailed view on each patient's health behavior. Next to standard measures such as weight and height, it includes data on a (i) patient's diet, (ii) their drinking and smoking habits, (iii) their level and kind of sporting activity as well as (iv) their sleeping behavior. Data on a patient's diet was collected by asking how often the patient eats how much of various kinds of food (carbs, fruit, salad, vegetables) per day. Similarly, for the patient's drinking habits, they are asked how often and how much they drink different kinds of alcoholic drinks (beer, red wine, white wine, sparkling wine, liquor) per week. Based on a conversion table supplied by the health check-up provider I calculated the amount of pure alcohol a patient drinks per week. With regard to a patient's smoking habits they are asked if they currently smoke and if so how many cigarettes per day they

[7] Tang and Ishwaran (2017) define Strawman imputation as imputing missing valued of continuous variables with the median of all non-missing values and missing values of categorical variables with the most frequently used value of all non-missing values.

Table 1: General Variables

| Variable | Measurement | Mean | SD |
|---|---|---|---|
| Age | years | 49.93 | 6.75 |
| Gender | 0(male), 1(female) | 0.42 | 0.49 |
| Partner | 0(no), 1(yes) | 0.82 | 0.38 |
| Personnel responsibility | 0(no), 1(yes) | 0.30 | 0.46 |
| Leadership | 0(no), 1(yes) | 0.33 | 0.47 |
| Children | count | 1.03 | 0.89 |
| Working hours | hours | 41.74 | 10.18 |
| Times visit | count | 1.53 | 0.72 |

*Notes: Age*, *Gender*: the patient's age, gender; *Partner*: whether or not the patient lives with a partner; *Personnel responsibility*: whether or not the patient has personnel responsibility at work; *Leadership*: whether or not the patient occupies a leading position at work; *Children*: the number of children the patient has; *Working hours*: the actual number of hours the patient works on average each week; *Times visit*: the number of times the patient has taken part in the health check-up.

usually smoke. Furthermore, they are asked if they have ever smoked in their life and for how long they have (not) been smoking to date. A patient's sporting activity is elicited by asking if and how often in a week, they engage in different kinds of cardio sports or different kinds of strength sports. I summarized all cardio sports into "minutes of cardio per week" and all strength sports into "minutes of strength sport per week". Moreover, it was asked whether all large muscle groups were trained during strength workouts. Finally, a patient's sleeping behavior is measured using several subquestions. She is asked at what time they usually go to bed, at what time they usually wake up and how many hours of sleep they usually get per night. Moreover, they are asked how often they wake up per night and how long it takes from the moment that they go to bed until they fall asleep. She is also asked to rate the quality of their sleep, how often they took any kind of sleeping pills, how often they had problems to stay awake during daytime activities and whether or not they felt like having enough energy to manage daily routines in the preceding four weeks. Table 2 summarizes all health behavior variables.

### 2.1.3 Psychological Data

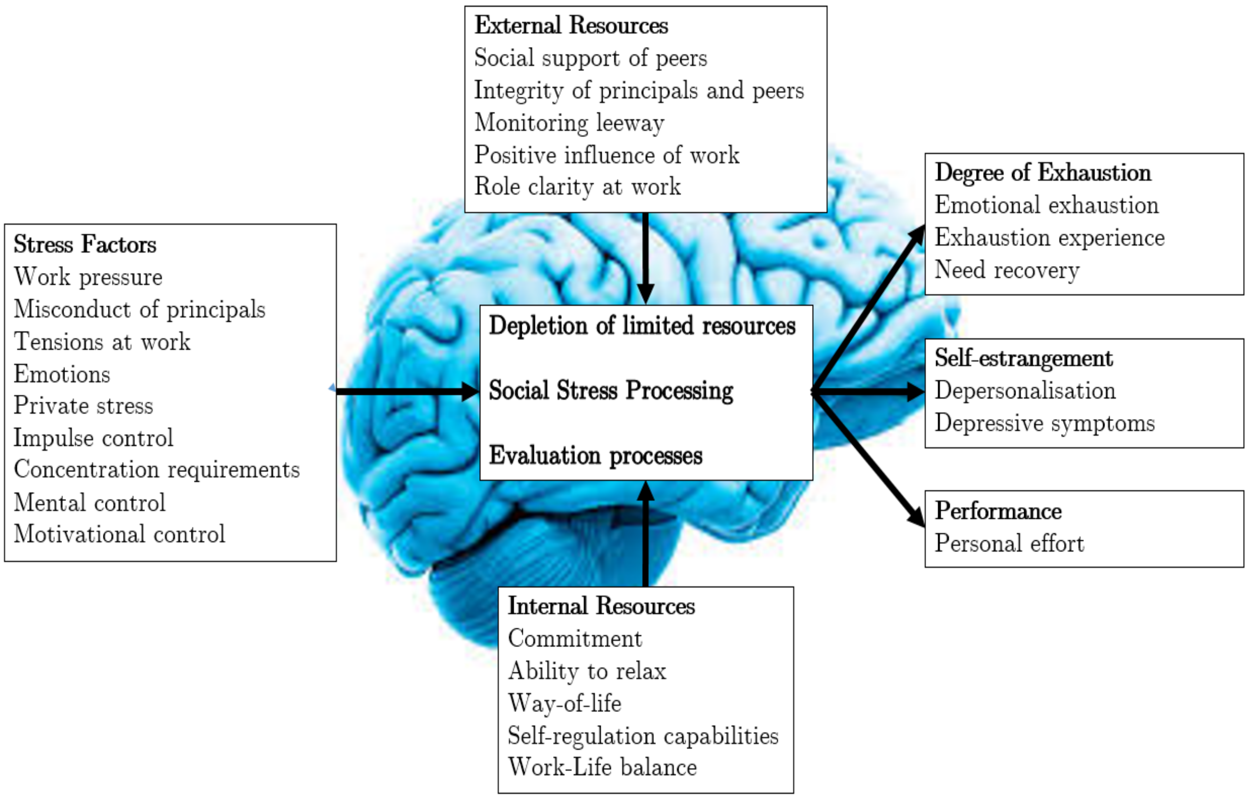
The psychological data measures a variety of individual characteristics that allow a comprehensive picture on the patient's mental condition in various settings. It includes 25 measures about (i) factors of personal, social or work related stress, (ii) internal and external psycho-

Table 2: Health Behavior Variables

| Variable | Measurement | Mean | SD |
|---|---|---|---|
| Weight | kg | 80.60 | 16.35 |
| Height | cm | 175.97 | 9.58 |
| Carbs | count | 3.01 | 1.49 |
| Fruits | count | 1.07 | 0.91 |
| Salad | count | 0.49 | 0.84 |
| Vegetables | count | 0.33 | 0.67 |
| Alcohol | liters | 0.11 | 0.12 |
| Smoke | 0(no), 1(yes) | 0.13 | 0.34 |
| Smoke amount | count | 14.79 | 6.06 |
| Ex-smoker | 0(no), 1(yes) | 0.27 | 0.45 |
| Never smoked | 0(no), 1(yes) | 0.59 | 0.49 |
| Not smoke duration | years | 3.98 | 8.31 |
| Smoke duration | years | 7.58 | 11.37 |
| Cardio | minutes | 161.66 | 154.31 |
| Strength | minutes | 74.44 | 52.42 |
| Strength detail | 0(no), 1(yes) | 0.36 | 0.48 |
| Time to bed | time | 22.29 | 8.28 |
| Time wake up | time | 6.24 | 0.69 |
| Time asleep | hours | 6.60 | 0.87 |
| Times woken up | count | 1.57 | 1.57 |
| Time to fall asleep | Minutes | 14.13 | 14.34 |
| Sleep quality | [1(bad), ..., 4(good)] | 2.90 | 0.65 |
| Sleeping pills | [1(rarely), ..., 4(often)] | 1.07 | 0.39 |
| Sleep daytime | [1(rarely), ..., 4(often)] | 1.42 | 0.72 |
| Energy | [1(no problems), ..., 4(huge problems)] | 2.09 | 0.89 |

*Notes: Weight, Height*: the patient's weight, height; *Carbs, Fruits, Salad, Vegetables*: the respective number of portions of carbs, fruits, salad, or vegetables the patient eats per day; *Alcohol*: the amount of pure alcohol a patient drinks per week; *Smoke, Smoke amount, Ex-smoker, Never smoked, Not smoke duration, Smoke duration*: whether or not the patient currently smokes, how many cigarettes per day, whether or not the patient has smoked in the past conditional on them currently being a non-smoker, whether or not the patient has ever smoked; for how long the patient has been not smoking to date; for how long the patient has been smoking to date; *Cardio, Strength*: average duration the patient engages in cardio and strength sport per week, respectively; *Strength detail*: whether or not all large muscle groups are trained during strength workouts;*Time to bed, Time wake up, Time asleep, Times woken up, Time to fall asleep*: the time the patient goes to bed, wakes up, the duration the patient is asleep, how often they wake up per night, how long they need to fall asleep from the moment they go to bed; *Sleep quality, Sleeping pills, Sleep daytime, Energy*: self-reported sleep quality, how often the patient takes sleeping pills, how often they had problems to stay awake during daytime activities, whether or not they felt like having enough energy to manage daily routines in the preceding four weeks.

Figure 1: Integrative Resource Model of Stress Coping and Prevention




*Notes:* This model is based on the model of the health check-up provider and updated to fit the dataset of this study.

logical resources and capabilities to deal with stress and (iii) implications of stress factors on private and professional life. All variables are measured by psychological scales and are based on multiple items each. Figure 1 depicts the "integrative resource model of stress coping and prevention" that serves as the basis for the psychological assessment.[8] The stress factors include work related factors, such as pressure at work, misconduct of principals and other tensions at work as well as private, internal factors such as individual emotions, perceptions of private stress, required impulse control as well as mental and motivational control, or concentration requirements in daily activities. These factors negatively influence an individual' state of mind, e.g by depleting limited mental resources or requiring a constant evaluation of

[8]This model is based on a model of the health check-up provider. Rohmert *et al.* (1975) set up a more general model for the evaluation of stress demands at the workplace. Their ideas can be considered the backbone of my model.

Table 3: Psychological Variables

| Variable | Measurement | Mean | SD |
|---|---|---|---|
| Work pressure | -3(good), ..., 3(bad) | -0.38 | 0.98 |
| Misconduct | -3(good), ..., 3(bad) | 0.02 | 0.96 |
| Emotions | -3(good), ..., 3(bad) | -0.09 | 0.99 |
| Tension workplace | -3(good), ..., 3(bad) | 0.12 | 1.00 |
| Private stress | -3(good), ..., 3(bad) | 0.14 | 1.05 |
| Impulse control | -3(good), ..., 3(bad) | -0.28 | 0.89 |
| Concentration requirements | -3(good), ..., 3(bad) | -0.16 | 0.96 |
| Mental control | -3(good), ..., 3(bad) | 0.04 | 0.99 |
| Motivational control | -3(good), ..., 3(bad) | 0.39 | 0.64 |
| Social support | -3(good), ..., 3(bad) | -0.15 | 1.00 |
| Integrity | -3(good), ..., 3(bad) | 0.14 | 0.98 |
| Monitoring leeway | -3(good), ..., 3(bad) | -0.19 | 0.94 |
| Positive influence | -3(good), ..., 3(bad) | -0.06 | 1.03 |
| Role clarity | -3(good), ..., 3(bad) | 0.06 | 1.05 |
| Commitment | -3(good), ..., 3(bad) | 0.00 | 1.08 |
| Ability to relax | -3(good), ..., 3(bad) | 0.09 | 1.00 |
| Way of life | -3(good), ..., 3(bad) | 0.16 | 0.96 |
| Self-regulation | -3(good), ..., 3(bad) | -0.07 | 0.96 |
| Work-life-balance | -3(good), ..., 3(bad) | -1.10 | 0.97 |
| Emotional exhaustion | -3(good), ..., 3(bad) | -0.06 | 0.99 |
| Exhaustion experience | -3(good), ..., 3(bad) | 0.04 | 1.01 |
| Need recovery | -3(good), ..., 3(bad) | -0.03 | 1.02 |
| Depersonalisation | -3(good), ..., 3(bad) | 0.01 | 0.92 |
| Depressive symptoms | -3(good), ..., 3(bad) | -0.10 | 0.96 |
| Personal effort | -3(good), ..., 3(bad) | -0.01 | 1.00 |

*Notes:* All psychological variables are based on multiple items each. Please see Appendix A.1 for the full set of items.

the current state of mind. However, certain external and internal resources and capabilities provide support in dealing with the mentioned stress factors. The dataset includes work and social context related resources such as social support of peers, integrity of principals and peers or a potential positive influence of the work on a patient's personal mental condition. Internal resources include, e.g. the abilities to relax, self-regulation capabilities as well as work-life balancing skills. Implications of the interplay of stress factors with external and internal resources include a patient's emotional exhaustion and need for recovery, feelings of depersonalization or depression and impacts on personal performances at work and in daily life. Table 3 shows a summary of all psychological variables.

### 2.1.4 Family Health History Data

Finally, my dataset includes information on a patient's family health history. Here, patients were asked whether or not any of various diseases has occurred among any of their siblings, parents or grandparents. In particular, they were asked about different forms of cancer, such as breast cancer, colon cancer, skin cancer, lung cancer, prostate cancer or any other form of cancer, as well as hypertension, dementia, colorectal polyps, narrowing of coronary vessels, any mental disease, diabetes, heart attack or stroke.

Table 4: Family Health History Variables

| Variable | Measurement | Mean | SD |
|---|---|---|---|
| Breast cancer (family) | 0(no), 1(yes) | 0.14 | 0.34 |
| Colon cancer (family) | 0(no), 1(yes) | 0.14 | 0.35 |
| Skin cancer (family) | 0(no), 1(yes) | 0.02 | 0.14 |
| Lung cancer (family) | 0(no), 1(yes) | 0.10 | 0.30 |
| Prostate cancer (family) | 0(no), 1(yes) | 0.10 | 0.30 |
| Other cancer (family) | 0(no), 1(yes) | 0.28 | 0.45 |
| Hypertension (family) | 0(no), 1(yes) | 0.58 | 0.49 |
| Dementia (family) | 0(no), 1(yes) | 0.16 | 0.37 |
| Colorectal polyp (family) | 0(no), 1(yes) | 0.04 | 0.20 |
| Coronary vessels (family) | 0(no), 1(yes) | 0.14 | 0.35 |
| Mental disease (family) | 0(no), 1(yes) | 0.08 | 0.27 |
| Diabetes (family) | 0(no), 1(yes) | 0.32 | 0.47 |
| Heart attack (family) | 0(no), 1(yes) | 0.28 | 0.45 |
| Stroke (family) | 0(no), 1(yes) | 0.25 | 0.43 |

*Notes:* All variables are dummy variables taking the value 1 if any of a patients' siblings, parents or grandparents has suffered from the respective condition and 0 otherwise.

## 2.2 Outcome Variables

Contrary to the explanatory variables based on survey questions, all outcome variables are based on objective medical test data. I define a number of outcome variables in order to predict various different health conditions. In particular, after extensive discussions with the check-up provider, I define the following seven outcome variables: blood pressure, blood glucose, cholesterol level, risk of a coronary heart disease, plaques, metabolic syndrome, and fitness. In collaboration with the check-up provider, I identify the respective classes for each of the outcome variables according to their applicability for medical practices. See Table 5 for an overview over all outcome variables.

### 2.2.1 Blood Pressure

Blood pressure is determined by the amount of blood a heart pumps and the amount of resistance to blood flow in an individual's arteries. The more blood a heart pumps and the narrower an individual's arteries are, the higher the individual's blood pressure. If an individual suffers from uncontrolled high blood pressure they are exposed to an increased risk of serious health problems, including heart attack and stroke (MFMER, 2020d). I divide the variable blood pressure into four commonly used classes, "optimal", "normal", "high-normal" and "high".

### 2.2.2 Diabetes

Diabetes describes a condition of excess blood sugar levels which can lead to serious health problems (MFMER, 2020c). I use several markers to assess an individual's risk for diabetes. Whereas the fasting blood sugar level serves as an indication for short-term blood sugar level variations, the HBA1c describes the average blood sugar levels for the last two to three months (WHO, 2011). Additionally, the homeostatic model assessment (HOMA) (Wallace *et al.*, 2004) assesses insulin resistance and hence serves as a complementing marker for risk of diabetes. I divide the variable diabetes into three commonly used classes, "no disorder", "at risk" and "diabetes".

### 2.2.3 Cholesterol

High levels of cholesterol pose another risk factor for a number of diseases. While there are different types of cholesterol (Low-density lipoprotein (LDL) and High-density lipoprotein

Table 5: Definition of Health Outcomes

| Outcome | Classes | Definition |
|---|---|---|
| **Blood pressure** | Optimal | $< 120/80$ mmHg |
| | Normal | [120/80 mmHg, 130/85 mmHg) |
| | High-Normal | [130/85 mmHg, 140/90) |
| | High | $\geq 120/80$ mmHg |
| **Diabetes** | No disorder | *If all of the following apply:* |
| | | Fasting blood sugar level: $\leq 99$ mg/dl |
| | | HBA1c: $\leq 5.8$ |
| | | HOMA-Index: $\leq 2.5$ |
| | At Risk | *If at least one of the following applies:* |
| | | Fasting blood sugar level: [100 mg/dl, 126 mg/dl] |
| | | HBA1c: [5.9, 6.0] |
| | | HOMA-Index: $> 2.5$ |
| | Diabetes | *If at least one of the following applies:* |
| | | Fasting blood sugar level: $> 126$ mg/dl] |
| | | HBA1c: $> 6.0$] |
| **Cholesterol** | Good | HDL: female $> 70$ mg/dl |
| | | male $> 60$ |
| | Normal | HDL: female (50 mg/dl, 70 mg/dl) |
| | | male (40 mg/dl, 60 mg/dl) |
| | Bad | HDL: female $< 50$ mg/dl |
| | | male $< 40$ mg/dl |
| **Coronary heart disease** | Low | Lifetime risk score: $< 30$ % |
| | High | Lifetime risk score: $\geq 30$ % |
| **Plaques** | No | Fatty deposit: $\leq 1.5$ mm |
| | Yes | Fatty deposit: $> 1.5$ mm |
| **Metabolic syndrome** | No | *If at least three of the following apply:* |
| | | Blood pressure: $> 130/85$ mmHg |
| | | Fasting blood sugar level: $> 100$ mg/dl |
| | | Tryglyceride: $> 150$ mg/dl |
| | | Waist circumference: female: $> 80$cm |
| | | male: $> 94$cm |
| | | HDL: female: $< 50$ mg/dl |
| | | male: $< 40$mg/dl |
| | Yes | *Otherwise* |
| **Fitness** | Good | Maximum oxygen uptake: $\geq 40^{th}$ Percentile |
| | Bad | Maximum oxygen uptake: $< 40^{th}$ Percentile |

*Notes:* The left column shows the various health outcomes that were predicted. The middle column displays the different classed that were defined for each health outcome. The right column shows the definitions of the respective classes.

(HDL)), only the former is potentially harmful. The LDL builds up in the walls of an individual's arteries, making them harder and narrower, and hence impedes blood to circulate through the respective arteries. The HDL usually picks up excess cholesterol and takes it back to the liver (MFMER, 2020b). I divide the variable cholesterol into three commonly used classes, "good", "normal" and "bad", with different thresholds for men and women.

### 2.2.4 Coronary Heart Disease

Using an adapted version of the "Reynold's Risk Score" from the health check-up provider ("Lifetime Risk Calculator"), I use the calculated risk that an individual suffers from a coronary heart disease (CHD) until the age of 80 (Berry *et al.*, 2012). This calculator comprises information about an individual's blood pressure, cholesterol levels (LDL and HDL), smoking behavior[9], risk of diabetes, and previous medication. I divide the score into two classes, "low" and "high" risk, using a threshold risk of 30%.

### 2.2.5 Plaques

Plaques can be described as fatty deposits that clog an individual's carotid artery. If the carotid artery is prevented from delivering blood to the individual's brain or head, the individual is at risk of a stroke (MFMER, 2020a). I define the binary variable "Plaques" such that it takes the value "Yes" if the fatty deposit in the carotid artery is greater than 1.5mm and "No" if the deposit is smaller than 1.5mm.

### 2.2.6 Metabolic Syndrome

Metabolic syndrome is a cluster of conditions that together form an increased risk for severe diseases. These conditions include high blood pressure, high blood sugar, high triglyceride levels, excess body fat around the waist, and abnormal cholesterol. I define the binary variable "Metabolic Syndrome" such that it takes the value "Yes" if at least three out of these five conditions are met. Otherwise, it takes the value "No". See Table 5 for the details of each condition.

[9]Since information about individuals' smoking behavior is also included in the explanatory variables this poses a spurious relationship between the explanatory variables and the risk for a coronary heart disease. However, since smoking behavior is only one of several factors of the risk score, I consider this spurious relationship sufficiently small.

### 2.2.7 Fitness

During the health check-up, all patients complete a fitness test (spiroergometry), during which the maximum oxygen uptake is measured. In the original dataset, they were then classified into performance quintiles. In collaboration with the check-up provider, I classify patients into the two classes "bad" and "good", depending on whether they belong to the bottom two or top three quintiles of the maximum oxygen uptake distribution among all patients of the same age group and gender.

# 3 Algorithm

Based on this data, I train and test an algorithm that predicts health conditions from easy to collect questionnaire data. For the selection of an appropriate algorithm that is able to predict the different health outcomes, several factors have to be taken into account. One of the most important considerations is the trade-off between bias and variance. An algorithm with a lower bias in its predictions has a higher variance of the predictions across samples, and vice versa. High bias can cause the algorithm to miss relevant relations between variables and outcome (under-fitting), whereas high variance can cause an algorithm to model the random noise in the data (over-fitting).[10] The trade-off implies trying to simultaneously minimize these two sources of error that prevent the algorithm from generalizing beyond the data on which it has been built. Another consideration that has to be taken into account is, that I intend to gain insights about the degree to which a variable is an important predictor (variable importance). In summary, I need an algorithm that is able to efficiently predict the defined health outcomes on out of sample observations while only applying a subset of available variables. While one could repeatedly hand-pick potential variables based on intuition and then use OLS regressions to identify the best combination of variables, I choose to apply a solely data-driven approach. Commonly used tools for these kinds of problems among economists are regularization methods, such as LASSO, ridge regression or elastic net. While LASSO penalizes the absolute size of identified variable coefficients and potentially gets rid of unimportant predictor variables, ridge regressions penalize the squared size of coefficients and hence only shrinks them towards zero. Therefore, I rule out ridge regressions as a potential candidate for identifying a subset of variables that are able to efficiently predict

[10]Brighton (2011) describes this trade-off as a matter of optimizing vs. satisficing, arguing that the complex and uncertain nature of health care questions can often be better addressed with simpler models.

health outcomes. Elastic net penalizes a mix of both, absolute and squared size but therefore incurs high computational costs in order to balance the weights between both constraints. Additionally, all of the regularization methods share the important disadvantage that they cannot easily express non-linear relationships in the data. Since the functional form of the underlying data generating process in my dataset is unknown, all of these three linear regularization methods lack the flexibility that is needed for accurate predictions. One flexible algorithm, that has proven to effectively manage the bias-variance trade-off under high-dimensional data while providing insights into the variable importance is the random forest (Breiman, 2001).

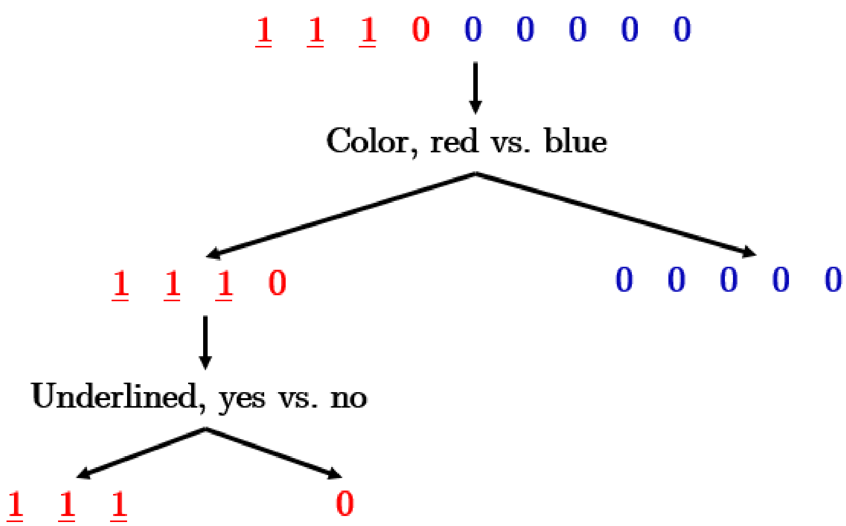
## 3.1 Random Forest: Intuition

Random forests belong to the class of supervised machine learning algorithms, that were developed to provide accurate (out-of-sample) predictions rather than model-based inferences (Mullainathan and Spiess, 2017). Contrary to parametric methods, that imply assumptions about the underlying data generating process, random forests do not rely on any functional form assumptions and are therefore able to pick up highly non-monotonous and complex patterns in data.

A random forest is a collection of decision trees. In order to understand the construction of a decision tree, consider the following simplified example based on Yiu (2019). A training dataset consists of the following nine observations (1, 1, 1, 0, 0, 0, 0, 0, 0). Now I would like to separate the observations into either 1 or 0, using the available two variables color (red and blue) and whether or not the observation is underlined. Since all but one of the 0s are blue, I first split the dataset by color and get the two subsamples (1, 1, 1, 0) and (0, 0, 0, 0, 0). Now I still need to separate the red 0 from the three red 1, 1, 1 in the first subsample in order to split the data perfectly. I apply the second variable whether or not an observation is underlined and get the two subsubsamples (1, 1, 1) and (0). Using the two variables, I was able to classify every observation into one of the two classes 1 or 0. Based on the information in this training data, the decision tree learned that observations that are red and underlined belong to class 1 and observations that are not underlined, either red or blue, belong to class 0. See Figure 2 for an illustration of the constructed decision tree.

As a collection of decision trees, the random forest fits a large number of decision trees to a training dataset using two layers of randomization. The first layer of randomization is that each tree of the random forest is built on a randomly drawn bootstrap sample of

Figure 2: Simplified Example of a Decision Tree




*Notes:* The goal of this decision tree is to divide the observations into 0s and 1s based on the attributes color and whether or not the observation is underlined or not.

observations[11] from the training set. The second layer of observation is that each split within a decision tree uses only a random subset of the available variables[12] in that training set (Liaw *et al.*, 2002). Additionally, there are a number of "hyperparameters" that can be optimized when building a random forest. To name a few, one can vary the number of trees, the depth of each decision tree, or the number of observations in each bootstrap sample.

Once the trees are built, all observations from a test dataset are run through the full set of trees, and the fraction falling into each class is treated as the predicted probability for that outcome. By averaging over a large number of uncorrelated trees, the variance of each individual decision tree is roughly averaged out. Together with its high flexibility to adapt to complex patterns in the data, this explains the typically low bias and moderate variance of the random forest predictions.

[11]In the example above the bootstrap sample was (1, 1, 1, 0, 0, 0, 0, 0, 0) for the construction of this particular decision tree.

[12]In the example above the dataset only consisted of two variables. I considered both variables for each split and decided that color was most appropriate for the first split, and whether or not the number is underlined was more appropriate for the second split. If a dataset consists of more variables, it is common to only consider a random subset of variables for each split.

## 3.2 Random Forest: Set-up

For each outcome variable, I follow similar steps in order to train the random forest and test and evaluate its predictions. First, I split the dataset into a training set and a test set. I allocate 25% of the 6344 observations to the test set, before I fit a set of decision trees (the random forest) to the training set. I optimize the hyperparameters of the random forest by a 5-fold cross-validated grid-search over a parameter grid in order to find the best performing specification. The applied parameter grid includes varying numbers of trees (10, 50, or 100), the maximum depth of each tree (either pure leaves or a maximum of 3 splits), the minimum number of samples required to split a node (2, 10, or 50) a well as the minimum numbers of samples required to be at a leaf node (1, 10, or 50). It also varies the maximum number of variables to consider when searching for the best split ($log_2 * p$ or $sqrt * p$, where p is the total number of available variables, i.e. 72), whether a bootstrap sample or the whole dataset is used to build a tree, and the function that measures the quality of a split (Gini impurity or information gain).[13] The contemplable specifications are evaluated using the $f_1-$score, as the harmonic mean of the model's precision and recall (see Section 3.3 for descriptions of these metrics). Having found the best model based on the training data, I run all observations of the test set through the full set of constructed trees in order to evaluate its performance on observations that have not been considered in the training of the model. I evaluate the resulting predictions on my test set by computing a variety of measures, such as accuracy, precision, recall, $f_1-$score, and a ROC curve in order to get an extensive view on the quality of the model predictions.

Finally, I rank the applied variables according to the extent to which each variable contributes to decreasing the weighted Gini impurity of the random forest. Gini impurity is defined as the probability that a randomly chosen data point in a node is classified incorrectly if it was classified by the distribution of samples in that node.[14] For example, consider a node with 10 observations of which 1 belongs to class A and 9 belong to class B. The Gini impurity would then be calculated as P(classified as A but belongs to B) + P(classified as B but belongs to A) which is P(classified as A) * P(belongs to B) + P(classified as B) * P(belongs to A) $= 0.9 * 0.1 + 0.1 * 0.9 = 0.18$.

[13] I chose this grid specification after a previous randomized tuning round, where I cross-validated a larger set of parameter combinations.

[14] Gini impurity$= \sum_{i=1}^{C} p(i) * (1 - p(i))$, where $C$ is the total number of classes and $p(i)$ is the probability of picking a data point that belongs to class i. A Gini impurity of 0 implies that all the samples in a node are from the same class.

## 3.3 Evaluation Metrics

I apply various metrics in order to evaluate the random forest predictions. One widely used evaluation criterion in the medical diagnosis and forecast verification literature is the area beneath the receiver operating characteristics (ROC-AUC) (Hanley and McNeil, 1982, 1983; Laking *et al.*, 2006; Mason and Graham, 2002). The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels. A threshold level is the probability for which an observation is classified as “true” vs. “false”. The area under the curve (AUC) therefore implies the probability that the random forest assigns a randomly chosen true observation a higher probability for being true than a randomly chosen false one. A perfect model would yield an AUC of 1, whereas random classifications would yield an AUC of 0.5.

However, depending on the desired information gain, other evaluation metrics might be more appropriate. Forming the basis for most other commonly used metrics, a confusion matrix provides a summary of the predictions. Each row of the matrix indicates the actual class and each column represents the predicted class. The cells on the diagonal represent correct predictions whereas the off-diagonal cells show false predictions. Based on the confusion matrix for each health outcome, I calculate the commonly used other evaluation metrics accuracy, precision, and recall. Accuracy gives a brief but incomplete view on the prediction quality as it simply refers to the percentage of correct classifications. However, accuracy does not differentiate between false positives and false negatives and can be very misleading in unbalanced samples. Consider a sample with 100 observations, of which only 1 observation is positive and 99 observations are negative. A simple rule that always classifies every observation as negative would yield an accuracy of 99% but is practically useless for the identification of positive observations – a phenomenon called “accuracy paradox”. In order to get more insights into information retrieval capacities of the classifications, I consider the model’s precision and recall. Whereas precision looks at all the positively classified cases, and counts how many of these are actually positive (in %), recall looks at all of the actually positive cases, and counts how many are classified as positive (in %). Taken together, these two metrics provide insights into how precise the model is in identifying positive cases and how complete this identification is. Depending on the prediction context, one or the other evaluation metric might be more applicable. In order to compare predictions of several health risks with rather balanced classes I consider the ROC-AUC score appealing as it does not depend on the baseline probability of a patient being at risk for a certain health condition. However, since for some health outcomes false negatives might have more severe consequences than false positives, it is important to additionally consider the metrics recall as the complement of false negatives and precision as a complement of false positives.

# 4 Results

In this section, I show the ROC curves and confusion matrices for the predictions of each of the seven health outcomes. Table 6 summarizes the ROC-AUC, accuracy, precision and recall scores of all seven predictions.

Table 6: Summary of Prediction Results

| Health outcome | ROC-AUC | Accuracy | Precision | Recall |
|---|---|---|---|---|
| Blood pressure | 0.69 | 68.92% | 55.90% | 17.79% |
| Diabetes | 0.73 | 69.33% | 60.93% | 57.63% |
| Cholesterol | 0.72 | 68.50% | 69.54% | 86.56% |
| Risk of CHD | 0.94 | 92.19% | 89.09% | 97.38% |
| Plaques | 0.76 | 82.33% | 32.70% | 34.16% |
| Metabolic syndrome | 0.79 | 85.13% | 43.48% | 43.90% |
| Fitness | 0.78 | 70.39% | 76.38% | 67.70% |

*Notes:* This tables shows the ROC-AUC, accuracy, precision and recall scores for the random forest predictions of all seven health outcomes.

## 4.1 Blood Pressure (Figure 3)

Blood pressure was initially categorized into the four separate classes "optimal", "normal", "high-normal", and "high" in order to create a medically relevant prediction problem. For the construction of the ROC curve and the evaluation of the information retrieval capacities of the predictions I pool the two classes optimal and normal under the new label "good" as well as the two classes high-normal and high which I label "bad". Thereby, I consider a binary classification problem in order to be able to compare results across outcomes.

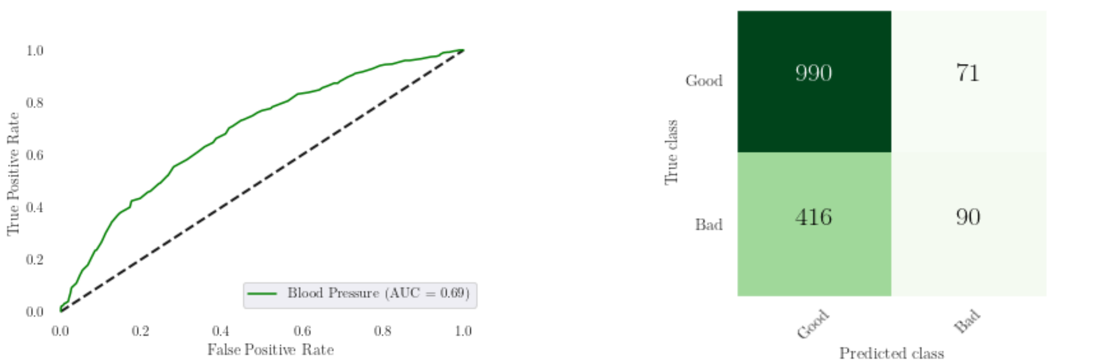
The ROC curve in Figure 3a shows that the random forest discriminates good from bad blood pressure levels rather well (ROC-AUC: 0.69). Looking at the confusion matrix in

Figure 3b, I find that the majority of observations is predicted to have good blood pressure. While this generally holds for the true distribution as well, the extent to which cases are either classified as good class exceeds the actual distribution. While this does not severely harm the prediction accuracy of 68.92% ($\frac{990+90}{990+90+416+71}$), this has larger effect on the information retrieval scores. The random forest only recalls 17.79% ($\frac{90}{90+416}$) of bad cases with 55.90% ($\frac{90}{90+71}$) of those cases being actually bad.

Figure 3: Blood Pressure

(a) ROC curve - Blood Pressure

(b) Confusion Matrix - Blood Pressure




*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions of blood pressure. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

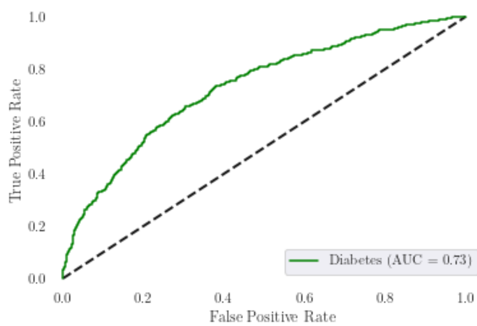
## 4.2 Diabetes (Figure 4)

Blood glucose levels were initially categorized into the three commonly used classes, “no disorder”, “at risk” and “diabetes”. Similarly to the pooling of blood pressure levels, I pool two classes in order to create a binary classification problem for the main analysis. In this case, I pool the two classes “at risk” and “diabetes” and label them “bad”, whereas I relabel the “no disorder” class as “good”.

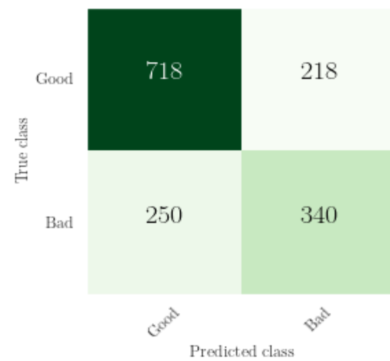
As was the case for the blood pressure predictions, the ROC curve for blood glucose levels shows a descent prediction quality for the random forest predictions (ROC-AUC: 0.73). While the confusion matrix again shows a rather skewed prediction count towards “good” classes, this time the true count shows that pattern to a similar extent. Therefore, next to an accuracy score of 69.33%, the precision of the random forest predictions of diabetes of 60.93% and also the recall of 57.63% indicate a quite well prediction quality.

Figure 4: Diabetes

(a) ROC curve - Diabetes




(b) Confusion Matrix - Diabetes




*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions of diabetes. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

## 4.3 Cholesterol (Figure 5)

For the main analysis of the predictions of cholesterol, I again binarize the outcome variable cholesterol into "good" cholesterol levels vs. the remaining two classes that I label as "bad".

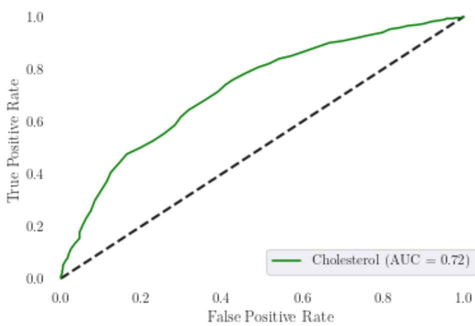
The ROC curve looks very similar to the one for the diabetes predictions with an area under curve of 0.72. However, contrary to the predictions for diabetes, where most classes are correctly predicted as "good", here the majority of classes is correctly predicted to suffer from bad cholesterol levels. Overall, the predictions score rather well on accuracy (68.50%), precision (69.54%) and especially recall, as the random forest finds 86.56% of all "bad" cases.

## 4.4 Coronary Heart Disease (Figure 6)

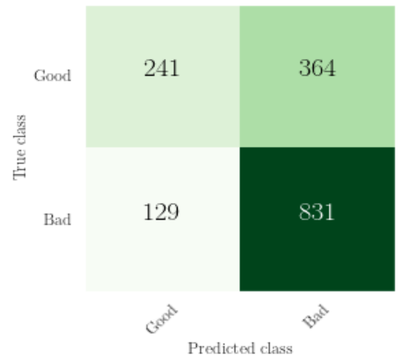
For the risk of suffering from a CHD, I find that the random forest performs strikingly well. With an ROC-AUC score of 0.94 the ROC curve comes rather close to perfect predictions. In addition, the confusion matrix shows an extremely good prediction quality of the random forest. With an accuracy of 92.19%, the CHD predictions outperform all other predictions. Moreover, 89.09% of all observations that the random forest classifies as "high risk" are indeed "high risk" cases. Finally, only 22 of 839 high risk classes are falsely classified as low risk, resulting in a recall score of 97.38%.

Figure 5: Cholesterol

(a) ROC curve - Cholesterol



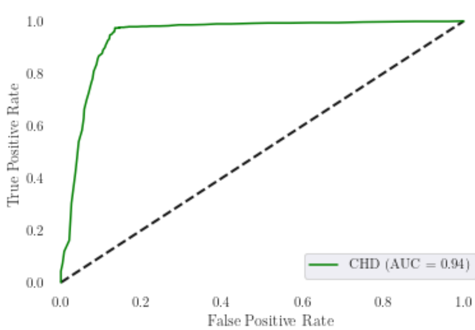

(b) Confusion Matrix - Cholesterol



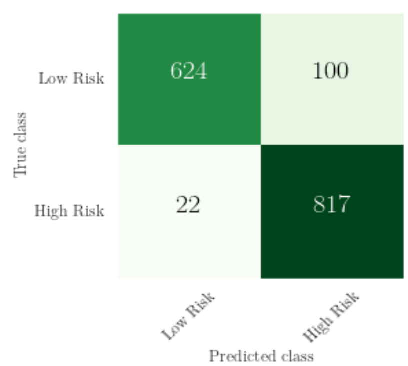

*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions of cholesterol levels. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

Figure 6: Coronary Heart Disease

(a) ROC curve - Coronary Heart Disease




(b) Confusion Matrix - Coronary Heart Disease



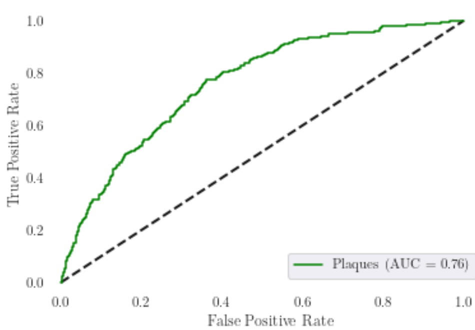

*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions for the risk of getting a coronary heart disease. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

## 4.5 Plaques (Figure 7)

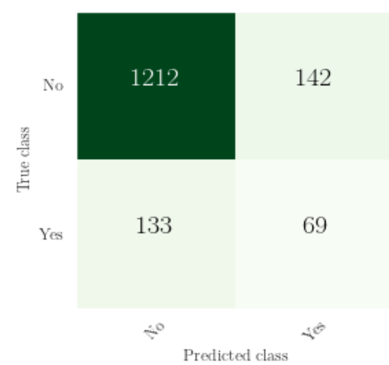
The metrics for the predictions for plaques show mixed results. While the ROC curve (ROC-AUC: 0.76) and the prediction accuracy of 82.33% point towards quite a good prediction performance, the precision and recall scores of 32.70% and 34.16%, respectively, suggest otherwise. The confusion matrix provides an explanation for these scores. A vast majority of cases (87%) in the used test set does not suffer from plaques. Under this circumstance, a classifier that mostly predicts negative cases shows good ROC and accuracy scores, while underperforming in precision and recall. This is an example of the previously mentioned "accuracy paradox" (see Section 3.3).

Figure 7: Plaques

(a) ROC curve - Plaques




(b) Confusion Matrix - Plaques




*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions for suffering from plaques. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

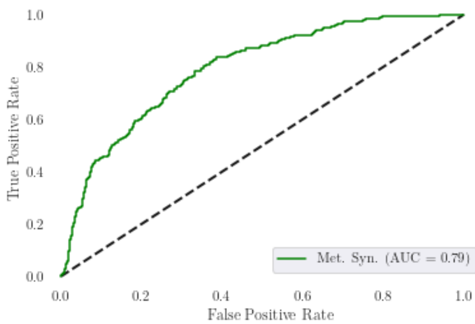
## 4.6 Metabolic Syndrome (Figure 8)

Similar to the results for plaques, the predictions for metabolic syndrome suggest a similar story, albeit not as distinctive. While the ROC curve (ROC-AUC: 0.79) and the prediction accuracy of 85.13% point towards a very good prediction performance, the precision score of 43.48% as well as the recall score of 43.90% raise some doubts on the prediction quality.

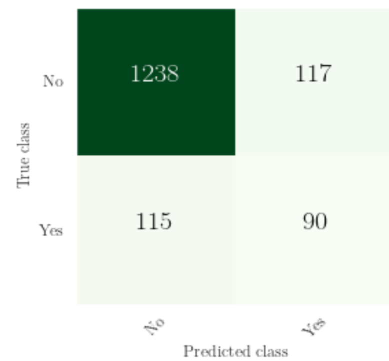
Again, the confusion matrix provides the explanation by showing a very unbalanced sample in which most patients do not suffer from a metabolic syndrome.

Figure 8: Metabolic Syndrome

(a) ROC curve - Metabolic Syndrome




(b) Confusion Matrix - Metabolic Syndrome



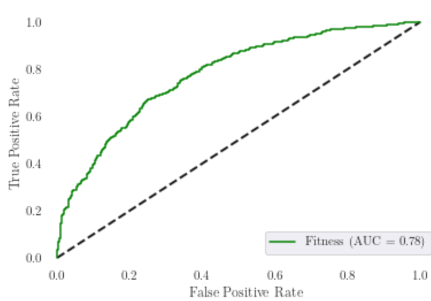

*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions for suffering from a metabolic syndrome. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

## 4.7 Fitness (Figure 9)

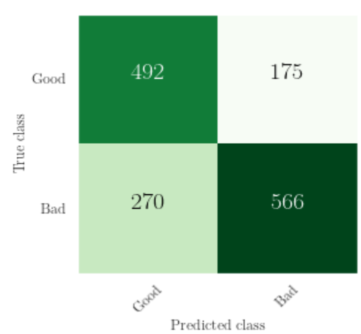
Finally, the ROC curve for the Fitness predictions shows a good prediction quality of the random forest (ROC-AUC: 0.78). Contrary to the confusion matrices for the plaques and metabolic syndrome predictions, the confusion matrix for fitness shows a more balanced sample. This is reflected in the consistently solid scores for accuracy (70.39%), precision (76.38%) and recall (67.70%).

Figure 9: Fitness

(a) ROC curve - Fitness




(b) Confusion Matrix - Fitness




*Notes:* The left panel shows the ROC curve that plots the true positive rate (TPR) against the false positive rate (FPR) for various threshold levels for the random forest predictions of a patient's relative fitness level. The right panel shows the confusion matrix based on which the accuracy, precision and recall scores are calculated.

## 4.8 Variable Importance

Having established the quality of the random forest predictions, I now highlight the most important predictor variables for each health outcome. Based on this, one can then set up a concise questionnaire tool that can be used for predicting a particular health outcome as the prediction performance is mainly determined by these variables.[15] As described in Section 3.2, I rank the applied variables according to their contribution in decreasing the weighted Gini impurity of the random forest. Table 7 summarizes the top ten important variables for each health outcome.

Overall, it can be seen that the characteristics weight, height and the amount of alcohol per week are among the most important variables for all of the predicted health outcomes. Additionally, age, hours of cardio sports per week and the actual working hours per week are among the top ten variables for six out of seven health outcomes. While the health outcomes fitness and metabolic risk rely mainly on health behavior variables such as smoking and sports behavior, the remaining five health outcomes all consider various psychological variables. Among these, the focus lies on the external resources integrity of principals and peers and a patient's role clarity at work as well as on internal resources such as a patient's self-regulation and commitment capabilities. Moreover, the overall perception of private stress is the most relevant stress factor. However, one needs to keep in mind that the importance decreases sharply with each variable, leaving only a minor part for the sixth to tenth important variables for most health outcomes. Still, the mentioned variables are consistently rated among the top ten important variables for varying parameter specifications of the random forest.

# 5 Random Forest vs. Physicians: Survey Design

In order to gain more insights into the applicability of the random forest results, I compare these predictions to predictions of physicians. I focus on the physicians that initially conducted the health check-up on the patients in the dataset. In particular, using the same set of patient observations as for the random forest predictions, I elicit physician predictions for all seven health outcomes. The survey was based on OTree (Chen *et al.*, 2016) and conducted

[15] I train and test random forest models only using the respective top ten variables and find roughly similar prediction results as under the full questionnaire data.

Table 7: Top 10 Important Variables

| Rank | Blood Pressure | Diabetes | Cholesterol | Risk of CHD |
|---|---|---|---|---|
| 1 | Weight | Weight | Weight | Gender |
| 2 | Alcohol | Age | Alcohol | Height |
| 3 | Height | Cardio | Cardio | Weight |
| 4 | Age | Height | Height | Working hours |
| 5 | Cardio | Alcohol | Age | Alcohol |
| 6 | Integrity | Integrity | Integrity | Age |
| 7 | Role clarity | Role clarity | Role clarity | Leadership |
| 8 | Working hours | Working hours | Working hours | Cardio |
| 9 | Private stress | Personal effort | Private stress | Role clarity |
| 10 | Commitment | Self-regulation | Self-regulation | Private stress |

| Rank | Plaques | Metabolic Syndrome | Fitness |
|---|---|---|---|
| 1 | Weight | Age | Weight |
| 2 | Height | Weight | Cardio |
| 3 | Cardio | Alcohol | Height |
| 4 | Age | Smoke duration | Strength |
| 5 | Working hours | Height | Strength detail |
| 6 | Smoke amount | Gender | Smoke duration |
| 7 | Alcohol | Smoke amount | Smoke amount |
| 8 | Strength | Working hours | Alcohol |
| 9 | Gender | Smoke | Age |
| 10 | Commitment | Not smoke duration | Working hours |

*Notes:* This table shows the top ten most important predictor variables for each health outcome, measured by their respective contribution to decreasing the weighted Gini impurity.

online between November 2018 and January 2019. In total four[16] physicians participated in the survey, resulting in a total of 156 physician predictions.

I provided each of the four participating physicians with 60 randomly drawn, numbered patient records from the test set in PDF format, that they could access online via a link to a secured online server.[17] These patient records entailed the exact same amount of infor-

[16] I contacted all nine physicians of the health check-up provider but unfortunately only four of them finally participated in the survey.

[17] See Appendix C.1 for an exemplary patient record (in German).

mation as was used for the predictions by the random forest. Furthermore, the physicians received detailed instructions about the procedure and the exact definitions of the outcome variables.[18] Finally, the physicians received a link to the survey.

Once a physician opened this link, they were provided with some more information and asked to open the first patient record. On the next screen, the physician was requested to provide the unique patient ID that was indicated at the top of the patient record. Subsequently, they were asked to provide probabilities that the patient belonged to a particular class of the pre-defined health outcomes. For example:

> *What is the probability, that this patient has optimal blood pressure*?
> *What is the probability, that this patient has normal blood pressure?*
> *What is the probability, that this patient has high-normal blood pressure?*
> *What is the probability, that this patient has high blood pressure?*

Below each question, I additionally displayed the exact definition of the respective class (thus, in this case the definitions for “optimal blood pressure”, “normal blood pressure”, “high-normal blood pressure”, “high blood pressure”).

Once a physician had completed the predictions of all seven health outcomes for a respective patient, they were asked to open the next patient record, until 30 patient records were processed. Next, for each health outcome the physician was asked to indicate at least three and at maximum ten variables that they considered when making the predictions. Finally, similar to before, the physician was asked to predict the health outcomes of another 30 randomly drawn patients. However, on each page of this second round, I additionally informed the physicians about the random forest prediction probabilities for each class of each health outcome before they could enter their predictions.

In order to motivate accurate predictions, I provided a monetary incentive for the physicians: First, I randomly selected 10 out of the 60 patient records that each physician has been provided with. For each of the three non-binary outcome variables, I compare the class that the physician assigned the highest probability to and check whether this class indeed reflected the true diagnosis in the data. If it did, the physician earned 50 cents for this prediction. If they missed the true class by one class, they still earned 25 cents. For each of the four binary outcome variables, the physician received 50 cent if they assigned at least

[18]See Appendix C.2 for the original instructions (in German).

50% probability to the true class. Choosing 10 patient records with seven outcomes each, the physician could thus earn a maximum of 35€. Additionally, each physician was paid a fixed fee of 25€.

In total, I received 96 completed patient records for the predictions without random forest support (Round 1) and 60 completed patient records for the predictions with random forest support (Round 2). Three physicians completed the variable importance round between the two prediction rounds.

# 6 Random Forest vs. Physicians: Results

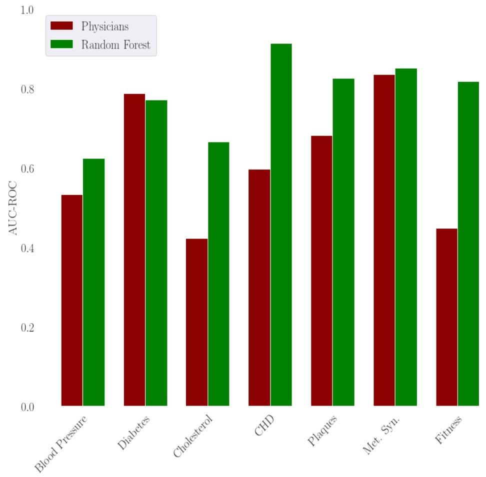
In order to compare the random forest predictions to the physician predictions, I first construct a comparable evaluation set from the random forest predictions. This means, I only consider the 156 observations from the initial test set, that also received a prediction from one of the physicians. I then compare the ROC curve as well as the accuracy, precision and recall scores between the predictions from the physicians and the predictions from the random forest on the same observations. Thereby, I split the analysis into two parts. I first only consider the patient predictions for which the physicians did not receive the random forest prediction as additional information and second I consider the predictions for which they received that additional piece of information. Finally, I highlight the variables that the physicians considered important for their predictions and compare them to the respective top ten important variables of the random forest.

## 6.1 Comparing the Prediction Performances

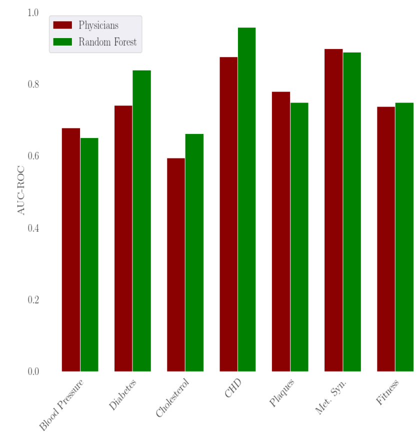
Figures 10a and 10b show the ROC-AUC scores for the physicians and the random forest for all health outcomes. For the predictions that were based on similar information sets, (i.e. for which the physicians did not receive the random forest prediction as additional information), the random forest outperforms the physicians in all health outcomes, except for diabetes and metabolic syndrome where the scores are roughly equal (Figure 10a). The largest differences can be found for the risk of getting a coronary heart disease, for which the random forest yielded almost perfect predictions and for the patients' relative fitness level. However, after providing the physicians with the random forest predictions, the physicians greatly improve (Figure 10b). I now find roughly equal ROC-AUC scores for almost all

Figure 10: ROC-AUC and Accuracy Scores

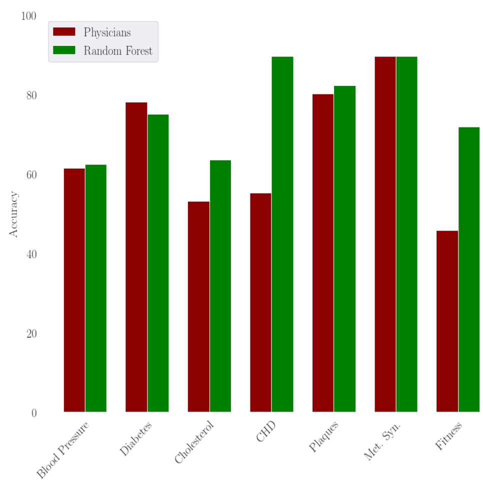
(a) ROC-AUC round 1 (no rf support)



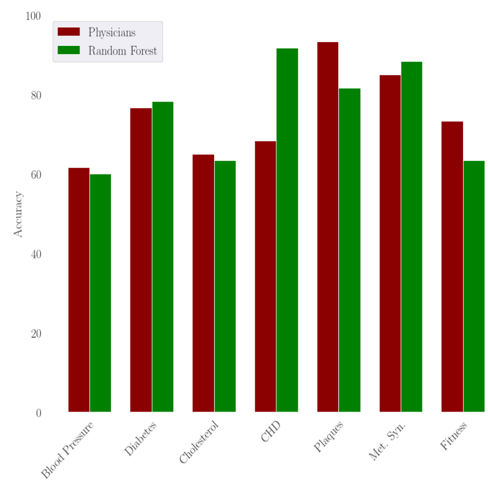

(b) ROC-AUC round 2 (with rf support)




(c) Accuracy round 1 (no rf support)




(d) Accuracy round 2 (with rf support)




*Notes:* The upper panel shows the ROC-AUC scores (the probability, that a randomly drawn positive case has a higher likelihood to be predicted positive than a randomly drawn negative case) for both physician (red) and random forest (green) predictions. The lower panel shows the accuracy scores (the percentage of correctly classified observation). The left column depicts results for round 1 (without the random forest predictions as additional information for the physicians), the right columns shows the results of round 2.

health outcomes. Thereby it is important to note that ROC-AUC scores for the random forest predictions stayed roughly the same between round 1 and round 2. This can be interpreted as a successful randomization check as the performance of the random forest should not be affected by drawing a different random subsample of the test set.

A very similar picture emerges when we consider the accuracy scores. Again, I find the random forest weakly outperforming the physicians across all health outcomes when physician predictions were based on the same information set as the random forest predictions (Figure 10c). Providing the physicians with the random forest predictions improved the accuracy of their predictions so that they even outperform the random forest slightly for plaques and fitness (Figure 10d). Again, the random forest prediction accuracy did not vary much between the two rounds.

The precision and recall scores show similar stories (Figure 11). While the random forest scores stay roughly constant between rounds the physicians greatly improve under the additional information of the random forest prediction probabilities.[19]

## 6.2 Comparing the Variable Importances

Finally, I compare the top ten important variables for each health outcome according to the physicians to the top ten important variables for each health outcome of the random forest predictions. For each health outcome, I pool the top ten variables over physicians and rank them according to their prevalence. This ranking constitutes the top ten for each health outcome for the physicians. Figure 12 plots the percentage use of each of the four variable categories (general, health behavior, psychological, and family health history) for physicians and random forest. I aggregated the top ten variables for each health outcome, allocated the variables into their respective categories[20] and divided the counts per category by the total amount of variables (72).

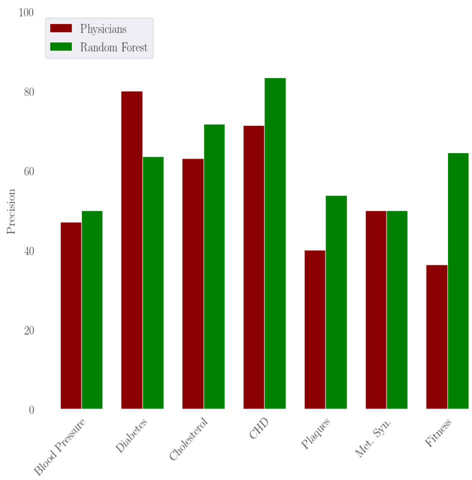
While both types of predictions used general and health behavior variables to similar extents, there are substantial differences in their use of psychological and family health history variables. The only psychological variable that physicians considered important was

[19]The random forest precision and recall scores for plaques and metabolic syndrome decrease from round 1 to round 2, but this can be explained by a few outlier observations. The fact that both scores are affected and the decreased physicians recall score for metabolic syndrome support this argument.

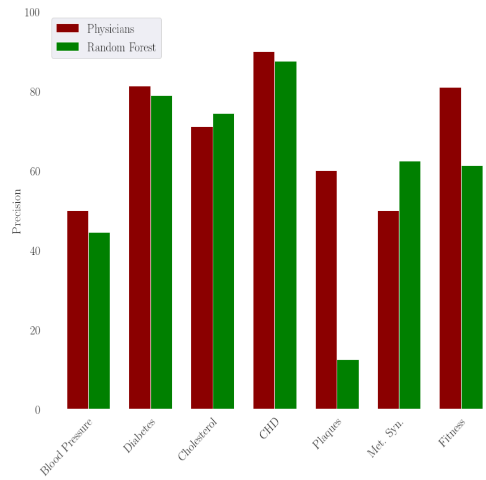
[20]I counted each variable only once, even it was mentioned for several outcomes. Figure 23 in Appendix C.4 shows the same plot with each variable counted only once even it was mentioned for more than one health outcome. Results are similar.

Figure 11: Precision and Recall Scores

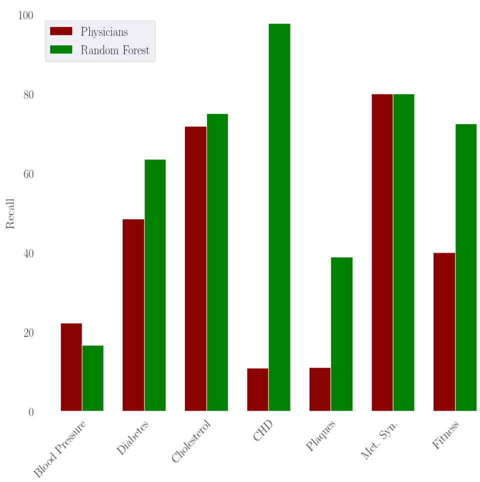
(a) Precision round 1 (no rf support)



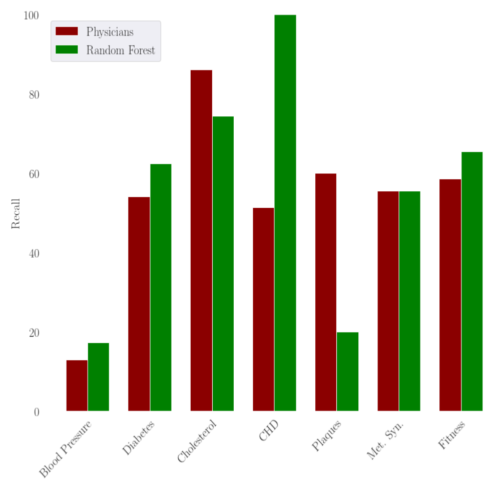

(b) Precision round 2 (with rf support)




(c) Recall round 1 (no rf support)

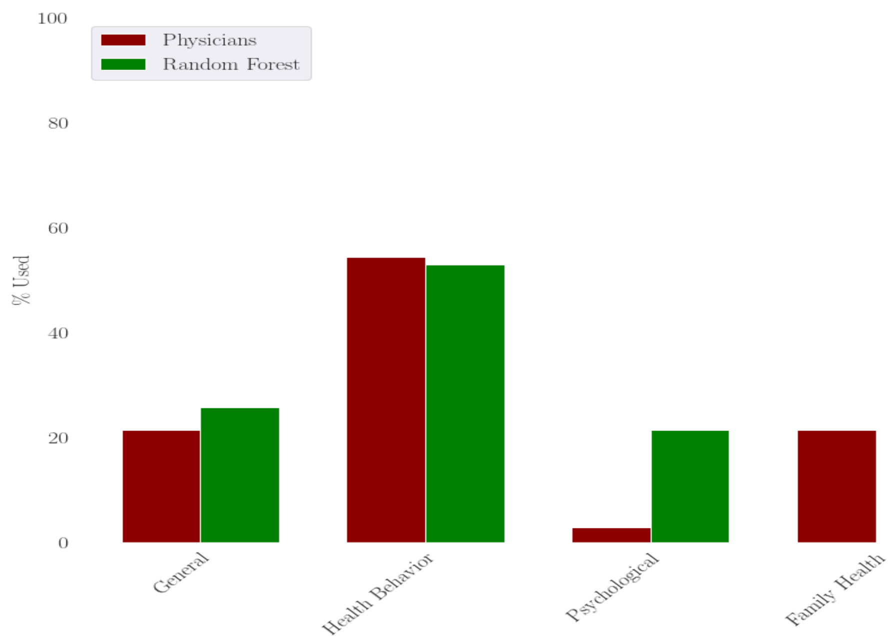



(d) Recall round 2 (with rf support)




*Notes:* The upper panel shows the precision scores (of all positively classified cases, what percentage is actually positive) of the physician predictions (red) vs. the random forest predictions (green). The lower panel shows the recall scores (of all actually positive cases, what percentage is classified as positive). The left column depicts results for round 1 (without the random forest predictions as additional information for the physicians), the right column shows the results of round 2.

Figure 12: Variables per Category




*Notes:* This figure shows the percentage use of each variable category for physicians (red) and random forest (green), based on the top 10 variables of each of the seven outcomes.

a patient's motivational control (for the prediction of fitness) whereas 20% of all variables that the random forest used were psychological variables. Contrary, the physicians placed much greater value on family health history variables 20%, whereas the random forest did not consider any family health history variable at all. Table 9 in Appendix C.4 lists the top ten variables for each health outcome for physicians and random forest.

# 7 Discussion

Predicting health outcomes from questionnaire data can be very useful in a variety of applications. One apparent application is the practical use of preventing costly treatments and pain by timely identification of potentially hazardous health conditions. Other potential applications are in the field of empirical research. Numerous studies base their results on self-reported health estimates, thereby risking a certain degree of measurement inaccuracy (Jürges, 2007). As an example, one potential source of biased self-reports is social desirability, implying that individuals underreport socially undesirable and overreport more desirable individual health information (Latkin *et al.*, 2017). Providing a validated questionnaire tool to predict health outcomes may be a first step towards more reliable estimates. Recent progress in prediction methods via machine learning enables us to improve upon previous health estimation and allows us to gather more accurate and reliable health information.

However, it is important to keep in mind that computer based predictions should not be seen as a substitute, but rather as a complement to physician visits. Potential treatments based on the identification of risk factors from computerized methods should subsequently be discussed and potentially implemented with physicians. Physicians usually base their final diagnosis predictions on much more information that they gather during actual patient visits, where such questionnaire data from an anamnesis only serve as complementary information.

As a necessary condition for machine learning methods to be effectively applied, physicians as well as patients need to have a certain degree of trust in the predictions. In light of the current use of machine learning methods in different fields, the question whether to trust machine learning predictions is widely debated (Spiegelhalter, 2020; Yeomans *et al.*, 2019). Even though this was not the central focus of this study, I provide first evidence towards physicians trust in a particular machine learning algorithm and that trusting the algorithm results in improved predictions. However, this result should by no means interpreted as a general trust among physicians towards machine learning algorithms in general. Based on the results I can only infer that a small sample of physicians incorporated information from a specific algorithm into account for predictions of particular health risk factors in a controlled study. I cannot infer whether or not they would also rely on machine learning algorithms in different tasks (e.g. the prediction of life threatening conditions) and settings (e.g. if prediction errors have severely harmful consequences for them or the patients).

While this paper shows a relatively distinct performance difference between computerized and physician predictions, it does not provide conclusive evidence of the underlying reasons for the inferior physician predictions. Wegwarth and Gigerenzer (2011) point out that not only patients suffer from statistical illiteracy in health (Gaissmaier and Gigerenzer, 2011), but also physicians often fail to understand health statistics. Consequently, this prevents informed shared decision making and impairs patients' knowledge (Gigerenzer *et al.*, 2009) in various health matters, such as screening effectiveness, diagnoses certainties, or treatment chances. Since our survey required physicians to express diagnoses in probabilities, statistical illiteracy, at least compared to perfect statistical literacy of computers, could explain parts of the inferior performance of the physicians. Moreover, the chosen patient attributes of section 6.2 give first insights into the different approaches of physicians and computers and show that physicians rely less on seemingly relevant information of psychological variables. However, the paper cannot distinguish whether the prediction differences are mainly driven by the different choices of variables between physicians and the random forest or rather by physicians' biased interpretation of certain variables. As an example, it may potentially be possible that while both consider the variable "height" an important attribute for a patient's blood pressure, physicians have a biased perception of the magnitude and/or direction of its impact.

In their paper on frictions and mental gaps, Handel and Schwartzstein (2018) argue, that two classes of policies can improve decision-making in the context of information processing. An "allocation policy" directly steers consumers to a specific action, e.g. through nudges, and does not rely on whether the initial information avoidance was a result of frictions or mental gaps. Contrary, a "mechanism policy" targets a particular friction or mental gap and hence critically depends on the role that the respective mechanism played in driving choices. Providing physicians with random forest predictions of a particular health risk is an example of a mechanism policy. One mechanism through which this policy could work is, that potential information processing cost of physicians are lowered. By aggregating information of 72 variables into one validated estimate, I reduce the cost of processing all the available information and enable cognitively constrained physicians (as compared to the random forest) to consider a larger amount of information. The effectiveness of the information provision in this setting can thus be considered as first evidence for frictions playing a role in the weaker prediction performance of physicians. In this case, ways that lower the cost of processing information further could additionally improve the physicians' predictions. However, this does by no means exclude other potential mechanisms. The random forest predictions might be perceived as a very salient piece of information, whereas the psychological variables were not and hence this new piece of information drives the physician predictions through a particular mental gap as described by Bordalo *et al.* (2013). In this case, an improved education about the analysis of questionnaire data could additionally alleviate the mental gap and improve physicians' predictions without any computer assistance. Future studies should have a closer look at the impact of frictions and mental gaps in this setting in order to derive meaningful and tailored policy implications.

# 8 Conclusion

I apply a flexible machine learning algorithm to predict various health outcomes from rather impersonal questionnaire data. I find that random forest predictions provide decent predictions across all applied health outcomes. Further, I identify a set of important predictor variables that contain meaningful information about individuals' health conditions. Comparing the random forest predictions to physician predictions, I find that the random forest predictions outperform physician predictions in almost all health outcomes as long as both are provided with the same set of information for their predictions. I further find that the random forest uses different patient characteristics for its predictions than physicians, which may partly explain the differences in prediction qualities. In order to shed more light onto the channel that can explain the differences, future research needs to look at the amount of information that is processed and the way in which information is interpreted by both the physicians and the random forest. When providing the physicians with random forest predictions, the physician predictions greatly improve, suggesting that physicians trust the random forest predictions and are willing to adapt their own estimates. Another fruitful area for future research in this context is to compare the random forest predictions to physician predictions during a patient visit (but before actual medical test) as this provides a more natural setting for physicians.

## Acknowledgements

I gratefully acknowledge the constant advice and assistance as well as the financial support from Michael Kosfeld, without whom this paper would not have been possible. I also wish to express my sincere gratitude to Dr. med. Jochen Haack and Dr. Nadine Schuster for providing the data and assisting with valuable discussions. Furthermore, I am thankful for many valuable comments and discussions by the participants of the Machine Learning Summer School 2018 in Madrid. I particularly thank Florian Hett, Simon Heß, Fabian Horst, Johannes Kasinger and Patrick Schmidt for detailed feedback on different parts of the paper.

# References


Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment.

Ahmed, Z., Mohamed, K., Zeeshan, S. and Dong, X. Q. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database : the journal of biological databases and curation*, **2020**.

Athey, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An agenda*, University of Chicago Press, pp. 507–552.

Bartling, B., Fehr, E. and Schunk, D. (2012). Health effects on children's willingness to compete. *Experimental Economics*, **15** (1), 58–70.

Berry, J. D., Dyer, A., Cai, X., Garside, D. B., Ning, H., Thomas, A., Greenland, P., Van Horn, L., Tracy, R. P. and Lloyd-Jones, D. M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, **366** (4), 321–329.

Bordalo, P., Gennaioli, N. and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, **121** (5), 803–843.

Breiman, L. (2001). Random forests. *Machine learning*, **45** (1), 5–32.

Brighton, H. (2011). The Future of Diagnostics From Optimizing to Satis¿ cing. In *Better Doctors, Better Patients, Better Decisions*, The MIT Press.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, **105** (7), 2183–2203.

Chen, D. L., Schonger, M. and Wickens, C. (2016). {oTree} – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, **9**, 88–97.

Chernozhukov, V., Demirer, M., Duflo, E. and Fernández-Val, I. (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments.

Cline, R. J. W. and Haynes, K. M. (2001). Consumer health information seeking on the Internet: the state of the art. *Health education research*, **16** (6), 671–692.

Collins, F. S. and Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, **372** (9), 793–795.

Diviani, N., van den Putte, B., Giani, S. and van Weert, J. C. M. (2015). Low health literacy and evaluation of online health information: a systematic review of the literature. *Journal of medical Internet research*, **17** (5).

Enke, B. and Zimmermann, F. (2019). Correlation Neglect in Belief Formation. *Review of Economic Studies*, **86** (1), 313–332.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, **129** (4), 1661–1710.

Gaissmaier, W. and Gigerenzer, G. (2011). When Misinformed Patients Try to Make Informed Health Decisions. In *Better Doctors, Better Patients, Better Decisions*, MIT Press, pp. 29–44.

Gigerenzer, G., Mata, J. and Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of the National Cancer Institute*, **101** (17), 1216–1220.

Goldman, N., Lin, I. F., Weinstein, M. and Lin, Y. H. (2003). Evaluating the quality of self-reports of hypertension and diabetes. *Journal of Clinical Epidemiology*, **56** (2), 148–154.

Handel, B. and Schwartzstein, J. (2018). Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care? *Journal of Economic Perspectives*, **32** (1), 155–178.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143** (1), 29–36.

— and — (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, **148** (3), 839–843.

Hu, Y., Zong, G., Liu, G., Wang, M., Rosner, B., Pan, A., Willett, W. C., Manson, J. A. E., Hu, F. B. and Sun, Q. (2018). Smoking cessation, weight change, type 2 diabetes, and mortality. *New England Journal of Medicine*, **379** (7), 623–632.

Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics*, **16** (2), 163–178.

Koszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics*, **128** (1), 53–104.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, **13**, 8–17.

Kriegsman, D. M., Penninx, B. W., Van Eijk, J. T. M., Boeke, A. J. P. and Deeg, D. J. (1996). Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *Journal of Clinical Epidemiology*, **49** (12), 1407–1417.

Laking, G., Lord, J. and Fischer, A. (2006). The economics of diagnosis. *Health economics*, **15** (10), 1109–1120.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A. and Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive behaviors*, **73**, 133–136.

Lee, S. I., Celik, S., Logsdon, B. A., Lundberg, S. M., Martins, T. J., Oehler, V. G., Estey, E. H., Miller, C. P., Chien, S., Dai, J., Saxena, A., Blau, C. A. and Becker, P. S. (2018). A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nature Communications*, **9** (1).

Liaw, A., Wiener, M. and Others (2002). Classification and regression by randomForest. *R news*, **2** (3), 18–22.

Manson, J. E., Stampfer, M. J., Colditz, G. A., Willett, W. C., Rosner, B., Hennekens, C. H., Speizer, F. E., Rimm, E. B. and Krolewski, A. S. (1991). Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *The Lancet*, **338** (8770), 774–778.

Margolis, K. L., Qi, L., Brzyski, R., Bonds, D. E., Howard, B. V., Kempainen, S., Liu, S., Robinson, J. G., Safford, M. M., Tinker, L. T. and Phillips, L. S.

(2008). Validity of diabetes self-reports in the Women's Health Initiative: Comparison with medication inventories and fasting glucose measurements. *Clinical Trials*, **5** (3), 240–247.

Mason, S. J. and Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **128** (584), 2145–2166.

McCall, J. J. (1970). Economics of Information and Job Search: Reply. *The Quarterly Journal of Economics*, **86** (1), 132.

MFMER (2020a). Carotid artery disease.

— (2020b). Cholesterol.

— (2020c). Diabetes.

— (2020d). hypertension.

Midthjell, K., Holmen, J., Bj0rndal, A. and Lund-Larsen, G. (1992). Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Tr0ndelag diabetes study. *J7ournal of Epidemiology and Community Health*, **46**, 537–542.

Mullainathan, S. and Obermeyer, Z. (2017). Does machine learning automate moral hazard and error? *American Economic Review*, **107** (5), 476–480.

— and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, **31** (2), 87–106.

Pokorska-Bocci, A., Stewart, A., Sagoo, G. S., Hall, A., Kroese, M. and Burton, H. (2014). 'Personalized medicine': What's in a name?

Rajkomar, A., Dean, J. and Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, **380** (14), 1347–1358.

Rohmert, W., Rutenfranz, J. and Luczak, H. (1975). Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen. In *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen*, Bundesminister für Arbeit und Sozialordnung, Referat Öffentlichkeitsarbeit, pp. 15–250.

Ryan, A. and Wilson, S. (2008). Internet healthcare: do self-diagnosis sites do more harm than good? *Expert opinion on drug safety*, **7** (3), 227–229.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, **12** (6), 1423–1452.

Seetharam, K., Shrestha, S. and Sengupta, P. P. (2019). Artificial Intelligence in Cardiovascular Medicine. *Current Treatment Options in Cardiovascular Medicine*, **21** (6).

Shickel, B., Tighe, P. J., Bihorac, A. and Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, **22** (5), 1589–1604.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, **50** (3), 665–690.

Spiegelhalter, D. (2020). Should We Trust Algorithms? *Harvard Data Science Review*.

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy*, **69** (3), 213–225.

Tang, F. and Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining*, **10** (6), 363–377.

Tretli, S., Lund-Larsen, P. G. and Foss, O. P. (1982). Reliability of questionnaire information on cardiovascular disease and diabetes: Cardiovascular disease study in Finnmark county. *Journal of Epidemiology and Community Health*, **36** (4), 269–273.

Wager, S. (2014). Asymptotic Theory for Random Forests. *arXiv preprint arXiv:1405.0352*.

— and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, **113** (523), 1228–1242.

Wallace, T. M., Levy, J. C. and Matthews, D. R. (2004). Use and abuse of HOMA modeling. *Diabetes care*, **27** (6), 1487–1495.

Wegwarth, O. and Gigerenzer, G. (2011). Statistical Illiteracy in Doctors. In *Better Doctors, Better Patients, Better Decisions*, MIT Press.

WHO (2011). *Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation*. Tech. rep., Geneva: World Health Organization.

Wong, D., Sahni, A., Bedford, J., Harris, S. and Moonesinghe, S. (2018). Man vs machine: how good are clinicians at predicting perioperative risk? *British Journal of Anaesthesia*, **121** (2), e28–e29.

Woodford, M. (2012). Inattentive Valuation and Reference-Dependent. *Unpublished Manuscript, Columbia University*, p. 89.

Yeomans, M., Shah, A., Mullainathan, S. and Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, **32** (4), 403–414.

Yiu, T. (2019). Understanding Random Forest. https://towardsdatascience.com/understanding-random-forest-58381e0602d2# (accessed on 10/10/2020).

# Appendix

## A Data

### A.1 Psychological Variables

This section provides the full set of items of all psychological variables by showing the original questionnaire regarding the psychological variables. Table 8 shows the sections and question numbers for each variable.

| I. 1 Arbeitstätigkeiten sind durch bestimmte Anforderungen charakterisiert. Geben Sie bitte an, in welchem Ausmaß die nachfolgenden Anforderungen auf Ihre Arbeit zutreffen. | trifft gar nicht zu<br>1 | trifft wenig zu<br>2 | teils/teils<br>3 | trifft überwiegend zu<br>4 | trifft völlig zu<br>5 |
|---|---|---|---|---|---|
| 1. Meine Arbeit verlangt von mir, niemals die Beherrschung zu verlieren. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Um meine Arbeitsziele zu erreichen, darf ich mich nicht ablenken lassen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Um mein Arbeitspensum zu schaffen, muss ich mich dazu zwingen, keine Zeit mit Nebensächlichkeiten zu vergeuden. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Wenn ich meine Arbeit erfolgreich bewältigen will, darf ich irgendwelchen Ablenkungen nicht nachgeben. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Auch wenn ich manchmal sehr gereizt bin, darf ich mir das auf keinen Fall anmerken lassen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Bestimmte Aufgaben in Angriff zu nehmen, kostet mich manchmal einiges an Überwindung. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Einige meiner Arbeitsaufgaben sind so, dass ich mich richtig zwingen muss, sie zu erledigen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Bei meiner Arbeit darf ich nie ungeduldig werden. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Einige meiner Arbeitsaufgaben kann ich nur gegen innere Widerstände bearbeiten. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Bei meiner Arbeit darf ich mich niemals gehen lassen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Bei der Arbeit stehe ich häufig unter Zeitdruck. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Ich habe zu viel Arbeit. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 13. Mein Arbeitspensum ist häufig nur zu schaffen, wenn ich auf Pausen verzichte. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 14. Bei meiner Arbeit gibt es Sachen, die zu kompliziert sind. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 15. Meine Arbeit stellt zu hohe Anforderungen an meine Konzentrationsfähigkeit. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 16. Meine Arbeitsaufgaben sind so vielfältig, dass man leicht den Überblick verliert. | ☐ | ☐ | ☐ | ☐ | ☐ |

| I. 2 An vielen Arbeitsplätzen ist es erforderlich, dass man in bestimmter Weise mit eigenen und fremden Gefühlen umgehen muss. Wie häufig treten in Ihrer Arbeit die folgenden Situationen auf? | sehr selten/nie<br>1 | selten<br>2 | gelegent-lich<br>3 | oft<br>4 | immer<br>5 |
|---|---|---|---|---|---|
| 1. Wie oft kommt es bei Ihrer Arbeit vor, dass Sie nach außen hin Gefühle zeigen müssen, die nicht mit dem übereinstimmen, was Sie anderen Menschen gegenüber tatsächlich fühlen? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Wie oft kommt es bei Ihrer Arbeit vor, dass Ihre eigentlich erlebten Gefühle nicht denen entsprechen, die Sie im Umgang mit anderen Menschen zeigen sollten? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Wie häufig erleben Sie bei Ihrer Arbeit einen Konflikt zwischen Ihren eigentlichen Gefühlen und den Gefühlen, die Sie nach außen hin anderen Menschen gegenüber zeigen sollten? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Wie häufig kommt es bei Ihrer Arbeit vor, dass Sie bestimmte Gefühle zum Ausdruck bringen müssen, die Sie eigentlich nicht empfinden? | ☐ | ☐ | ☐ | ☐ | ☐ |

| I. 3 Im Folgenden finden Sie einige Aussagen zu Vorgesetzten und Kollegen. Inwieweit treffen diese Aussagen auf Ihren Vorgesetzten/Ihre Kollegen zu? | trifft gar nicht zu<br>1 | trifft eher nicht zu<br>2 | teils/teils<br>3 | trifft eher zu<br>4 | stimmt völlig<br>5 |
|---|---|---|---|---|---|
| 1. Mein Vorgesetzter teilt die angenehme Arbeit immer bestimmten Leuten zu. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Mein Vorgesetzter spielt die Kollegen gegeneinander aus. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Wenn ein Fehler passiert, findet der Vorgesetzte ihn immer bei uns, nie bei sich. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Es ist offensichtlich, dass es unter meinen Arbeitskollegen persönliche Konflikte gibt. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Die Beziehungen unter meinen Arbeitskollegen sind nicht immer harmonisch. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Unter meinen Arbeitskollegen bestehen zwischenmenschliche Spannungen. | ☐ | ☐ | ☐ | ☐ | ☐ |

| I. 4 Persönliche Belastungen sind Ärgernisse, die von geringfügigen Beschwerden zu erheblichen Problemen oder Schwierigkeiten reichen können. Sie können einige Male oder häufig auftreten. Bitte geben Sie an, inwiefern die unten aufgeführten persönlichen Belastungen in den letzten Monaten Sie bedrängt haben. | überhaupt nicht<br>0 | etwas schwer<br>1 | mäßig schwer<br>2 | äußerst schwer<br>3 |
|---|---|---|---|---|
| 1. Soziale Verpflichtungen | ☐ | ☐ | ☐ | ☐ |
| 2. Gesundheit eines Familienmitglieds | ☐ | ☐ | ☐ | ☐ |
| 3. Finanzielle Verantwortung für jemanden, der nicht bei Ihnen lebt | ☐ | ☐ | ☐ | ☐ |
| 4. Zu viele Verantwortungen | ☐ | ☐ | ☐ | ☐ |
| 5. Schwierigkeit Entscheidungen zu treffen | ☐ | ☐ | ☐ | ☐ |
| 6. Besorgnis über Unfälle | ☐ | ☐ | ☐ | ☐ |
| 7. Finanzielle Unsicherheit | ☐ | ☐ | ☐ | ☐ |
| 8. Körperliche Erkrankungen | ☐ | ☐ | ☐ | ☐ |
| 9. Freunde oder Verwandte sind zu weit entfernt | ☐ | ☐ | ☐ | ☐ |
| 10. Besorgnis über körperliche Funktionen | ☐ | ☐ | ☐ | ☐ |
| 11. Probleme mit alternden Eltern | ☐ | ☐ | ☐ | ☐ |
| 12. Probleme mit den eigenen Kindern | ☐ | ☐ | ☐ | ☐ |
| 13. Überlastet mit familiären Verantwortungen | ☐ | ☐ | ☐ | ☐ |
| 14. Probleme mit Scheidung oder Trennung | ☐ | ☐ | ☐ | ☐ |
| 15. Sorgen über innere Konflikte | ☐ | ☐ | ☐ | ☐ |
| 16. Schwierigkeiten mit Freunden | ☐ | ☐ | ☐ | ☐ |
| 17. Nicht genug Zeit für die Familie | ☐ | ☐ | ☐ | ☐ |
| 18. Zu viele berufsbedingte Reisen | ☐ | ☐ | ☐ | ☐ |

| **II. 1 Jetzt geht es darum, wie Sie sich geistig und körperlich fühlen.** | **fast nie**<br>**1** | **manchmal**<br>**2** | **oft**<br>**3** | **fast immer**<br>**4** |
|---|---|---|---|---|
| 1. Es fällt mir zunehmend schwer, mich auf etwas zu konzentrieren. | ☐ | ☐ | ☐ | ☐ |
| 2. Ich fühle mich antriebslos. | ☐ | ☐ | ☐ | ☐ |
| 3. Ich fühle mich, als hätte ich keine Willenskraft mehr. | ☐ | ☐ | ☐ | ☐ |
| 4. Es fällt mir zunehmend schwer, meine Aufmerksamkeit aufrechtzuerhalten. | ☐ | ☐ | ☐ | ☐ |
| 5. Ich finde es schwierig, mich am Ende des Arbeitstages zu entspannen. | ☐ | ☐ | ☐ | ☐ |
| 6. Ich kann wenig Interesse für andere Menschen aufbringen, wenn ich eben erst nach Hause gekommen bin. | ☐ | ☐ | ☐ | ☐ |
| 7. Mein Job bewirkt, dass ich nach einem Arbeitstag völlig erschöpft bin. | ☐ | ☐ | ☐ | ☐ |
| 8. Nach der Arbeit habe ich Schwierigkeiten, mich bei meinen Freizeitaktivitäten zu konzentrieren. | ☐ | ☐ | ☐ | ☐ |
| 9. Es kommt vor, dass ich nach einem Arbeitstag so müde bin, dass ich nicht mehr zu anderen Dingen komme. | ☐ | ☐ | ☐ | ☐ |

| **II. 1 Wie fühlen Sie sich generell?** | **nie**<br>**1** | **selten**<br>**2** | **ab und zu**<br>**3** | **häufig**<br>**4** | **sehr häufig**<br>**5** |
|---|---|---|---|---|---|
| 1. Verärgert | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Bekümmert | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Feindselig | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Gereizt | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Nervös | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Durcheinander | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Ängstlich | ☐ | ☐ | ☐ | ☐ | ☐ |

| II. 2 In diesem Abschnitt des Fragebogens geht es um Ihr Arbeitsverhalten. Bitte geben Sie an, inwieweit die folgenden Aussagen auf Sie persönlich zutreffen. | trifft überhaupt nicht zu<br>1 | 2 | 3 | 4 | 5 | 6 | trifft voll und ganz zu<br>7 |
|---|---|---|---|---|---|---|---|
| 1. Ich mache innovative Vorschläge zur Verbesserung der Qualität in meinem Arbeitsbereich. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ich informiere mich über neue Entwicklungen im Unternehmen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ich beachte Vorschriften und Arbeitsanweisungen mit größter Sorgfalt. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Ich ergreife die Initiative, um das Unternehmen vor möglichen Problemen zu bewahren. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Ich wirke bei auftretenden Meinungsverschiedenheiten ausgleichend auf Kollegen/Kolleginnen ein. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Ich bemühe mich aktiv darum, Schwierigkeiten mit Kollegen/Kolleginnen vorzubeugen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| II. 3 Wie häufig treten die nachfolgend beschriebenen Situationen bzw. Gefühle auf? | Die Situation/das Gefühl tritt<br>überhaupt nicht auf<br>1 | 2 | 3 | 4 | 5 | sehr oft auf<br>6 |
|---|---|---|---|---|---|---|
| 1. Ich fühle mich durch meine Arbeit ausgebrannt. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Der direkte Kontakt mit Menschen bei meiner Arbeit belastet mich. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Den ganzen Tag mit Menschen zu arbeiten, ist für mich wirklich anstrengend. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Ich glaube, dass ich manche Menschen bei der Arbeit so behandle, als wären sie unpersönliche „Objekte". | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Ich fühle mich durch meine Arbeit emotional erschöpft. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Im Laufe der Zeit hat mein Interesse an den Menschen in meiner Arbeitsumgebung deutlich nachgelassen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Am Ende eines Arbeitstages fühle ich mich verbraucht. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Ich habe das Gefühl, dass manche Menschen im Umfeld meiner Arbeit mir die Schuld für einige ihrer Probleme geben. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Ich fühle mich wieder müde, wenn ich morgens aufstehe und den nächsten Arbeitstag vor mir habe. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Ich befürchte, dass mich die Arbeit emotional verhärtet. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Ich habe das Gefühl, dass ich an meinem Arbeitsplatz zu hart arbeite. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Mir fällt es zunehmend schwer, mich in andere Menschen hineinzuversetzen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 13. Ich fühle mich durch meine Arbeit frustriert. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 14. Seitdem ich diese Arbeit ausübe, bin ich gefühlloser im Umgang mit anderen Menschen geworden. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| II. 3 Bitte geben Sie zu jeder Aussage an, wie häufig Sie die genannte Stimmung oder Sichtweise erleben? | fast nie | | | | | fast immer |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 1. Ich sehe mutlos in die Zukunft. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ich habe Schuldgefühle. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ich bin von mir enttäuscht. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Ich werfe mir Fehler und Schwächen vor. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Ich fühle mich gereizt und verärgert. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Ich muss mich zu jeder Tätigkeit zwingen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Ich bin müde und lustlos. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 1 In diesem Abschnitt bitten wir Sie, einige Fragen zu beantworten, die mit der Klarheit über verschiedene Aspekte Ihrer Arbeit zu tun haben. | stimmt überhaupt nicht | | | | | | stimmt vollständig |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. Ich weiß genau, welche **Arbeitsleistung** meinen unmittelbaren Vorgesetzten zufriedenstellt | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Meine Arbeit ist so, dass ich immer genau weiß, **wann** ich eine bestimmte Aufgabe während meiner Arbeit zu erledigen habe. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Meine Arbeit ist so, dass ich immer genau weiß, **wie** ich vorgehen muss, um meine Aufgaben gut zu erledigen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Mir ist klar, **was** mein unmittelbarer Vorgesetzter als zufriedenstellende **Arbeitsleistung** ansieht. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Meine Arbeit ist so, dass ich immer genau weiß, in welcher **zeitlichen Abfolge** ich meine Arbeit zu erledigen habe. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Meine Arbeit ist so, dass ich immer genau weiß, welches die beste **Vorgehensweise** bei der Erledigung meiner Aufgaben ist. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | stimmt überhaupt nicht 1 | 2 | 3 | 4 | 5 | 6 | stimmt vollständig 7 |
|---|---|---|---|---|---|---|---|
| 7. Ich weiß genau, welches **Leistungsniveau** von meinem unmittelbaren Vorgesetzten als akzeptabel betrachtet wird. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Meine Arbeit ist so, dass ich immer genau weiß, **wann** eine bestimmte Arbeitshandlung auszuführen ist. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Meine Arbeit ist so, dass ich immer genau weiß, auf welche **Art und Weise** ich meine Aufgaben zu erledigen habe. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **III. 2 In diesem Abschnitt des Fragebogens finden Sie einige Aussagen zu Einflussmöglichkeiten auf Arbeitsablauf und Arbeitsmethoden. Geben Sie bitte an, in wieweit diese Aussagen auf Ihre eigene Arbeit zutreffen.** | nein, überhaupt nicht 1 | 2 | 3 | ja, weitgehend 4 |
|---|---|---|---|---|
| 1. Meine Arbeit gibt mir die Möglichkeit, selbst zu bestimmen, wann ich mit der Erledigung einer Aufgabe beginne. | ☐ | ☐ | ☐ | ☐ |
| 2. Können Sie Ihre Arbeit auf unterschiedliche Art und Weise erledigen? | ☐ | ☐ | ☐ | ☐ |
| 3. Meine Arbeit gibt mir die Möglichkeit, selbst zu bestimmen, wann ich eine Aufgabe zum Abschluss bringe. | ☐ | ☐ | ☐ | ☐ |
| 4. Meine Arbeit ist so, dass ich selbst bestimmen kann, wie ich bei der Erledigung meiner Aufgaben vorgehe. | ☐ | ☐ | ☐ | ☐ |
| 5. Bei meiner Arbeit kann ich selbst mein Arbeitstempo bestimmen. | ☐ | ☐ | ☐ | ☐ |
| 6. Können Sie planen, wie Sie bei der Erledigung Ihrer Arbeit vorgehen? | ☐ | ☐ | ☐ | ☐ |

| III. 3 In diesem Abschnitt geht es um das Verhalten Ihres Vorgesetzten. Bitte bewerten Sie die folgenden Aussagen. | Trifft überhaupt nicht zu<br>1 | 2 | 3 | 4 | Trifft voll und ganz zu<br>5 |
|---|---|---|---|---|---|
| 1. Mein unmittelbarer Vorgesetzter investiert viel Zeit, um gute Beziehungen zu den Mitarbeitern aufzubauen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Mein unmittelbarer Vorgesetzter erzeugt ein Zusammengehörigkeitsgefühl unter den Mitarbeitern. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Mein unmittelbarer Vorgesetzter lässt sich in seinen Entscheidungen von den Ansichten der Mitarbeiter beeinflussen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Mein unmittelbarer Vorgesetzter versucht bei wichtigen Entscheidungen, einen Konsens unter den Mitarbeitern herzustellen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Mein unmittelbarer Vorgesetzter nimmt auf die außerberufliche Lebenssituation der Mitarbeiter Rücksicht. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Für meinen unmittelbaren Vorgesetzten ist die persönliche Weiterentwicklung der Mitarbeiter ein vorrangiges Ziel. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Mein unmittelbarer Vorgesetzter hält die Mitarbeiter zur Einhaltung hoher moralischer Standards an. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Mein unmittelbarer Vorgesetzter hält, was er verspricht. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Mein unmittelbarer Vorgesetzter verknüpft Alltagsangelegenheiten mit langfristigen Plänen für die Zukunft. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Mein unmittelbarer Vorgesetzter verfügt über weitreichende Kenntnisse bei der Bewältigung von Arbeitsproblemen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Mein unmittelbarer Vorgesetzter gibt mir das Gefühl, dass ich mit ihm und nicht für ihn arbeite. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Mein unmittelbarer Vorgesetzter arbeitet hart daran, andere dabei zu unterstützen, ihr Bestes zu geben. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 13. Mein unmittelbarer Vorgesetzter ermutigt die Mitarbeiter, sich an gemeinnützigen und ehrenamtlichen Aktivitäten außerhalb der Arbeit zu beteiligen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 14. Mein unmittelbarer Vorgesetzter betont die Notwendigkeit, für das gesellschaftliche Wohl einen Beitrag zu leisten. | ☐ | ☐ | ☐ | ☐ | ☐ |

| **III. 4 Nun einige Fragen zur erlebten Unterstützung von Kollegen.** | **gar nicht**<br>**1** | **wenig**<br>**2** | **ziemlich**<br>**3** | **völlig**<br>**4** |
|---|---|---|---|---|
| 1. Wie sehr können Sie sich auf Ihre Kollegen verlassen, wenn es in der Arbeit schwierig wird? | ☐ | ☐ | ☐ | ☐ |
| 2. Wie sehr sind Ihre Kollegen bereit, Ihre Probleme im Zusammenhang mit der Arbeit anzuhören? | ☐ | ☐ | ☐ | ☐ |
| 3. Wie sehr unterstützen Ihre Kollegen Sie, so dass Sie es in der Arbeit leichter haben? | ☐ | ☐ | ☐ | ☐ |

| **III. 5 Die folgenden Aussagen beschreiben das Ausmaß, in dem man davon überzeugt ist, dass die eigene Arbeit für andere Menschen nützlich ist und sie in positiver Weise beeinflusst.** | **trifft gar nicht zu**<br>**1** | **trifft eher nicht zu**<br>**2** | **teils/teils**<br>**3** | **trifft eher zu**<br>**4** | **stimmt völlig**<br>**5** |
|---|---|---|---|---|---|
| 1. Ich glaube, dass meine Arbeit das Leben anderer Menschen in positiver Weise verändert. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ich weiß genau, in welcher Weise meine Arbeit anderen Menschen nützt. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ich bin mir des positiven Einflusses meiner Arbeit auf andere Menschen sehr bewusst. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Meine Arbeit macht das Leben anderer wirklich besser. | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 6 Nun geht es um Eigenschaften und Fähigkeiten von Ihnen. Bitte geben Sie jeweils an, wie sehr die nachfolgenden Aussagen auf Sie persönlich zutreffen. | trifft gar nicht zu 1 | trifft eher nicht zu 2 | teils/teils 3 | trifft eher zu 4 | stimmt völlig 5 |
|---|---|---|---|---|---|
| 1. Ich bin gut darin, Versuchungen zu widerstehen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Es fällt mir schwer, schlechte Gewohnheiten abzulegen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ich sage häufig unüberlegte Dinge. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Andere Menschen würden mich als impulsiv beschreiben. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Andere Menschen würden sagen, dass ich eine eiserne Selbstdisziplin habe. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Vergnügen und Spaß halten mich manchmal davon ab, meine Arbeit zu tun. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. Ich verliere zu schnell die Geduld. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Ich unterbreche oft andere Menschen. | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 6 Bestimmte Formen der Arbeitsbelastung können sich in einer Reihe von Alltagsfehlern niederschlagen. Wie häufig sind Ihnen die aufgeführten Missgeschicke unterlaufen? | nie 1 | 2 | 3 | 4 | sehr häufig 5 |
|---|---|---|---|---|---|
| 1. Wie häufig ist es Ihnen bei der Arbeit passiert, dass Sie etwas gesagt haben und erst dann merkten, dass es beleidigend gewesen sein könnte? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Wie häufig haben Sie bei der Arbeit die Nerven verloren und es dann bedauert? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Wie häufig haben Sie bei der Arbeit Schwierigkeiten gehabt, sich zu entscheiden? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Wie häufig haben Sie am Tage geträumt, wenn Sie eigentlich zuhören sollten? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Wie häufig haben Sie bei der Arbeit angefangen, etwas zu tun, wurden abgelenkt und haben dann plötzlich etwas anderes getan, was Sie gar nicht beabsichtigt hatten? | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Wie häufig haben Sie das Gefühl gehabt, dass Ihnen etwas „auf der Zunge liegt", aber Sie kamen nicht darauf? | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 7 Die folgenden Aussagen beziehen sich auf das Verhältnis zwischen Ihrem Privat- und Berufsleben. | trifft überhaupt nicht zu 1 | 2 | 3 | 4 | 5 | trifft voll zu 6 |
|---|---|---|---|---|---|---|
| 1. Ich wünschte, ich könnte mein Berufs- und Privatleben besser trennen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Der Wechsel zwischen Berufs- und Privatleben fällt mir leicht. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Mein Beruf nimmt mir die Kraft für regelmäßige Hobbys. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Mein Beruf kostet mich häufig so viel Energie, dass mein Privatleben zu kurz kommt. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Nach der Arbeit kann ich mich sofort meinem Privatleben zuwenden. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 8 Im Folgenden finden Sie Aussagen darüber, wie Sie am Feierabend über Ihre Arbeit denken und inwiefern Sie sich entspannen können. | trifft gar nicht zu 1 | trifft eher nicht zu 2 | teils/teils 3 | trifft eher zu 4 | stimmt völlig 5 |
|---|---|---|---|---|---|
| 1. Am Feierabend vergesse ich die Arbeit. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Am Feierabend denke ich überhaupt nicht an meine Arbeit. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Am Feierabend gelingt es mir, mich von meiner Arbeit zu distanzieren. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Am Feierabend gewinne ich Abstand zu meinen beruflichen Anforderungen. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. Am Feierabend kann ich mich auf Dinge einlassen, bei denen ich mich entspanne. | ☐ | ☐ | ☐ | ☐ | ☐ |

| III. 9 Bitte beantworten Sie nun einige Fragen zu Ihrer Verbundenheit mit Ihrem Betrieb/ Unternehmen. | stimme überhaupt nicht zu 1 | 2 | 3 | 4 | 5 | 6 | stimme vollständig zu 7 |
|---|---|---|---|---|---|---|---|
| 1. Ich empfinde mich als „Teil der Familie" meines Unternehmens. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ich fühle mich emotional stark mit dem Unternehmen verbunden. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Dieses Unternehmen hat eine große persönliche Bedeutung für mich. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Ich empfinde ein starkes Gefühl der Zugehörigkeit zu meinem Unternehmen. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**III. 10 Im Folgenden sind zwei Möglichkeiten einander gegenübergestellt, wie man über sein Leben nachdenken kann oder wie man vorgehen kann, um zu erreichen, was einem wichtig ist. Bitte wählen Sie jeweils immer eine Alternative aus, die auf Ihre persönliche Lebensplanung und -gestaltung am zutreffendsten ist.**

| Alternative A | Alternative B |
|---|---|
| Ich konzentriere meine ganze Energie auf wenige Dinge. ☐ | Ich verteile meine Energie auf viele Dinge. ☐ |
| Ich verfolge immer nur einen Plan nach dem anderen. ☐ | Ich verfolge immer viele Pläne auf einmal. ☐ |
| Wenn ich mir überlege, was ich will, lege ich mich auf ein oder zwei wichtige Ziele fest. ☐ | Auch wenn ich mir überlege, was ich eigentlich will, lege ich mich nicht endgültig fest. ☐ |
| Wenn die Dinge nicht mehr so gut laufen wie bisher, lege ich mich auf ein oder zwei wichtige Ziele fest. ☐ | Wenn die Dinge nicht mehr so gut laufen wie bisher, versuche ich trotzdem, all meine Ziele beizubehalten. ☐ |
| Wenn ich etwas Wichtiges nicht mehr so tun kann wie bisher, suche ich nach einem neuen Ziel. ☐ | Wenn ich etwas Wichtiges nicht mehr so tun kann wie bisher, verteile ich meine Zeit und Energie auf viele andere Dinge. ☐ |

| Alternative A | Alternative B |
|---|---|
| Wenn mir etwas nicht mehr so gelingt wie früher, überlege ich ganz genau, was mir wichtig ist. ☐ | Wenn mir etwas nicht mehr so gelingt wie früher, lasse ich die Dinge erst einmal auf mich zukommen. ☐ |
| Ich probiere so lange, bis mir gelingt, was ich mir vorstelle. ☐ | Wenn mir nicht gleich gelingt, was ich mir vorstelle, probiere ich nicht mehr lange andere Möglichkeiten durch. ☐ |
| Ich setze alles daran, meine Pläne zu verwirklichen. ☐ | Ich warte lieber ab, ob sich meine Pläne nicht vielleicht von selbst verwirklichen. ☐ |
| Wenn mir an etwas sehr gelegen ist, setze ich mich voll und ganz dafür ein. ☐ | Auch wenn mir an etwas sehr gelegen ist, lasse ich mich dennoch nicht voll und ganz darauf ein. ☐ |
| Wenn die Dinge nicht mehr so gut laufen wie bisher, suche ich nach anderen Wegen, um zum Ziel zu kommen. ☐ | Wenn die Dinge nicht mehr so gut laufen wie bisher, gebe ich mich auch damit zufrieden. ☐ |
| Wenn etwas nicht mehr so gut klappt wie bisher, bitte ich andere um Rat oder Hilfe. ☐ | Wenn etwas nicht mehr so gut klappt wie bisher, verzichte ich lieber darauf, als andere um Rat oder Hilfe zu bitten. ☐ |
| Wenn mich etwas daran hindert, weiterzumachen wie bisher, dann gebe ich mir erst recht Mühe. ☐ | Wenn mich etwas daran hindert, weiterzumachen wie bisher, verzichte ich lieber darauf. ☐ |

Table 8: Psychological Variables - Table of Contents

| Variable | Section | Questions |
|---|---|---|
| Impulse control | I 1 | 1 - 5 |
| Motivational control | I 1 | 6 - 10 |
| Work pressure | I 1 | 11 - 13 |
| Concentration requirements | I 1 | 14 - 16 |
| Emotions | I 2 | 1 - 4 |
| Misconduct | I 3 | 1 - 3 |
| Tension workplace | I 3 | 4 - 6 |
| Private stress | I 4 | 1 - 18 |
| Exhaustion experience | II 1 | 1 - 4 |
| Need recovery | II 1 | 5 - 9 |
| Personal effort | II 2 | 1 - 6 |
| Emotional exhaustion | II 3 | 1 - 8 |
| Depersonalisation | II 3 | 9 - 14 |
| Depressive symptoms | II 3 | 15 - 22 (1 - 7) |
| Role clarity | III 1 | 1 - 9 |
| Monitoring leeway | III 2 | 1 - 6 |
| Integrity | III 3 | 1 - 14 |
| Social support | III 4 | 1 - 3 |
| Positive influence | III 5 | 1 - 4 |
| Self-regulation | III 6 | 1 - 8 |
| Mental control | III 6 | 9 - 14 (1 - 6) |
| Work-life-balance | III 7 | 1 - 5 |
| Ability to relax | III 8 | 1 - 5 |
| Commitment | III 9 | 1 - 4 |
| Way of life | III 10 | 1 - 12 |

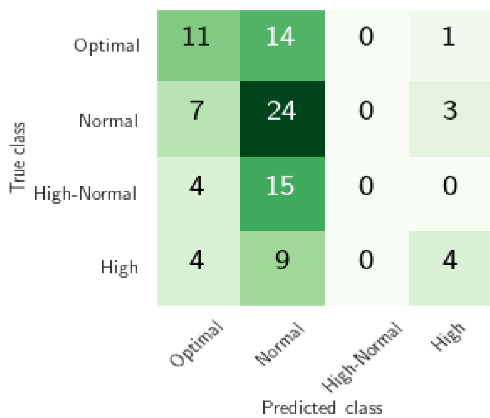
*Notes:* This table shows the sections and question numberings of the subsequent questionnaire, in which the respective psychological variables are elicited. Numbers in brackets are original item numberings where numberings are duplicated in the original questionnaire.

# B Random Forest: Results

This section shows the confusion matrices for all predicted classes (before pooling) for blood pressure, diabetes and cholesterol.

## B.1 Blood Pressure

Figure 13: Confusion Matrix Blood Pressure




## B.2 Diabetes

## B.3 Cholesterol

Figure 14: Confusion Matrix Diabetes

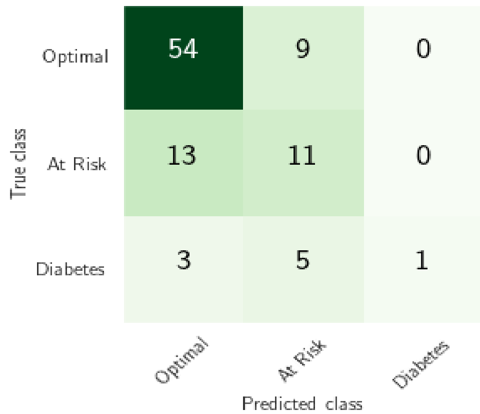
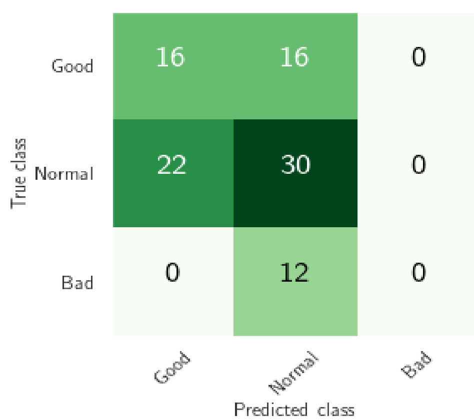



Figure 15: Confusion Matrix Cholesterol



# C Random Forest vs. Physicians

This section shows an exemplary patient record that was given to the physicians in Section C.1 as well as the original survey instructions in Section C.2.

## C.1 Random forest vs. Physicians: Patient Record

### ALLGEMEINE ANGABEN

| | |
|---|---|
| **Patienten-ID** | **11425** |
| **Geschlecht** | männlich |
| **Alter** | 61 |
| **Größe (in cm)** | 169 |
| **Gewicht (in kg)** | 61,8 |

### FAMILIENVORGESCHICHTE

| **Welche der folgenden Risikofaktoren oder Erkrankungen kamen in Ihrer Familie bei Geschwistern, Eltern und/oder Großeltern vor?** | |
|---|---|
| Bluthochdruck | Nein |
| Zuckerkrankheit (Diabetes) | Nein |
| Herzkranzgefäßverengung | Nein |
| Herzinfarkt | Nein |
| Schlaganfall | Ja |
| Dickdarmkrebs | Ja |
| Dickdarmpolypen | Nein |
| Lungenkrebs | Nein |
| Prostatakrebs | Nein |
| Brustkrebs | Nein |
| Andere Krebserkrankungen | Ja |
| Psychische Erkrankungen | Nein |
| Demenz/Alzheimer | Nein |

### EIGENE VORERKRANKUNGEN

| **Welche der folgenden Risikofaktoren oder Erkrankungen sind bei Ihnen bekannt?** |
|---|
| Bluthochdruck |
| Diabetes Typ 2 |
| Koronare Herzerkrankung |
| Vorhofflimmern |

1

## ANGABEN ZU MEDIKAMENTEN

| Nehmen Sie folgende Medikamente ein? | |
|---|---|
| Beta-Blocker | Nein |
| Lipidsenker | Nein |
| Schilddrüsenhormon | Nein |
| Sonstige Antihypertonika | Nein |
| Sonstiges | Ja |

## RAUCHENVERHALTEN

| | |
|---|---|
| Rauchen Sie? | Nein |
| Falls ja, wie viele Jahre sind Sie bereits Raucher? | 10 |
| Falls nein, wie viele Jahre sind Sie bereits Nichtraucher? | 30 |
| Falls Raucher/Exraucher wie viele Zigaretten pro Tag rauchen Sie oder haben Sie geraucht? | 10 |

## KÖRPERLICHE AKTIVITÄT

| | |
|---|---|
| Ausdauersport in Minuten pro Woche | 150 |
| Kraftsport in Minuten pro Woche | 90 |
| Wenn Sie Kraftsport betreiben, trainieren Sie dann alle [großen] Muskelgruppen (Bein-, Brust-, Rücken- und Rumpfmuskulatur)? | Ja |

## INFORMATIONEN ZUR ERNÄHRUNG & ALKOHOLKONSUM

| | |
|---|---|
| Portionen Gemüse pro Woche | 0 |
| Portionen Kohlenhydrate pro Woche | 2 |
| Portionen Obst pro Woche | 2 |
| Portionen Salat pro Woche | 0 |
| Liter (reiner) Alkohol pro Woche<br>(Hier wurde zur Grundlage genommen, dass Bier 5% Alkohol enthält, Wein 12%, Sekt 11%, und Spirituosen 37%) | 0,02 |

2

## PSYCHOMENTALE RESSOURCEN

| **Schlafgewohnheiten & subjektive Schlafqualität** | |
|---|---|
| Wann sind Sie während der letzten vier Wochen gewöhnlich morgens aufgestanden? | 6,5 |
| Wann sind Sie während der letzten vier Wochen gewöhnlich abends zu Bett gegangen? | 22,5 |
| Wie oft sind Sie während der letzten vier Wochen pro Nacht zwischendurch aufgewacht? | 3 |
| Wie lange hat es während der letzten vier Wochen gewöhnlich gedauert bis Sie nachts eingeschlafen sind? (in min) | 30 |
| Wie viele Stunden haben Sie während der letzten vier Wochen pro Nacht tatsächlich geschlafen? (Effektive Schlafzeit (Stunden) pro Nacht) | 5,5 |
| Wie würden Sie insgesamt die Qualität Ihres Schlafes während der letzten vier Wochen beurteilen? (1=sehr gut, 2=ziemlich gut, 3=ziemlich schlecht, 4=sehr schlecht) | 2 |
| Wie oft haben Sie während der letzten vier Wochen Schlafmittel eingenommen? (1=gar nicht, 2=weniger 1/Wo, 3=1-2x/Woche, 4=3x od. häufiger/Woche) | 1 |
| Wie oft hatten Sie während der letzten vier Wochen Schwierigkeiten wachzubleiben, etwa beim Autofahren, beim Essen oder bei gesellschaftlichen Anlässen? (1=gar nicht, 2=weniger 1/Wo, 3=1-2x/Woche, 4=3x od. häufiger/Woche) | 1 |
| Hatten Sie während der letzten vier Wochen Probleme, mit genügend Schwung Alltagsaufgaben zu erledigen? (1=keine Probleme, 2=kaum Probleme, 3=etwas Probleme, 4=große Probleme) | 3 |

| **Lebensgestaltung** | |
|---|---|
| Leben Sie mit Partner/in? | Ja |
| Wie viele Kinder haben Sie? | 0 |
| Wie empfinden Sie Ihr Erschöpfungserleben? *(Überbeanspruchung mentaler Ressourcen, Antriebslosigkeit, Konzentrationsschwierigkeiten)* | 0,4 |
| Wie beurteilen Sie Ihre Entspannungsfähigkeit? *(individuelle Verhaltensstrategie zur Regeneration mentaler Ressourcen)* | 0,5 |
| Wie beurteilen Sie Ihre Persönliche Lebensgestaltung? *(optimales Zusammenspiel von Selektion, Optimierung und Kompensation)* | -0,8 |
| Bedürfnis nach Erholung *(Völlige Erschöpfung nach der Arbeit und daraus resultierend fehlende Konzentration/Motivation für Freizeitgestaltung)* | 2,3 |
| Motivationskontrolle | -0,3 |
| Emotionskonflikte am Arbeitsplatz | 0,1 |
| Private Belastungen | 0,7 |

3

| Arbeitsgestaltung | |
|---|---|
| Wie viele Stunden beträgt Ihr tatsächliche Arbeitszeit pro Woche? | 29 |
| Sind Sie in einer leitenden Funktion tätig? | |
| Haben Sie in Ihrem Job Personalverantwortung? | |
| Arbeitsdruck | -0,3 |
| Commitment<br>*(Identifikation/Verbundenheit mit Arbeitsplatz)* | 1,1 |
| Integrität der Führung?<br>*(Verpflichtung zu moralischen Werteprinzipien gegenüber seinen Mitarbeitern und der Unternehmung)* | 0,3 |
| Kontrollspielräume<br>*(individuelle Beeinflussung der Arbeitsgestaltung und des zeitlichen Arbeitsablaufs)* | 0,4 |
| Rollenklarheit<br>*(aufgabenbezogene Sicherheit durch eindeutige Information der Arbeitsrolle)* | 0,7 |
| Positiver Einfluss Ihrer Arbeit<br>*(erlebte Sinnhaftigkeit der eigenen Arbeit)* | 1,2 |
| Work-Life-Balance<br>*(Betrachtung des Menschen in Berufs- und Privatleben)* | -0,3 |
| Emotionale Erschöpfung<br>*(Belastung am Arbeitsplatz durch andere Menschen; Gefühl verbraucht und ausgebrannt zu sein)* | 1,3 |
| Persönlicher Einsatz<br>*(Arbeitsverhalten, das über das vereinbarte Maß hinausgeht)* | 1,4 |
| Spannung des Arbeitsumfelds | -0,1 |
| Soziale Unterstützung durch die Kollegen<br>*(Unterstützung durch Kollegen & Vorgesetzte zur erfolgreichen Arbeitsbewältigung)* | -0,6 |
| Fehlverhalten<br>*(Beziehung zwischen Vorgesetzten und Mitarbeitern in Konflikt / Fehlsituationen)* | 1,1 |
| Mentale Kontrolle<br>*(innere Kontrolle und Ordnung über Dinge bei der Arbeit)* | 0,1 |
| Selbstregulationskompetenz<br>*(Fähigkeit, innere Widerstände zu überwinden)* | -0,9 |
| Impulskontrolle | -1,8 |
| Konzentrationsanforderungen | 0,2 |
| Depersonalisation<br>*(Abstumpfung der Persönlichkeit durch Belastung am Arbeitsplatz)* | -0,4 |
| Depressive Symptome?<br>*(z. B. Irritation, Traurigkeit, Müdigkeit, mangelnde Initiative)* | -0,2 |

| | |
|---|---|
| Wie oft waren Sie schon beim Check-Up von Preventon?<br>*(0 = Erstuntersuchung, 1 = Erste Nachfolgeuntersuchung, 2 = Zweite Nachfolgeuntersuchung, … )* | **1** |

## C.2 Random Forest vs. Physicians: Instructions

### Erklärungen zur Studie

Vielen Dank, dass Sie sich bereit erklären, an unserer Studie teilzunehmen. Alle Daten dieser Studie sind nur den Studienleitern der Goethe-Universität zugänglich und werden nicht an Dritte weitergegeben. Die Daten werden ausschließlich in anonymisierter und aggregierter Form ausgewertet. Ein Rückschluss auf Ihr persönliches Verhalten ist damit nicht möglich.

**Ablauf:** Sie haben von uns Zugriff auf zwei Ordner mit jeweils 30 „Patientenakten" als PDFs bekommen. Die Informationen in diesen Patientenakten basieren auf dem Fragebogen, der während des prevent.on Check-Ups von verschiedenen – von uns zufällig ausgewählten – Patienten/Patientinnen beantwortet wurde und fassen für jede Akte die von dem/der jeweiligen Patienten/in tatsächlich gegebenen Antworten zusammen.

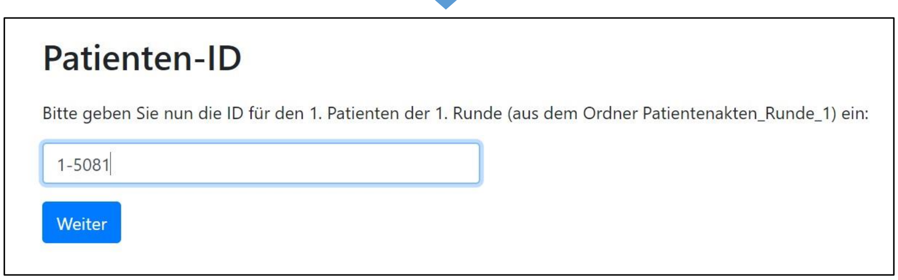
Außerdem haben Sie von uns per E-Mail einen Link zur Teilnahme an dieser Studie bekommen. Wenn Sie auf diesen Link klicken, erscheint zuerst eine Willkommensseite inkl. einer Kurzfassung des Studienablaufs. Nachdem Sie dort auf „Weiter" geklickt haben, beginnt die Studie.

**Bitte beachten Sie: Der Link ist insgesamt drei Wochen gültig.** Sie können die Beantwortung während dieser zwei Wochen jederzeit unterbrechen und zu einem späteren Zeitpunkt fortsetzen. Öffnen Sie den Link dazu einfach erneut – Sie werden automatisch zu der Stelle gebracht, an der Sie unterbrochen haben.

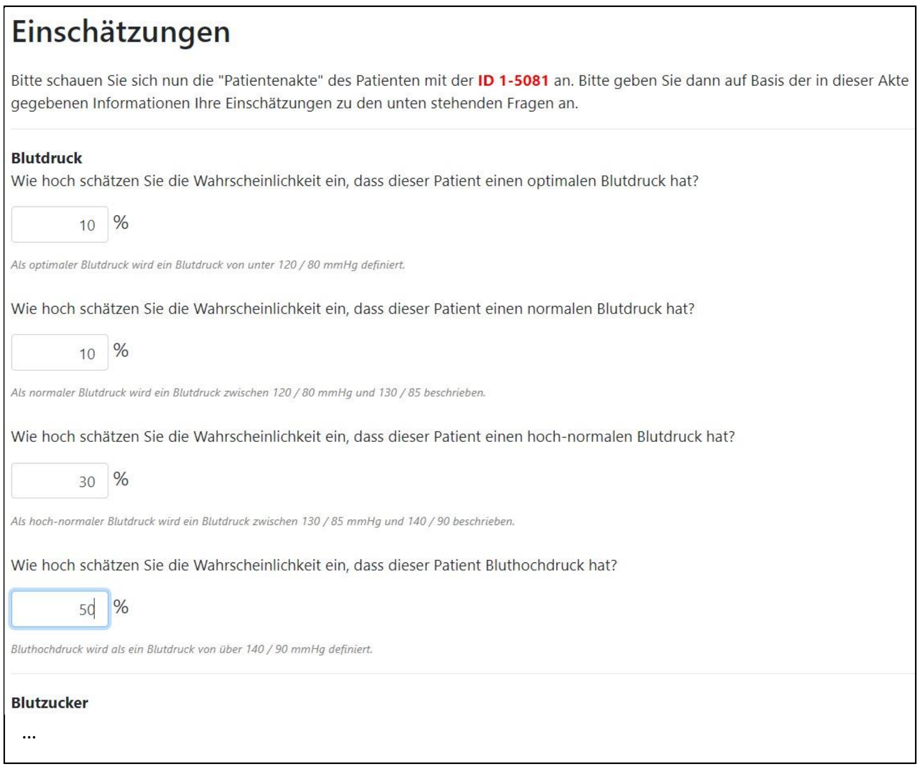
Schritt 1: Nachdem Sie den Link zur Studie geöffnet haben, nehmen Sie zunächst bitte die Patientenakten aus dem Ordner **„Patientenakten_Runde_1"** zur Hand. Auf der dann erscheinenden Seite werden Sie gebeten, die ID des Patienten einzugeben, dessen Gesundheit Sie zunächst evaluieren möchten (aus dem Ordner „Patientenakten_Runde_1"). Geben Sie dazu also bitte diese Patienten-ID ein.

ALLGEMEINE ANGABEN

| | |
|---|---|
| **Patienten-ID** | **1-5081** |
| **Geschlecht** | männlich |
| **Alter** | 43 |
| **Größe (in cm)** | 183 |
| **Gewicht (in kg)** | 113,0 |



Schritt 2: Auf der nächsten Seite finden Sie dann die Eingabemaske für Ihre Einschätzungen. Geben Sie dort bitte für den Patienten, dessen ID Sie auf der Seite zuvor eingegeben haben, Ihre Einschätzungen zu den jeweiligen Fragen ab. Nehmen Sie dazu bitte die jeweilige Patientenakte als Grundlage und versuchen Sie, die Fragen so gut wie möglich zu beantworten. Je besser Ihre Einschätzungen den tatsächlich während des Check-Ups diagnostizierten Gesundheitszuständen des Patienten entsprechen, desto höher ist Ihre Auszahlung am Ende der Studie (siehe dazu die weiteren Informationen zum Thema „Auszahlung“ unten).


Einschätzungen

Bitte schauen Sie sich nun die "Patientenakte" des Patienten mit der **ID 1-5081** an. Bitte geben Sie dann auf Basis der in dieser Akte gegebenen Informationen Ihre Einschätzungen zu den unten stehenden Fragen an.

**Blutdruck**
Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass dieser Patient einen optimalen Blutdruck hat?

10 %

*Als optimaler Blutdruck wird ein Blutdruck von unter 120 / 80 mmHg definiert.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass dieser Patient einen normalen Blutdruck hat?

10 %

*Als normaler Blutdruck wird ein Blutdruck zwischen 120 / 80 mmHg und 130 / 85 beschrieben.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass dieser Patient einen hoch-normalen Blutdruck hat?

30 %

*Als hoch-normaler Blutdruck wird ein Blutdruck zwischen 130 / 85 mmHg und 140 / 90 beschrieben.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass dieser Patient Bluthochdruck hat?

50 %

*Bluthochdruck wird als ein Blutdruck von über 140 / 90 mmHg definiert.*

**Blutzucker**

...

Schritt 3: Nachdem Sie Ihre Einschätzungen für den ersten Patienten abgegeben haben, klicken Sie bitte auf „Weiter“. Dann wiederholen Sie bitte Schritt 1 und Schritt 2 für alle Patientenakten aus dem Ordner „Patientenakten_Runde_1“.

Schritt 4: Nachdem Sie für alle 30 Patientenakten aus dem Ordner „Patientenakten_Runde_1“ Ihre Einschätzungen abgegeben haben, werden Sie gebeten, für jeden der sieben verschiedenen Gesundheitsoutcomes anzugeben, welche Faktoren / Fragen für Ihre Einschätzungen am Wichtigsten waren. Geben Sie hierzu bitte bis zu 10 der für Sie wichtigsten Faktoren an. Wenn Sie weniger als 10 Faktoren für Ihre Einschätzungen benötigt haben, geben Sie bitte nur so viele an, wie für Sie relevant waren, mindestens jedoch die Top 3. Klicken Sie auf „Weiter“ sobald Sie Ihre Eingaben gemacht haben.

## Blutdruck

Welche Faktoren haben ihre Einschätzung bzgl. **des Blutdrucks** am meisten beeinflusst?

*Geben Sie bitte mindestens 3 bis maximal 10 Faktoren an.*

- ☐ Geschlecht
- ☐ Alter
- ☐ Größe (in cm)
- ☐ Gewicht (in kg)
- ☐ Bluthochdruck in Familie
- ☐ Diabetes in Familie
- ☐ Herzkranzgefäßverengung in Familie
- ☐ Herzinfarkt in Familie
- ☐ Schlaganfall in Familie
- ☐ Dickdarmkrebs in Familie
- ☐ Dickdarmpolypen in Familie
- ☐ Lungenkrebs in Familie

...

Schritt 5: Nun beginnt die zweite Runde. In dieser werden Sie wieder gebeten, die ID eines Patienten einzugeben, dessen Gesundheit Sie evaluieren möchten. Nehmen Sie dazu nun bitte die Patientenakten aus dem Ordner **„Patientenakten_Runde_2"** zur Hand und geben Sie bitte die ID eines Patienten dieses Ordners ein.

ALLGEMEINE ANGABEN

| | |
|---|---|
| **Patienten-ID** | **2-3056** |
| **Geschlecht** | weiblich |
| **Alter** | 44 |
| **Größe (in cm)** | 173 |
| **Gewicht (in kg)** | 67,0 |

## Patienten-ID

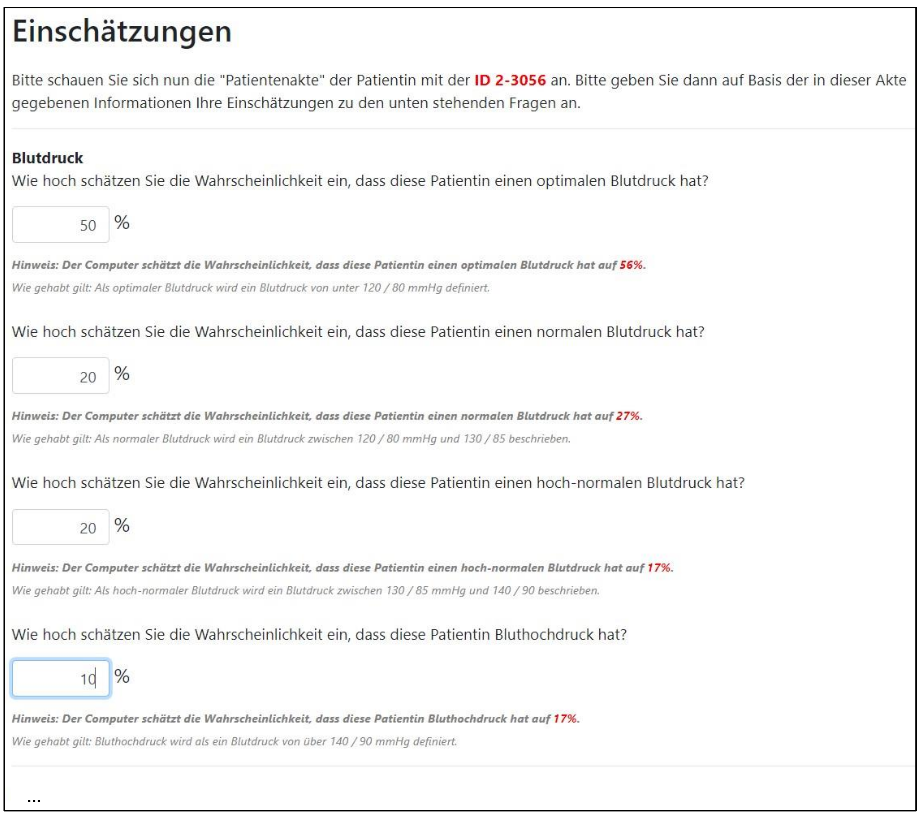
Bitte geben Sie nun die ID für den 1. Patienten der zweiten Runde (aus dem Ordner Patientenakten_Runde_2) ein:

Patienten ID:

2-3056

Weiter

Schritt 6: Auf der nächsten Seite finden Sie dann wieder die Eingabemaske für Ihre Einschätzungen. Allerdings werden diesmal für jede Frage die Antworten des Computers angezeigt, der auf Basis einer statistischen Analyse aller Patientenakten zuvor ebenfalls die Gesundheitsoutcomes geschätzt hat. Ansonsten gibt es keine Änderung im Vergleich zur Runde 1. Geben Sie daher also bitte hier wieder für den Patienten, dessen ID Sie auf der Seite zuvor eingegeben haben, Ihre Einschätzungen zu den jeweiligen Fragen ab. Nehmen Sie dazu die jeweilige Patientenakte als Grundlage und versuchen Sie, die Fragen so gut wie möglich zu beantworten. Je besser Ihre Einschätzungen den tatsächlich während des Check-Ups diagnostizierten Gesundheitszuständen des Patienten entsprechen, desto höher ist Ihre Auszahlung am Ende der Studie (s.u.).


## Einschätzungen

Bitte schauen Sie sich nun die "Patientenakte" der Patientin mit der **ID 2-3056** an. Bitte geben Sie dann auf Basis der in dieser Akte gegebenen Informationen Ihre Einschätzungen zu den unten stehenden Fragen an.

**Blutdruck**

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass diese Patientin einen optimalen Blutdruck hat?

50 %

***Hinweis: Der Computer schätzt die Wahrscheinlichkeit, dass diese Patientin einen optimalen Blutdruck hat auf 56%.***

*Wie gehabt gilt: Als optimaler Blutdruck wird ein Blutdruck von unter 120 / 80 mmHg definiert.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass diese Patientin einen normalen Blutdruck hat?

20 %

***Hinweis: Der Computer schätzt die Wahrscheinlichkeit, dass diese Patientin einen normalen Blutdruck hat auf 27%.***

*Wie gehabt gilt: Als normaler Blutdruck wird ein Blutdruck zwischen 120 / 80 mmHg und 130 / 85 beschrieben.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass diese Patientin einen hoch-normalen Blutdruck hat?

20 %

***Hinweis: Der Computer schätzt die Wahrscheinlichkeit, dass diese Patientin einen hoch-normalen Blutdruck hat auf 17%.***

*Wie gehabt gilt: Als hoch-normaler Blutdruck wird ein Blutdruck zwischen 130 / 85 mmHg und 140 / 90 beschrieben.*

Wie hoch schätzen Sie die Wahrscheinlichkeit ein, dass diese Patientin Bluthochdruck hat?

10 %

***Hinweis: Der Computer schätzt die Wahrscheinlichkeit, dass diese Patientin Bluthochdruck hat auf 17%.***

*Wie gehabt gilt: Bluthochdruck wird als ein Blutdruck von über 140 / 90 mmHg definiert.*

...

Schritt 7: Nachdem Sie Ihre Einschätzungen für den ersten Patienten des Ordners „**Patientenakten_Runde_2“** abgegeben haben, klicken Sie bitte auf „Weiter“. Dann wiederholen Sie bitte Schritt 5 und Schritt 6 für alle Patientenakten aus dem Ordner „Patientenakten_Runde_2“.

Auszahlung: Je besser Ihre Einschätzungen sind, desto mehr Geld können Sie in dieser Studie verdienen. Für die Auszahlung werden am Ende 10 der insgesamt 60 von Ihnen bewerteten Patientenakten zufällig ausgewählt. Für jede dieser 10 Patientenakten gelten folgende Auszahlungsregeln.

Pro Patientenakte werden Sie gebeten Wahrscheinlichkeiten für insgesamt sieben Gesundheitsdimensionen anzugeben: Blutdruck, Blutzucker, Cholesterin, KHK, Plaques, Metabolisches Syndrom, Fitness (siehe dazu die genauen Definitionen der Gesundheitsdimensionen unten). Manche dieser Dimensionen bestehen aus zwei Kategorien (z.B. Plaques – nein, ja), andere aus drei Kategorien (z.B. Blutzucker – optimal, beginnende Störung, Diabetes) oder auch aus vier Kategorien (z.B. Blutdruck – optimal, normal, hoch-normal, hoch).

Grundsätzlich gilt: Die Kategorie, der Sie jeweils die höchste Wahrscheinlichkeit zuordnen, wird am Ende mit dem **tatsächlichen Ergebnis** verglichen, welches der Patient während des prevent.on Gesundheits-Check-Ups erhalten hat.

Im Fall von zwei Kategorien bekommen Sie 50 Cent, falls Ihre Vorhersage mit dem tatsächlichen Ergebnis übereinstimmt, andernfalls bekommen Sie 0 Cent. Im Fall von drei oder vier Kategorien bekommen Sie 50 Cent, falls Ihre Vorhersage mit dem tatsächlichen Ergebnis übereinstimmt. Liegen Sie nur eine Kategorie daneben, bekommen Sie 25 Cent. Liegen Sie zwei oder drei Kategorien daneben, bekommen Sie 0 Cent.

**Beispiele:**

1) Angenommen Sie geben folgende Wahrscheinlichkeiten bei den Fragen nach Blutdruck an:

- Optimaler Blutdruck: **10%**
- Normaler Blutdruck: **10%**
- Hoch-normaler Blutdruck: **30%**
- Bluthochdruck**: 50%**

Das bedeutet, dass Sie „Bluthochdruck“ die höchste Wahrscheinlichkeit zugeordnet haben. Wenn der Patient nun tatsächlich Bluthochdruck hat, bekommen Sie 50 Cent für diese Einschätzung. Wenn der Patient in Wirklichkeit einen hoch-normalen Blutdruck hat (d.h. Sie liegen eine Kategorie daneben), bekommen Sie noch 25 Cent für diese Einschätzung. Wenn der Patient in Wirklichkeit einen normalen oder optimalen Blutdruck hat, bekommen Sie 0 Cent für diese Einschätzung.

2) Angenommen Sie geben folgende Wahrscheinlichkeit bei der Frage nach dem metabolischen Syndrom an:

- Wahrscheinlichkeit für ein metabolisches Syndrom**: 70%**

Das bedeutet, dass Sie mit 70%iger Wahrscheinlichkeit einschätzen, dass der Patient unter einem metabolischen Syndrom leidet. Wenn der Patient nun tatsächlich unter einem metabolischen Syndrom leidet, bekommen Sie 50 Cent für diese Einschätzung. Wenn der Patient in Wirklichkeit nicht unter einem metabolischen Syndrom leidet, bekommen Sie 0 Cent für diese Einschätzung.

Pro Patientenakte können Sie also bis zu 3,50 Euro verdienen (wenn alle sieben Einschätzungen richtig sind). Bei insgesamt 10 zufällig ausgewählten Patientenakten bedeutet dies, dass Sie insgesamt 35 Euro verdienen können. Zu diesem Verdienst erhalten Sie unabhängig von Ihren Einschätzungen eine Teilnahmeprämie von 25 Euro. Damit ist eine Gesamtprämie von bis zu 60 Euro in dieser Studie möglich.

Für die Auszahlung Ihrer Prämie werden wir Sie nach Durchführung der Studie per E-Mail kontaktieren.

Definitionen der Gesundheitsdimensionen:

**Blutdruck (4 Kategorien: optimal, normal, hoch-normal, hoch)**

Als optimaler Blutdruck wird ein Blutdruck von unter 120 / 80 mmHg definiert.

Als normaler Blutdruck wird ein Blutdruck von mindestens 120 / 80 mmHg aber unter 130 / 85 definiert.

Als hoch-normaler Blutdruck wird ein Blutdruck von mindestens 130 / 85 mmHg aber unter 140 / 90 definiert.

Bluthochdruck wird als ein Blutdruck von mindestens 140 / 90 mmHg definiert.

**Blutzucker (3 Kategorien: nein, beginnende Störung, Diabetes)**

Ein Patient hat optimale Blutzuckerwerte, wenn alle drei der folgenden Kriterien zutreffen:

- Nüchternblutzucker ≤ 99 mg/dl
- HBA1c ≤ 5,8
- HOMA-Index ≤ 2,5

Eine beginnende Störung liegt vor, wenn mindestens eins der drei folgenden Kriterien zutrifft:

- Nüchternblutzucker zwischen 100 mg/dl und 126 mg/dl
- HBA1c zwischen 5,9 und 6,0
- HOMA-Index > 2,5

Diabetes wird diagnostiziert, wenn mindestens eines der beiden folgenden Kriterien zutrifft:

- Nüchternblutzucker > 126 mg/dl
- HBA1c > 6,0

**Cholesterin (3 Kategorien: gut, normal, schlecht)**

- Hier werden Sie nach dem HDL-Cholesterinwert gefragt.
- Als „Gut" wird klassifiziert: Weiblich: HDL-Cholesterin über 70 mg/dl<br>Männlich: HDL-Cholesterin über 60 mg/dl
- Als „Normal" wird klassifiziert: Weiblich: HDL-Cholesterin zwischen 50 mg / dl und 70 mg/dl<br>Männlich: HDL-Cholesterin zwischen 40 mg / dl und 60 mg/dl
- Als „Schlecht" wird klassifiziert: Weiblich: HDL-Cholesterin unter 50 mg/dl Männlich: HDL-Cholesterin unter 40 mg/dl

**Koronare Herzerkrankung (2 Kategorien: niedrig = kleiner als 30%, hoch = mindestens 30%)**

Das Risiko für eine koronare Herzerkrankung wird durch den "Lifetime-Risk-Rechner" berechnet.

**Plaques (2 Kategorien: nein, ja)**

Als Plaques wurde eine maximale Höhe der Ablagerungen in der Halsschlagader von mehr als 1,5 mm definiert.

**Metabolisches Syndrom (2 Kategorien: nein, ja)**

Ein Metabolisches Syndrom liegt vor, wenn mindestens drei Kriterien der fünf folgenden Kriterien zutreffen:

- Blutdruck > 130/85 mmHg
- Nüchternblutzucker > 100 mg/dl
- Triglyceride > 150 mg/dl
- Bauchumfang: weiblich > 80 cm; männlich > 94 cm
- HDL Cholesterin: weiblich < 50 mg/dl; männlich < 40 mg/dl

**Fitness (2 Kategorien: schlecht = unterhalb der 40. Perzentile der Cooper-Daten, gut = über der 40. Perzentile der Cooper-Daten)**

Hier werden Sie nach dem Anteil an Patienten der gleichen Altersgruppe gefragt, die geringere Werte bei der maximalen Sauerstoffaufnahme während einer Spiroergometrie bei prevent.on erzielt haben.

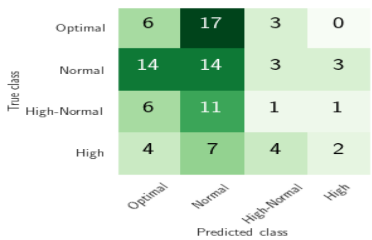
## C.3 Random Forest vs. Physicians: Prediction Performance

This section shows the ROC-curves and confusion matrices for the physician and random forest predictions for each health outcome and round.

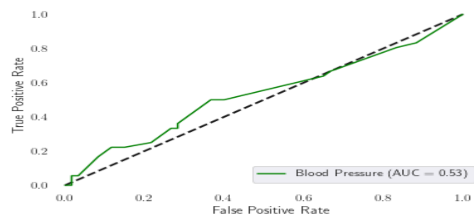
### C.3.1 Blood Pressure

Figure 16: Blood Pressure - Confusion Matrices and ROC-Curves

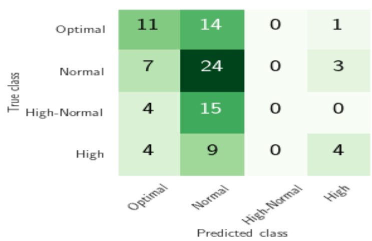
(a) CM Physicians - Round 1



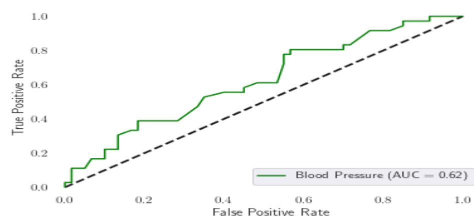

(b) ROC-Curve Physicians - Round 1



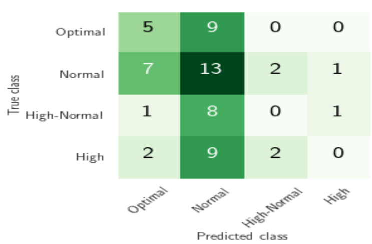

(c) CM Random Forest - Round 1



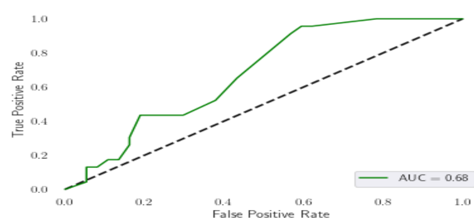

(d) ROC-Curve Random Forest - Round 1



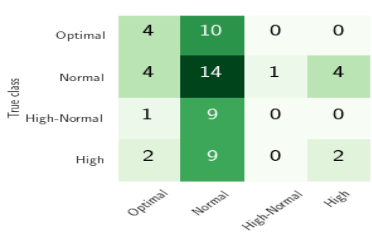

(e) CM Physicians - Round 2



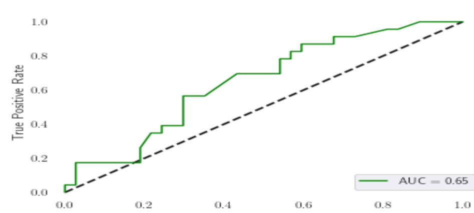

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

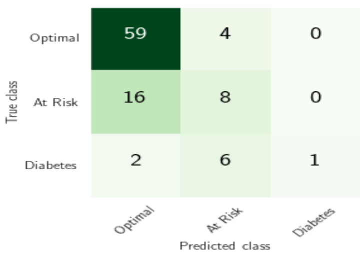



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves after pooling the classes “Optimal” and “Normal” as well as “High-Normal” and “High”, respectively.

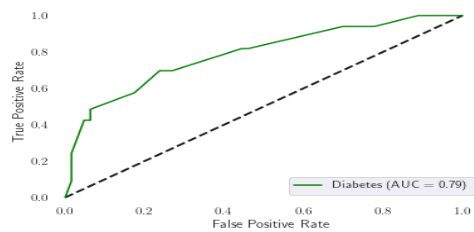
### C.3.2 Diabetes

Figure 17: Diabetes - Confusion Matrices and ROC-Curves

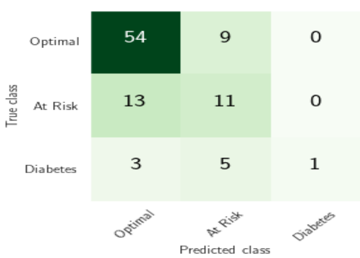
(a) CM Physicians - Round 1



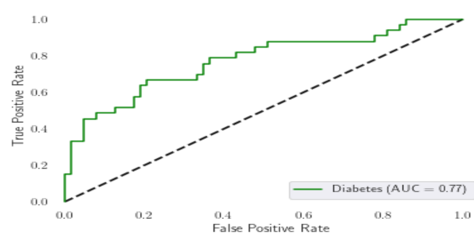

(b) ROC-Curve Physicians - Round 1



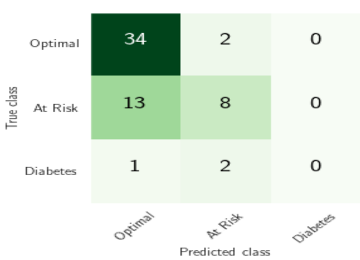

(c) CM Random Forest - Round 1



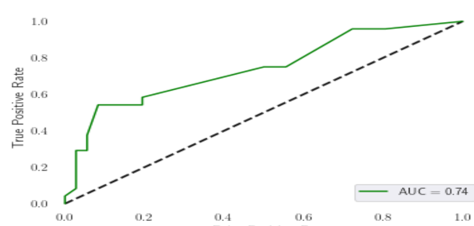

(d) ROC-Curve Random Forest - Round 1



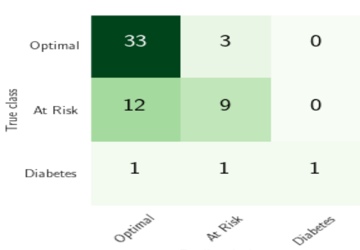

(e) CM Physicians - Round 2



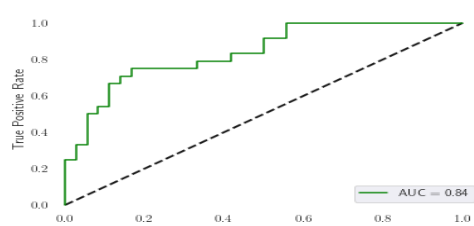

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

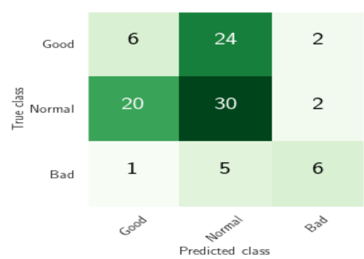



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves after pooling the classes "At Risk" and "Diabetes"

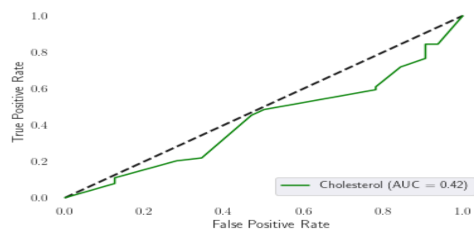
### C.3.3 Cholesterol

Figure 18: Cholesterol - Confusion Matrices and ROC-Curves

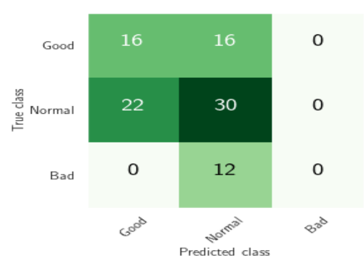
(a) CM Physicians - Round 1



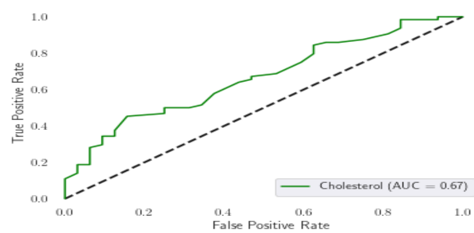

(b) ROC-Curve Physicians - Round 1



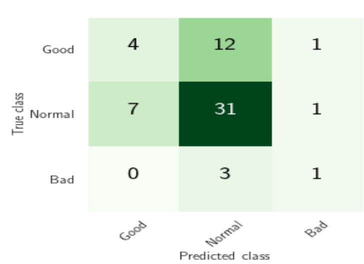

(c) CM Random Forest - Round 1



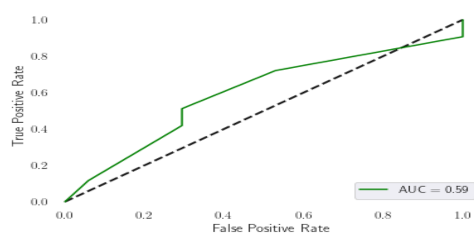

(d) ROC-Curve Random Forest - Round 1



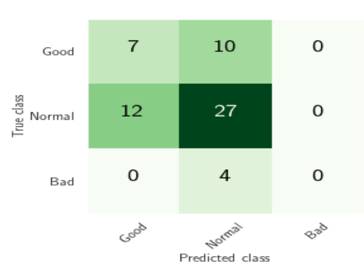

(e) CM Physicians - Round 2



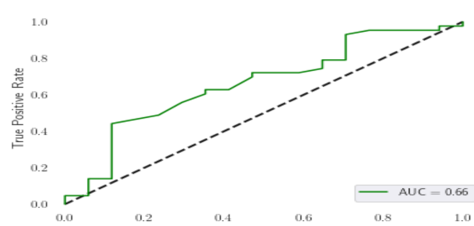

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

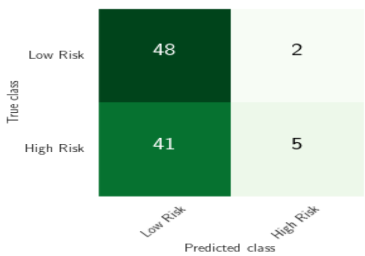



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves after pooling the classes "Normal" and "Bad".

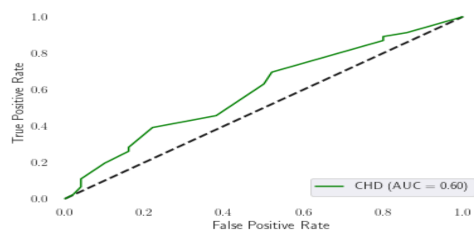
### C.3.4 Coronary Heart Disease

Figure 19: Coronary Heart Disease - Confusion Matrices and ROC-Curves

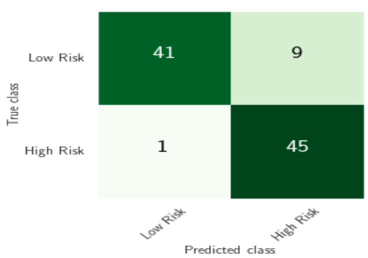
(a) CM Physicians - Round 1



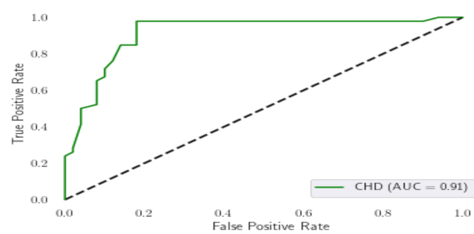

(b) ROC-Curve Physicians - Round 1



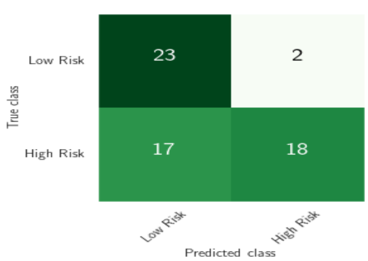

(c) CM Random Forest - Round 1



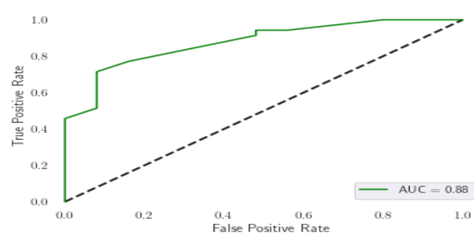

(d) ROC-Curve Random Forest - Round 1



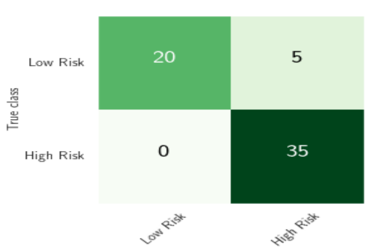

(e) CM Physicians - Round 2



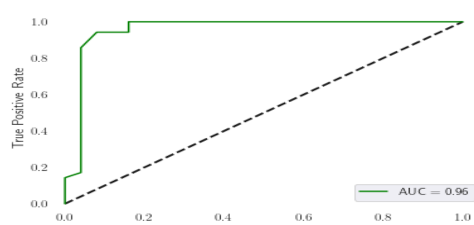

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

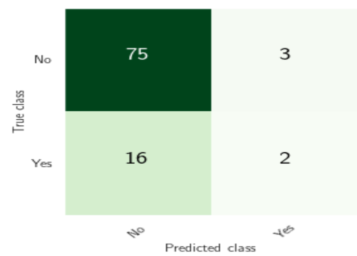



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves.

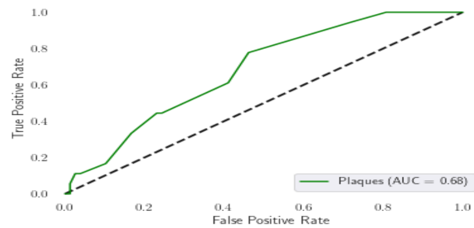
### C.3.5 Plaques

Figure 20: Plaques - Confusion Matrices and ROC-Curves

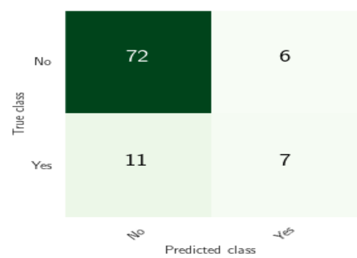
(a) CM Physicians - Round 1



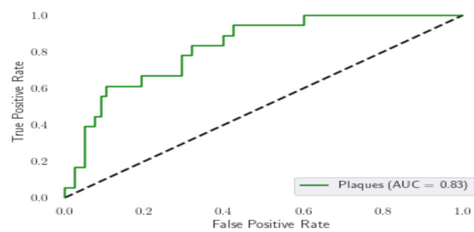

(b) ROC-Curve Physicians - Round 1



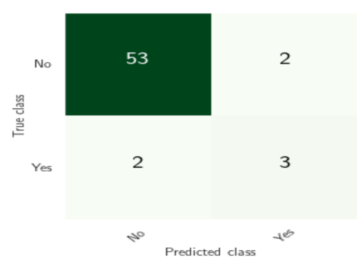

(c) CM Random Forest - Round 1



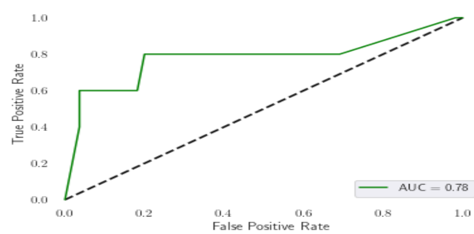

(d) ROC-Curve Random Forest - Round 1



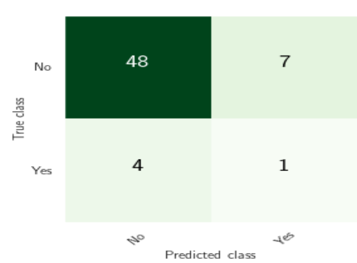

(e) CM Physicians - Round 2



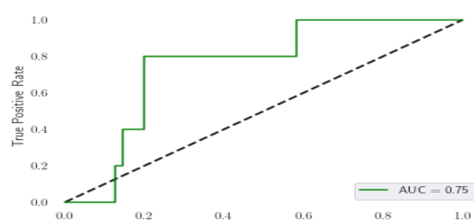

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

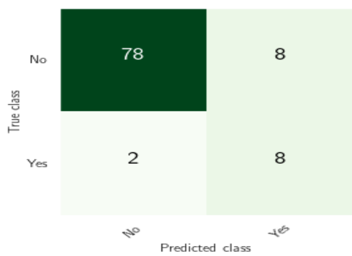



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves.

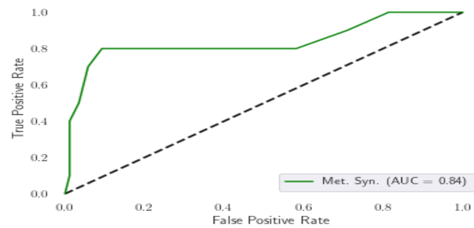
### C.3.6 Metabolic Syndrome

Figure 21: Metabolic Syndrome - Confusion Matrices and ROC-Curves

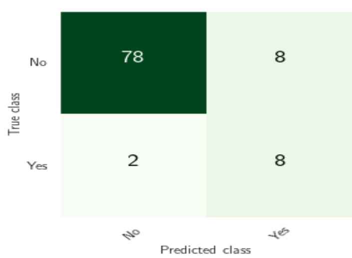
(a) CM Physicians - Round 1



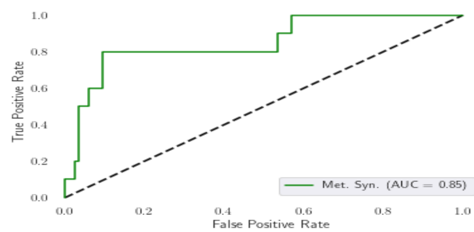

(b) ROC-Curve Physicians - Round 1



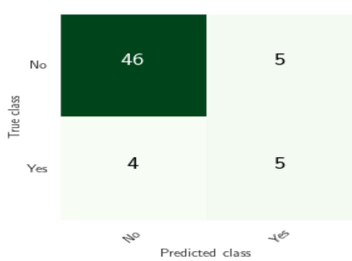

(c) CM Random Forest - Round 1



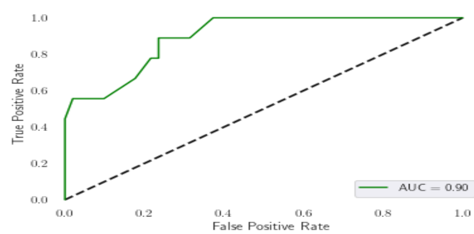

(d) ROC-Curve Random Forest - Round 1



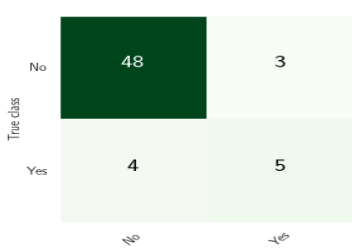

(e) CM Physicians - Round 2



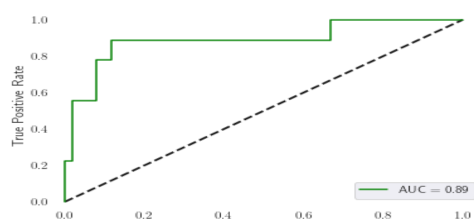

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

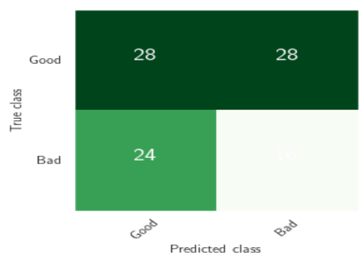



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves.

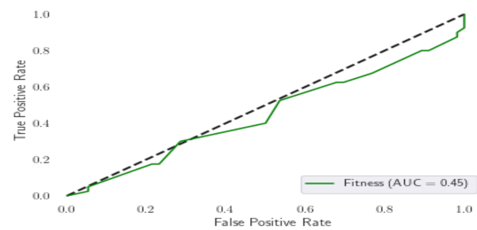
### C.3.7 Fitness

Figure 22: Fitness - Confusion Matrices and ROC-Curves

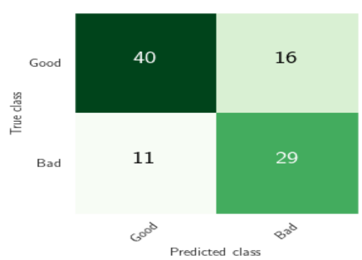
(a) CM Physicians - Round 1



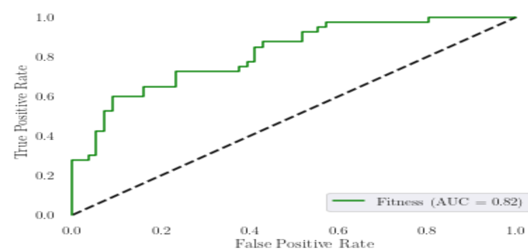

(b) ROC-Curve Physicians - Round 1



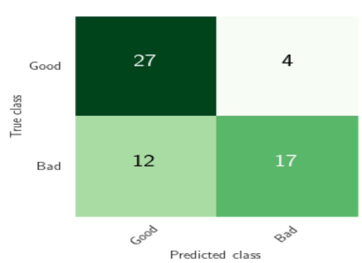

(c) CM Random Forest - Round 1



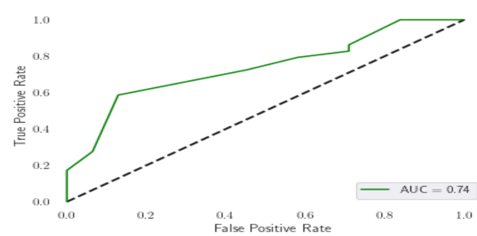

(d) ROC-Curve Random Forest - Round 1



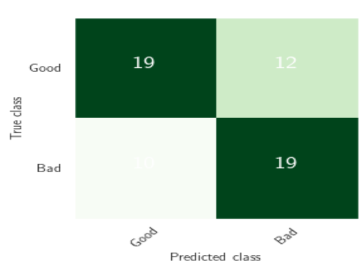

(e) CM Physicians - Round 2



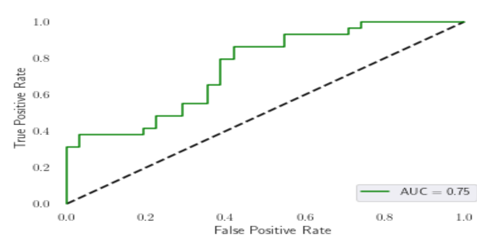

(f) ROC-Curve Physicians - Round 2




(g) CM Random Forest - Round 2




(h) ROC-Curve Random Forest - Round 2

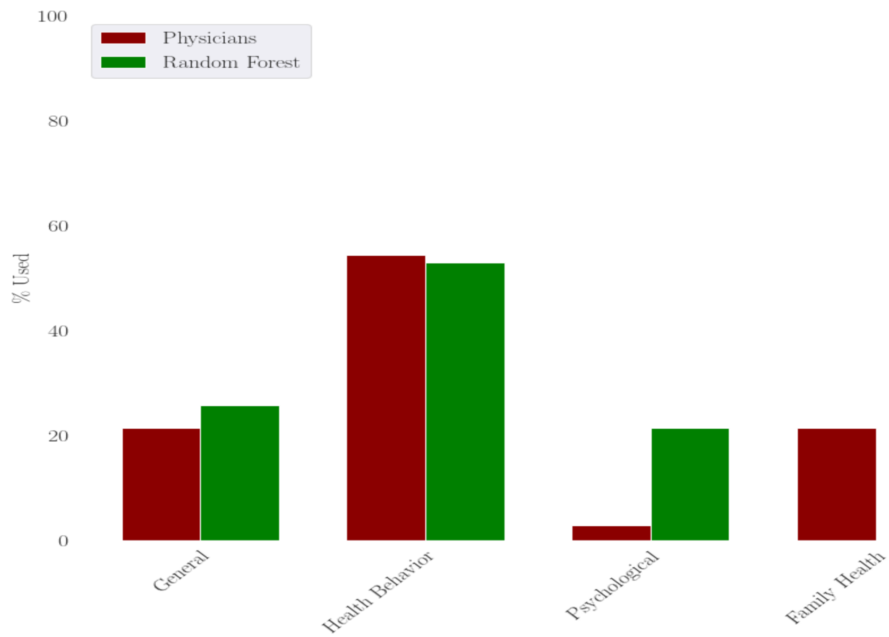



*Notes:* The left panel shows the confusion matrices (CM) for physicians (a) and (e) and random forest (c) and (g) for rounds 1 (upper half) and 2 (lower half). The right panel shows the respective ROC-curves.

## C.4 Random Forest vs. Physicians: Variable Importance

This section shows the differences between physicians and the random forest with respect to the kinds of variables they used for their predictions as well as their respective top 10 variables for each of the seven health outcomes.

Figure 23: Variables per Category




*Notes:* This figure shows the percentage use of each variable category for physicians (red) and random forest (green), based on the top 10 variables of each of the seven outcomes. Each variable was only counted once (in case it was considered important in more than one outcome).

Table 9: Variable Importance

| | Physicians | Random forest |
|---|---|---|
| Blood pressure | weight, gender, age, hypertension (family), smoke, cardio, ability to relax, height, strength, vegetables | weight, alcohol, height, age, cardio, integrity, role clarity, working hours, private stress, commitment |
| Diabetes | weight, gender, age, height, diabetes (family), cardio, Strength, carbs, fruits, smoke | weight, age, cardio, height, alcohol, integrity, role clarity, working hours, personal effort, self-regulation |
| Cholesterol | gender, weight, age, height, cardio, strength, vegetables, heart attack (family), stroke (family), smoke, | weight, alcohol, cardio, height, age, integrity, role clear, working hours, private stress, self-regulation |
| CHD | age, weight, coronary vessels (family), smoke, gender, height, heart attack (family), stroke (family), cardio, strength | gender, height, weight, working hours, alcohol, age, leadership, cardio, role clarity, private stress |
| Plaques | age, coronary vessels (family), smoke, gender, weight, stroke (family), smoke duration, height, hypertension (family), diabetes (family), | weight, height, cardio, age, working hours, smoke amount, alcohol, strength, gender, commitment |
| Metabolic syn. | weight, cardio, strength, gender, age, height, diabetes (family), hypertension (family), coronary vessels (family), smoke, | age, weight, alcohol, smoke duration, height, gender, smoke amount, working hours, smoke, not smoke duration |
| Fitness | weight, cardio, strength, gender, age, height, smoke, strength detail, motivational control, leadership | weight, cardio, height, strength, alcohol, smoke duration, smoke amount, strength detail, role clarity, private stress |

*Notes:* The table shows for each outcome, the pooled top ten most important variables for physicians and the ten most important variables for the random forest, respectively.