Aus der Radiologischen Universitätsklinik Tübingen

Abteilung für Diagnostische und Interventionelle Radiologie

Evaluation of MRI-derived radiomics features of hepatic fat as biomarkers of type 2 diabetes mellitus and the metabolic syndrome

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Medizin

der Medizinischen Fakultät

der Eberhard Karls Universität

zu Tübingen

vorgelegt von

Gutmann, Daniel Andreas Philipp

2021

Dekan:	Professor Dr. B. Pichler
1. Berichterstatter:	Professor Dr. F. Bamberg
2. Berichterstatter:	Professorin Dr. D. Thorwarth
Tag der Disputation:	11.08.2021

Gewidmet:

Marianne Gutmann-Dotzauer,

Gerold E. A. Gutmann,

Dawson

&

Rakia

Table of contents

List of Tables 1
List of figures 1
1 Abbreviations 2
2 Introduction 5
2.1 Metabolic syndrome 5
2.1.1 Origins of the metabolic syndrome 5
2.1.2 Definition, epidemiology, economic and clinical aspects 6
2.1.2.1 Diagnostic Definitions of the metabolic syndrome 6
2.1.2.2 Epidemiology and risk of complications 6
2.1.2.3 Economic impact of metabolic syndrome 10
2.1.2.4 The components of the metabolic syndrome 11
2.1.2.4.1 Impaired glucose metabolism 11
2.1.2.4.2 Impaired lipid homeostasis 13
2.1.2.4.3 Obesity 14
2.1.2.4.4 Hypertension 15
2.2 The liver and the metabolic syndrome 16
2.3 MR for assessment of hepatic fat 18
2.4 Radiomics: extracting more from plain sight 20
2.5 Aim and scientific hypothesis 21
3 Materials and Methods 23
3.1 Study population 23
3.2 Covariate collection 25
3.3 MRI acquisition parameters 25
3.4 Segmentation of volumetric liver masks 26
3.5 Image and radiomics data processing 26
3.5.1 Generation liver volumes of interest 26
3.5.2 Computation of rfwc maps 27
3.5.3 Artificial test-retest and inter-rater scenarios 27
3.5.4 Extraction of radiomics features 28
3.5.5 Selection of stable radiomics features 28
3.5.6 Selection of stable features for predictive models 30
3.5.7 Training of benchmark RF models 31

3.5.8 Training and validation statistics of random forest models 31
3.6 Statistical analysis 32
4 Results 33
4.1 General Results 33
4.2 Quality management 33
4.3 Epidemiological characteristic of study sample 36
4.4 Analysis of selected radiomics features 37
4.5 Predictive performance of radiomics features and benchmark parameters 43
4.6 Stability of radiomics RF model features in the validation set 45
5 Discussion 51
6 Summary 58
7 Bibliography 61
8 Erklärung zum Eigenanteil 74
9 List of publications 76
10 Acknowledgments 77

List of Tables

Table 1	Diagnostic definitions of the metabolic syndrome	8
Table 2	Description of radiomics features	29
Table 3	Summary of the correlation and Bland-Altman analysis	36
Table 4	Summary of epidemiological characteristics	38
Table 5	Radiomics features stability selection	39
Table 6	Radiomics features regression analysis	46
Table 7	Radiomics features of predictive random forest models	47
Table 8	Random forest model performance for T2DM and MetS	48

List of figures

Figure 1	Structure of the KORA study and participant selection	24
Figure 2	Hepatic fat fraction quantified by T1-DED compared to T1-MED and MRS	34
Figure 3	Correlation and Band-Altman analysis of hepatic fat fraction quantified by T1-DED compared to T1-MED and MRS	35
Figure 4	Feature stability associations in training and validation sets	40
Figure 5	Association of radiomics feature coefficient of variation in validation and training sets	41
Figure 6	Association of the coefficient of variation with feature stability	42
Figure 7	Box-plots of feature stability in the validation set	43
Figure 8	Feature selection cross-validation scores	44
Figure 9	Random forest receiver operating characteristic curves	49
Figure 10	Pair plots of radiomic feature stability of in training and validation set.	50

1 Abbreviations

95 % CI: 95 % confidence interval

95 % CI – lb: lower bound of the 95 % confidence interval

$Accuracy_B$: balanced accuracy

AGE: advanced glycosylation end product

au: arbitrary units

AUROC: area under the curve of the receiver operating characteristic

BMI: body mass index

CoV: coefficient of variation

CV: cross-validation

CVD: cardiovascular disease

DNL: de-novo lipid synthesis

FFA: free fatty acids

FPG: fasting plasma glucose

GLCM: gray level co-occurrence matrix

gldbw: gray level discretization bin width

GLDM: gray level dependence matrix

GLRLM: gray level run length matrix

GLSZM: gray level size zone matrix

HOMA-IR: homeostasis model assessment of insulin resistance

HCC: hepatocellular carcinoma

HDL-C: high-density lipoprotein cholesterol

ICC: intraclass correlation coefficient

ICC(1,1): test-retest reliability

ICC(3,k): inter-rater reliability

IDF: international diabetes federation

KORA: Cooperative Health Research in the Region of Augsburg (Kooperative Gesundheitsforschung in der Region Augsburg)

LDL-C: low-density lipoprotein cholesterol

MetS: metabolic syndrome

MR: magnetic resonance

MRI: magnetic resonance imaging

MRS: ^{1}H-magnetic resonance spectroscopy

NAFLD: non-alcoholic fatty liver disease

NASH: non-alcoholic steatohepatitis

NCEP-ATPIII: National Cholesterol Education Program Adult treatment Panel III

NGTDM: neighboring gray tone difference matrix

NO: nitric oxide

OGTT: oral glucose tolerance test

OR: odds ratio

PDFF: proton density fat fraction

RF: random forest

RF_{RAD}: radiomics random forest

rfwc: relative fat water content

$rfwc_N$: noise augmented relative fat water content

SAT: subcutaneous adipose tissue

SD: standard deviation

S_{fat}: fat-signal

S_{IP}: in-phase-signal

SMOTE: synthetic minority over-sampling technique

S_{OP}: out-of-phase-signal

S_{Water}: water-signal

T: Tesla

T1-DED: T1-weighted dual-echo Dixon

T1-MED: T1-weighted multi-echo Dixon

T2-HASTE: T2-weighted half-Fourier singleshot turbo spin-echo

T2DM: type 2 diabetes mellitus

TE: echo time

TR: repetition time

VAT: visceral adipose tissue

VIBE: volume interpolated breath hold

VLDL: very low density lipoprotein

VOI: volume of interest

VOI_D: deformed volume of interest

WC: waist circumference

WHO: world health organization

2 Introduction

2.1 Metabolic syndrome

2.1.1 Origins of the metabolic syndrome

Over the course of the last century the world has witnessed a constant and relentless global trend towards sedentary lifestyles and surplus dietary intake. Once limited to industrialized countries, such lifestyles have spread to the developing world and are now near ubiquitous (Saklayen, 2018). Our metabolism evolved eons ago in an age of caloric scarcity and it thus poorly adapted to current times. Beyond the ongoing global rise in obesity (NCP Risk Factor Collaboration, 2017), an unforeseen consequence of these shifting life-styles, which took its origin in western industrial societies, has been an ever rising proportion of individuals suffering arterial hypertension, elevated triglycerides, low high-density lipoprotein cholesterol and impaired glucose metabolism (Kylin, 1923, Vague, 1947, Haller H. and M., 1975). These conditions are all recognized as risk factors for cardiovascular disease (CVD) and occur in combination more frequently than predicted by chance (Eckel et al., 2005). This cluster of cardiovascular risk factors has been recognized as a clinical entity in its own right and is now referred to as metabolic syndrome (MetS) (Hanefeld and Leonhardt, 1981, Reaven, 1988, Eckel et al., 2005).

Over the past 4 decades have seen a striking increase in the worldwide prevalence of MetS. This surge is closely associated with the global epidemics of obesity and type 2 diabetes mellitus (T2DM) (Ng et al., 2014, Khan et al., 2020). MetS is an established risk factor for progression of prediabetes to diabetes (Wang et al., 2007, Cameron et al., 2008) and many common cancers (Esposito et al., 2012). Yet despite the enormous human and economic cost of MetS (Nichols and Moler, 2011, Sullivan et al., 2007), its cause and interplay of risk factors at a basic biological level is still too poorly understood to provide personalized risk models and tailored plans for intervention.

2.1.2 Definition, epidemiology, economic and clinical aspects

2.1.2.1 Diagnostic Definitions of the metabolic syndrome

Several international bodies, including the World Health Organization (WHO) in 1999, the National Cholesterol Education Program Adult treatment Panel III (NCEP-ATPIII) in 2001, and the International Diabetes Federation (IDF) (Alberti et al., 2005) in 2005 have conceptualized the MetS in terms of overlapping diagnostic criteria. Their common denominator is the required combination of at least 3 of 5 cardiometabolic risk factors. Whilst NCEP-ATPIII criteria allow for any three risk factors to satisfy diagnostic criteria, the definitions formulated by the WHO and the IDF require either impaired glucose metabolism or central obesity respectively in addition to two further risk factors (Table 1). The multiplication of definitions made a unification of criteria desirable to facilitate comparison across cohorts, hence differences in thresholds for fasting plasma (FPG) glucose levels and WC for abdominal obesity were resolved to yield the consensus worldwide IDF criteria in 2006 (Alberti et al., 2006), which has found widespread application next to the NCEP-ATPIII definition and is employed in this study (Table 1).

2.1.2.2 Epidemiology and risk of complications

A comprehensive review by Saklayen (2018) estimated the global prevalence of MetS in 2018 at approximately 25 %. In western cohorts, the levels of MetS prevalence are plateauing in or above this range, whilst developing countries have approached similar levels over the last two decades. According to the most commonly used NCEP-ATPIII and IDF criteria, the overall prevalence of MetS in the United States ranges from 33.0 - 39.0 % (Ford, 2005a, Aguilar et al., 2015). In Europe the general age-adjusted prevalence is 24.3 %, with

considerable variation amongst countries (Scuteri et al., 2015). Mediterranean countries such as Spain and Italy report the highest prevalence of MetS with 31 % (Fernández-Bergés et al., 2012) and 33 % (Tocci et al., 2015), respectively. Estimates have put the prevalence of MetS in Germany at 22.7 % (Moebus et al., 2008).

In the early 2000s, incidence and prevalence in Asian countries were trailing their western counterparts considerably. At that point in China, Japan and Thailand, 9.5 %, 15.3 % and 12.8 % of the respective populations had met the criteria for MetS (Feng et al., 2006, Ishizaka et al., 2005, Lohsoonthorn et al., 2006). However, longitudinal data from Korea (1999 - 2007) and China (1991 - 1995 to 2011 - 2015) exposed a growing MetS prevalence of approximately 0.6 % and 1 % per year (Lim et al., 2011, Huang et al., 2018), whilst similar studies from the United States show no significant change in MetS prevalence in the last two decades (Aguilar et al., 2015, Palmer and Toth, 2019).

The prevalence of MetS is highly dependent on the composition of the studied population in terms age, sex and-ethnicity in addition to the definition used for analysis, resulting in varying estimates even within countries and regions (Cameron et al., 2004, Gray et al., 2000, Alberti et al., 2006). However, an increase of MetS prevalence with age is observed regardless of cohort, peaking after around 60 years of age (Hirode and Wong, 2020, Ford, 2005a, Lim et al., 2011).

MetS accounts for up to 34 % of incident CVD and between 47 % to 62 % of diabetes over 8 years follow-up (Wilson et al., 2005). The cumulative lifetime incidence and prevalence must be considerably higher, thus the global rise in diabetes prevalence thought to be largely driven by MetS (Grundy, 2008). A meta-analysis of longitudinal data from several US cohorts with follow-up between 3 and 13.5 years attributes ~ 6 – 7 % of all-cause mortality to MetS (Ford, 2005b). A separate US-based investigation concluded that MetS conferred a 2.82 to 5.52 fold risk for cardiovascular mortality, more than double the odds ratio (OR) of all-cause mortality reported by the same investigation (Ho et al., 2008).

Table 1 - Diagnostic definitions of the metabolic syndrome

BMI, body mass index; IDF, international diabetes federation; M, men; NCEP-ATPIII, National Cholesterol Education Program Adult treatment Panel III; W, women; WHO, World Health Organization.

Risk factor	WHO	NCEP-ATPIII 2005	IDF 2006
Comment	Insulin resistance/ impaired glucose metabolism plus at least two risk factors.	Presence of at least three risk factors.	Abdominal obesity in terms of waist circumference and at least two more risk factors.
Central/ Abdominal obesity (Waist circumference/ body mass index)	Waist/hip ratio M: > 0.9, W: > 0.85 OR BMI > 30 kg/m2	Waist circumference M: > 102 cm, W: > 88 cm	Waist circumference (Europids)§ M: > 94 cm, W: > 80 cm
Blood pressure	> 140/90 mmHg	> 130/85 mmHg OR drug treatment for hypertension	> 130/85 mmHg OR drug treatment for hypertension
Insulin resistance/ impaired glucose metabolism	blood glucose > 6.1 mmol/L (110 mg/dl) OR 2 h blood glucose > 7.8 mmol (140 mg/dl) (required)	Blood glucose > 5.6 mmol/L (100 mg/dl) OR drug treatment for elevated blood glucose	Blood glucose > 5.6 mmol/L (100 mg/dl) OR drug treatment for elevated blood glucose

Triglycerides	> 1.7 mmol/L (150 mg/dl)	> 1.7 mmol/L (150 mg/dl) OR drug treatment for elevated triglycerides	> 1.7 mmol/L (150 mg/dl) OR drug treatment for elevated triglycerides
High-density lipoprotein cholesterol	M: < 0.9 mmol/L (35 mg/dL), F: < 1.0 mmol/L (40 mg/dL)	M: < 1.0 mmol/L (40 mg/dl), F: < 1.3 mmol/L (50 mg/dl) OR drug treatment for elevated HDL-cholesterol	M: < 1.0 mmol/L (40 mg/dl), F: < 1.3 mmol/L (50 mg/dl) OR drug treatment for elevated HDL-cholesterol

§ cutoffs for waist circumference are ethnicity- and gender-specific.

Vascular complications are not limited to CVD. Individuals with MetS are also at higher risk of ischemic stroke (adjusted OR 1.5); based on findings of a US cohort that attributes ~ 20 % of ischemic strokes to the MetS (Boden-Albala et al., 2008). Furthermore, in both Chinese and Scandinavian longitudinal cohort studies accelerated age-related decline of glomerular filtration rate was found to be associated with MetS after adjusting for epidemiological confounders (Chen et al., 2007, Stefansson et al., 2018).

MetS is also linked to increased incidence of many frequent and rare cancers (Esposito et al., 2012), including prostate (Lund Haheim et al., 2006), breast (Dibaba et al., 2018), and early onset colorectal cancers (Chen et al., 2020a).

2.1.2.3 Economic impact of metabolic syndrome

The economic perspective on MetS is equally disconcerting. Although there is no recent study estimating the global cumulative economic burden of MetS, the economic impact of its most severe complications T2DM and CVD have been thoroughly investigated.

In 2015, the global economic burden of T2DM in adults, from medical costs including complications such as CVD and indirect costs due to lost productivity, disability and death was estimated at $1.31 trillion or 1.8 % of global gross domestic product (GDP) (Einarson et al., 2018). To put this staggering sum in perspective, global expenditure for armed violence, war, terrorism and military budgets amounted to $2.1 trillion (2.8 % GDP) in 2012 (Dobbs et al., 2014).

In addition, MetS incurs increased medical costs regardless of incident cardiovascular complications and hospitalization (Nichols and Moler, 2011). A recent economic analysis of MetS components in a large US cohort (n = 43,037) revealed that individuals were up to 75 % more likely to miss days at work, up to 39 % more likely to require emergency room care, and saw and increase of direct

medical cost of 117 % after adjusting for epidemiological confounders (Mcqueen et al., 2016).

2.1.2.4 The components of the metabolic syndrome

This section provides a brief overview of the five cardiometabolic risk factors that serve as diagnostic criteria for the MetS and their pathophysiological mechanisms contributing to increased risk for cardiovascular disease and other comorbidities. By their very nature, occurring as a cluster and seldom in isolation, a disentanglement of their relative impact on disease mechanism is currently not possible.

2.1.2.4.1 Impaired glucose metabolism

Elevated blood glucose profiles are the diagnostic definitions of both prediabetes and diabetes. The insulin resistance of target tissues has been identified as root cause of T2DM and has been reviewed in detail by (Yaribeygi et al., 2019).

The diagnostic criteria for prediabetes and diabetes have been defined in a consensus report by the WHO & IDF (2006). The criteria for prediabetes are satisfied by diagnosis of impaired glucose tolerance (IGT) > 5.6 mmol/l (100 mg/dl) and/or impaired fasting glucose (IFG) 5.6 – 6.9 mmol/l (100 – 125 mg/dl) and 2-hour plasma glucose of 7.8 – 11.0 mmol/l (140 – 199 mg/dl) in an oral glucose tolerance test (OGTT). Diabetes is diagnosed by an OGTT with 2-h plasma glucose levels of ≥ 11.1 mmol/l (200 mg/dl) and/or a fasting glucose levels of ≥ 7.0 mmol/l (126 mg/dl) .

Elevated glucose levels correlate with carotid intima media thickness, a common surrogate of atherosclerotic disease progression, in a dose dependent manner (Bulut and Avci, 2019). Prediabetes is significantly associated with total plaque area and number across all major vascular territories only in the presence of at least two other MetS components (Sanchez et al., 2019). Carotid intima media thickness however is independently associated with prediabetes (Bamberg et al., 2017). Prediabetes and diabetes mellitus are further associated with coronary atherosclerotic burden (Acar et al., 2019, Scicali et al., 2016). Diabetes also predicts progress of coronary atherosclerotic plaque volume (Yang et al., 2019).

Glucose is a reactive substance developing relevant toxic effects at the increased concentrations found in prediabetes and diabetes. This glucotoxicity underlies complications including CVD, diabetic neuropathy, nephropathy, retinopathy and cataract formation. Details have been reviewed in detail elsewhere (Singh et al., 2014). In brief, glucose spontaneously undergoes a non-enzymatic conversion forming reactive intermediaries, so-called advanced glycation end products (AGEs). AGE formation is accompanied by oxidative stress, which favors the generation of specific AGEs through glycoxidation of amino acid side chains but also cross-linking of proteins. Both the vascular system and in particular the endothelial cells are exposed to hyperglycemia, resulting in the accumulation of AGEs and cellular damage, further causing inflammation, macrophage invasion and tissue remodeling (Rhee and Kim, 2018). AGEs also cause the uncoupling of nitric oxide synthase, resulting in diminished production of the vasoprotective NO and instead superoxide anion (O_2^-) formation, which further impairs endothelial repair (Forstermann and Sessa, 2012). In conjunction these processes all participate in the progression of atherosclerosis as evidenced by increased hypertension, carotid intima-media thickness, severity of plaque formation as well as worse clinical outcomes in terms of incident cardiovascular and other vascular complications (Hegab et al., 2012).

The current standard of care aims at slowing the progress of prediabetes to diabetes through lifestyle modifications (American Diabetes Association,

2020). These measures including exercise, dietary adjustment and weight loss can reduce the relative risk of progress to T2DM by approximately 60 % (Perreault et al., 2012).

For individuals that are diagnosed with T2DM, the latest European and American consensus guidelines recommend a combination of lifestyle modifications and pharmaceutical intervention comprising oral hypoglycemic agents and subcutaneous insulin for glycemic control (Davies et al., 2018).

2.1.2.4.2 Impaired lipid homeostasis

The abnormal concentration of lipids in circulation, or dyslipidemia, includes increased concentrations of plasma triglycerides > 1.7 mmol/L (150 mg/dl) (hypertriglyceridemia), low levels of high-density lipoprotein cholesterol (HDL-C) with sex-specific cutoffs (men: < 1.0 mmol/L [40 mg/dl], women: < 1.3 mmol/L [50 mg/dl]), and the appearance of low-density lipoprotein cholesterol (LDL-C). Both hypertriglyceridemia and low HDL-C have been shown to be independent predictors of cardiovascular risk through acceleration of atherosclerotic plaque formation (Hokanson and Austin, 1996, Yarnell et al., 2001, Barter and Genest, 2019). Both lipid abnormalities often appear in tandem (Fruchart et al., 2004, Castelli, 1992). Furthermore, insulin resistance has been recognized as a contributor in the development of hypertriglyceridemia and low HDL-C (Reaven and Chen, 1988, Reaven et al., 1967, Ginsberg et al., 2005).

Current treatment guidelines for dyslipidemia recommend lifestyle modifications and prescription of lipid lowering medication (Mach et al., 2020). Statins are the first line medication for treating both hypertriglyceridemia HDL-C (Collins et al., 2016). Statin therapy can be further supplemented with other drugs such as ezetimibe (Vavlukis and Vavlukis, 2018) and/or PCSK9 inhibitors (Kaddoura et al., 2020) if statins alone are not sufficient to normalize the blood lipid profile.

2.1.2.4.3 Obesity

The excess accumulation of fat in adipose tissue to an extend that is associated with deleterious effects on health is referred to as obesity (Haslam and James, 2005). According to the WHO (2020), individuals with a body mass index (BMI) as > 30 kg/m^2 are considered obese.

It has been recognized that beneath the clinical presentation of obesity, the distribution and relative amount of fat stored in the adipose tissue in different anatomical sites including subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) are superior to anthropometric aggregate parameters such as WC and BMI at stratifying and assessing cardiometabolic risk (Kwon et al., 2017, Kaess et al., 2012, Liu et al., 2010). Furthermore, ectopic fat depots, referring to fat storage in non-adipose tissue such as the liver or the pericardial sac have emerged more recently as CVD risk factors (Rosito et al., 2008, Cai et al., 2020, Rado et al., 2019).

Even though the association of obesity and insulin resistance has not been disputed since they were formally linked in an early description of MetS by Gerald Reaven (1988) the pathological mechanism is elusive. In broader terms there is agreement that obesity is marked by an accumulation of inflammatory immune cells that contribute to insulin resistance and various tissue. Various hypotheses regarding direction of causality and the pathobiology have been put forward in this context and have been reviewed recently (Wu and Ballantyne, 2020).

The treatment of obesity is based on the clinical disease model that regards the net surplus of calories as principle modifiable factor in the pathological built-up of fat. Thus, current guidelines recommend exercise and dietary intervention as therapeutic entry point (Yumuk et al., 2014). For pharmaceutical intervention orlistat, lorcaserin and phentermine/topiramate have been approved for long term weight management (Toplak et al., 2015). In those individuals that are either highly obese (≥ 40 kg/m^2) or present with obesity (≥ 30 kg/m^2) and comorbidities, escalating treatment to bariatric surgery should be considered to improve long-term outcomes (Fried et al., 2013).

2.1.2.4.4 Hypertension

Hypertension is caused by a mismatch of cardiac output and vascular resistance resulting in elevated systolic and/or diastolic blood pressure. Both European (Williams et al., 2018) and American (Unger et al., 2020) guidelines agree on critical thresholds of bloop pressure (BP) classification. A BP < 130/85 mmHg is defined as normal and BP in the range 130 - 139/85 - 89 mmHg is considered high-normal. The diagnostic threshold for hypertension is set at 140/90 mmHg (Williams et al., 2018, Unger et al., 2020). A meta-analysis of 61 prospective studies with more than 1 million adults concluded that a 20 mmHg increase in systolic blood pressure is associated with a more than twofold risk of death from each stroke, ischemic heart disease and all other vascular causes (Lewington et al., 2002).

In MetS, insulin resistance stimulates the sympathetic nervous system, upregulates angiotensin II receptors and reduces the production of NO (Mendizabal et al., 2013). In combination these effects cause hypertension by simultaneously increasing cardiac output on one hand and vascular resistance on the other. Approximately 80 % of individuals with MetS also suffer from hypertension (Katsimardou et al., 2020). MetS is also independently associated with an increased risk of poorly controlled blood pressure and conveys an increased risk of cardiovascular complications compared to individuals with isolated hypertension (Schillaci et al., 2004).

Despite such findings, recently updated guidelines on the treatment of hypertension do not contain specific recommendations for individuals with MetS (Williams et al., 2018, Unger et al., 2020). Hypertension, like the other 4 components of MetS, can be treated with lifestyle interventions that overlap in terms of weight loss, physical activity but also embrace dietary salt restriction (Goit and Yang, 2019). Treatment can be escalated with common anti-hypertensive drugs such as thiazide diuretics, angiotensin converting enzyme inhibitors, angiotensin receptor blockers and calcium channel blockers. These oral agents have a proven track record of lowering blood pressure and reducing

the risk of cardiovascular events (ALLHAT, 2002). For full guidelines of hypertensive treatment please refer to Williams et al. (2018) and Unger et al. (2020).

2.2 The liver and the metabolic syndrome

The liver plays host to multiple metabolic pathways implied in the pathophysiology of MetS. Hepatocytes host the necessary enzymatic pathways for carbohydrate and lipid homeostasis (Trefts et al., 2017) and synthesizes a variety of systemically active vasoregulatory and vasoprotective enzymes (Anavi et al., 2017). In healthy individuals the liver achieves physiological glucose homeostasis by lowering (postbrandially) elevated blood glucose concentrations through glucose utilization and storage by glycogenesis and raising low blood glucose concentrations (during fasting) by releasing glucose generated by glycolysis and gluconeogenesis (Gerich, 1993). It is estimated that the liver is able to supply up to 90 % of blood glucose during prolonged fasting (Konig et al., 2012). The homeostatic response is under hormonal control of insulin, the only hormone known to lower blood glucose and its main opponents glucagon and epinephrine (Gerich, 1993). The liver is likewise a central hub of lipid metabolism. Hepatocytes convert free fatty acids (FFA) to triglycerides for storage and are the only source of lipoproteins that enable lipid circulation (Nguyen et al., 2008).

Insulin resistance, the key endocrine dysfunction underlying all components of MetS including T2DM also manifests in the liver. Here, insulin is no longer able to moderate the net output of glucose from the liver contributing to increased blood glucose levels (Trefts et al., 2017). At the same time, insulin loses its ability to suppress the production of triglyceride-rich very low density lipoprotein (VLDL) particles, whilst the stimulatory effect of insulin on de-novo lipid synthesis (DNL) is preserved (Smith et al., 2020) and the rate of cholesterol synthesis even increases (Pihlajamaki et al., 2004).

Whilst DNL is a minor pathway in healthy individuals, it is highly upregulated by the combination of elevated insulin and glucose concentrations, i.e. in a state of insulin resistance (Trefts et al., 2017). DNL is the major contributor to pathological fat accumulation by hepatocytes (Smith et al., 2020). In metabolically healthy young adults the hepatic fat fraction determined by clinical imaging rarely exceeds 4-5 % (Ulbrich et al., 2015), a value similar to skeletal muscle (Trinh et al., 2017). Once pathological DNL has been triggered by insulin resistance hepatic fat fraction frequently exceeds 5 %. In the absence of significant alcohol consumption or steatogenic medication this form of hepatic steatosis is referred to as non-alcoholic fatty liver disease (NAFLD) (Chalasani et al., 2018). Furthermore, the catalytic steps of DNL are a major source of oxidative stress and toxic lipid metabolites such as palmitate. A complication of the resulting irreversible cell damage and activation of pro-inflammatory signaling pathways is the increased incidence of liver fibrosis in individuals with NAFLD. The acute inflammatory form of NAFLD is the non-alcoholic steatohepatitis (NASH), which is marked by overt liver inflammation and accelerates progress to cirrhosis and liver failure (Chalasani et al., 2018). The increasing prevalence of liver fibrosis due to NAFLD (and NASH) also translates into higher incidence rates of hepatocellular carcinoma (HCC) (Younossi et al., 2015).

NAFLD receives increasing recognition as a mediator of MetS, T2DM and CVD (Yki-Jarvinen, 2014). The interplay of NAFLD and cardiometabolic risk factors was addressed by multiple longitudinal studies, which came to the consistent conclusion that NAFLD independently predicts incident prediabetes (Zelber-Sagi et al., 2013), T2DM (Ballestri et al., 2016) and MetS (Ballestri et al., 2016). In particular, the effect on T2DM incidence was dose dependent and reduction of steatosis grade resulted in decreased risk of new-onset T2DM (Yamazaki et al., 2015, Fukuda et al., 2016, Lee et al., 2020). Longitudinal cohort studies that relied on plasma enzyme biomarkers, an established semi-quantitative marker of NAFLD, found a decreased survival in long term follow-ups (Nasr et al., 2020). Furthermore, prospective cohort studies concluded that NAFLD predicts subclinical atherosclerosis (Zhou et al., 2018) as well as CVD (Targher et al., 2016).

The mechanisms underlying these associations between NAFLD, MetS and CVD can be coherently explained, at least in part, in terms of the aberrant hepatocyte function described above. The hepatic overproduction of glucose contributes to hyperglycemia in prediabetes and T2DM. Steatotic hepatocytes also overproduce coagulation factors and fibrinogen thereby inducing a systemic procoagulatory state (Targher et al., 2009, Sookoian et al., 2010). The increased synthesis of triglyceride-rich VLDL and cholesterol synthesis are complicit in atherosclerotic plaque formation (Nordestgaard, 2016).

The epidemiology of NAFLD is tightly entwined with obesity and MetS (Yki-Jarvinen, 2014) and has emerged as the most common liver disease worldwide, with approximately 25 % of adults in the US and Europe being currently affected (Younossi et al., 2018). In respect to both its shared epidemiology and pathophysiology, NAFLD can be regarded as the hepatic manifestation of MetS.

2.3 MR for assessment of hepatic fat

Two non-invasive magnetic resonance (MR)-based imaging techniques, magnetic resonance imaging (MRI) and ^{1}H-magnetic resonance spectroscopy (MRS) are available for the quantification of hepatic fat fraction (Bohte et al., 2011). Both methods take advantage of the different resonance frequencies (chemical shifts) between water- and fat-bound protons to decompose the liver signal into its component water- (S_{Water}) and fat-signals (S_{Fat}). The (per voxel) signal fraction attributable to fat, also known as signal fat fraction or relative fat water content (rfwc) can be calculated according to equation 1 (Reeder et al., 2011).

$$\text{(equ. 1) } rfwc\ (\%) = \frac{S_{Fat}}{S_{Fat} + S_{Water}} \times 100$$

MRI employs so-called chemical shift-based methods (named “Dixon”-sequences named after its discoverer (Dixon, 1984)) to obtain the water- and fat-

components. Due to the differences in chemical shift, water- and fat-bound protons precess at slightly different frequencies ($\Delta \approx 440$ Hz at 3 Tesla [T]). For that reason their magnetization vectors will alternate periodically between pointing in the same direction ("in-phase") resulting in signal summation and pointing in the opposite direction ("out-of-phase") with converse (partial) signal cancellation. Acquisition sequences thus acquire at least two echo times (TEs) (as in T1-weighted dual-echo Dixon [T1-DED] MRI) beginning with the first out-of-phase signal (S_{OP}; 1.23 ms at 3T) and the first in-phase signal (S_{IP}; 2.46 ms at 3T) in order to maximize the signal-to-noise ratio and minimize signal loss from T2* decay. The S_{Water} and S_{Fat} components can then be derived using equations 2 a and b (Dixon, 1984).

$$\text{(equ. 2 a)} \quad S_{Water} = |S_{IP} + S_{OP}|$$

$$\text{(equ. 2 b)} \quad S_{Fat} = |S_{IP} - S_{OP}|$$

MRS by comparison relies on direct acquisition of proton signal spectrums from manually placed large voxels (eg. $3 \times 3 \times 3$ cm^3). The S_{Water} and S_{Fat} components are then quantified as the areas under the curves from the spectral traces (Thomsen et al., 1994).

Both methods suffer effects from a common set of confounders that need to be adjusted in order to convert the rfwc to the more accurate proton density fat fraction (PDFF) (Reeder et al., 2011). The confounding factors are the T1 bias, the T2 bias/T2* decay, and the spectral complexity of fat (Reeder et al., 2011).

The T1 bias leads to the overestimation of the T1 fat-signal which relaxes faster than the corresponding T1 water-signal. This effect is countered by operating MRI sequences at small flip angles (Fishbein et al., 1997) to reduce T1-weighting and by using long repetition times (TR) for MRS (Kim et al., 2008), respectively.

T2 bias is a phenomenon arising because even at the smallest possible TE some degree of T2 relaxation occurs before echo acquisition. A common strategy is to sample multiple echos to directly estimate and adjust for the T2 bias (Sharma et al., 2009). T2* decay is the apparent T2 relaxation, confounded by

local magnetic field inhomogeneities. A robust approximation and correction of T2* related signal decay can be achieved by applying echo trains with multiple TEs (Meisamy et al., 2011), as implemented for example in T1 multi-echo Dixon (T1-MED) MRI (Hetterich et al., 2016).

Importantly, both T2 bias and T2* decay are amplified by hepatic iron overload, thus adjustment permits more accurate quantification of the hepatic fat fraction (Lee et al., 2011, Reeder et al., 2011).

Modern MRI post-processing models also correct for the spectral complexity of fat, ensuring that smaller peripheral peaks also contribute to the fat signal (Bydder et al., 2008, Meisamy et al., 2011).

2.4 Radiomics: extracting more from plain sight

Imaging is an important branch of medical science and is used in clinical practice to guide decision making (Aerts et al., 2014). As part of a broader evolution in the medical field, imaging is being actively developed to contribute towards personalized precision medicine (Hood and Friend, 2011). In this context, radiomics is one of the central technologies that enable the high-throughput extraction of vast amounts of mineable image features from standard of care imaging data (Kumar et al., 2012, Lambin et al., 2012).

Essentially, radiomics can be described as an umbrella term for equations that transform medical images - or to be more precise - regions of interests describing lesions, round-masses or organs to a set of quantitative high-dimensional image descriptors (i.e. radiomics features) that capture salient information relating to interactions of voxel intensities and their respective locations. The quantitative nature of radiomics features is also particularly valuable in integrated “-omics” approaches. Radiomics data, which is both complementary and distinct from qualitative image features can be combined with

other patient information (epidemiology, genomic data, treatment) to improve decision making (Gatenby et al., 2013).

Since the inception of the term “radiomics” in 2010 (Gillies et al., 2010), the field has moved rapidly and shown many promising results in oncological diagnostic support, prediction of therapy response and clinical endpoints (Nougaret et al., 2019).

The main challenge faced by this field is the integration of radiomics information from different patient examinations and study centers to create standardized and robust models to guide clinical evaluation and decision making (Lambin et al., 2017). The utility of radiomics data can be greatly diminished by the variance introduced by reader disagreement, changes in patient positioning between sessions but also vendor specific hard- and software implementations, in addition to random noise (Zwanenburg et al., 2019). Several computational approaches have been presented that allow the assessment of feature robustness, also referred to as stability, without performing time consuming and costly physical test-retests and multiple manual delineations to gauge the effects of test-retest and inter-rater variance on radiomics model performance (Gevaert et al., 2014, Bologna et al., 2018, Zwanenburg et al., 2019). These concepts to simulate re-test and inter-rater variance were adapted by this study to select stable radiomics features, as described in later sections.

2.5 Aim and scientific hypothesis

Individuals with NAFLD are more likely to develop insulin resistance, T2DM, MetS and CVD and suffer from higher mortality compared to the general population. These entwined cardiometabolic disorders and their comorbidities place a considerable burden on global societies and healthcare systems, both in terms of human cost and economic expenditure.

The clinical use of MetS as a tool for individual risk stratification is limited due to its susceptibility to various confounders including age, gender and ethnicity that are difficult to adjust for resulting in broad confidence intervals in the OR of all-cause, cardiovascular and cancer mortality (Shi et al., 2020). Meanwhile researchers have widened their focus from isolated components of MetS to an integrated “-omics” approach that draws on liquid-biopsy and medical imaging to identify useful biomarkers to improve risk stratification and allow for targeted and personalized intervention. In recent years MRI has become the imaging modality of choice for the necessary large-scale cohort studies. In contrast to CT, MRI image acquisition is radiation free and generates images with the high soft-tissue contrast necessary for accurate assessment of the inner organs and fat deposits linked to cardiometabolic risk. Regarding the hepatic tissue aberrations of NAFDL, MRI-based imaging provided highly accurate estimates of hepatic fat fraction (HFF) (Bonekamp et al., 2014, Noureddin et al., 2013) and outperforms serological markers in the detection of early stage fibrosis (Park et al., 2019).

The aim of this study was to assess the microheterogeneity of hepatic fat in a cross-sectional cohort of 400 individuals from southern Germany without prior cardiovascular disease and representative prevalence of cardiometabolic risk factors including T2DM, MetS and NAFLD. Our analysis was based on volumetric segmentations of the entire liver, using T1-weighted dual-echo Dixon (T1-DED) image data to obtain per-voxel estimates of hepatic fat fraction. We hypothesized that hepatic fat heterogeneity, captured by radiomics features would allow to predict the participants’ T2DM and MetS status. To that end, radiomics features of hepatic fat fraction maps were extracted and assessed for stability in computationally simulated test-retest and inter-rater setups. Stable features were evaluated as biomarkers for T2DM and MetS in predictive machine learning models.

Parts of the microheterogeneity analysis have previously been published (Gutmann et al., 2020).

3 Materials and Methods

3.1 Study population

This study used image data collected from participants (*n* = 400) of the Cooperative Health Research in the Region of Augsburg (KORA) MRI substudy carried out in 2013 - 2014. The cohort was nested within the 14 year follow-up FF4 study (*n* = 2,279) of the prospective population-based KORA survey S4 (1999-2001; *n* = 4,261). T

Participants of the FF4 study were eligible to be included in the MRI substudy if they had no history of prior cardiovascular events including myocardial infarction, stroke or revascularization, were < 73 years of age and had no contraindications to contrast-enhanced MRI examinations. These included medical implants, claustrophobia, pregnancy or breast-feeding, renal impairment (serum creatinine ≥ 1.3 mg/dl) and known allergy to contrast agents. The criteria were previously discussed (Bamberg et al., 2017).

Full written consent was obtained from all individuals that were willing to participate. Ethical approval was also obtained from the institutional review board of the Medical Faculty of the Ludwig Maximilians University Munich, Germany. This ethical approval was reviewed and confirmed by the institutional review board of the Medical Faculty of the Eberhard Karls University Tübingen, Germany under the project number 576/2016BO2. Figure 1 provides a full overview of the study design.

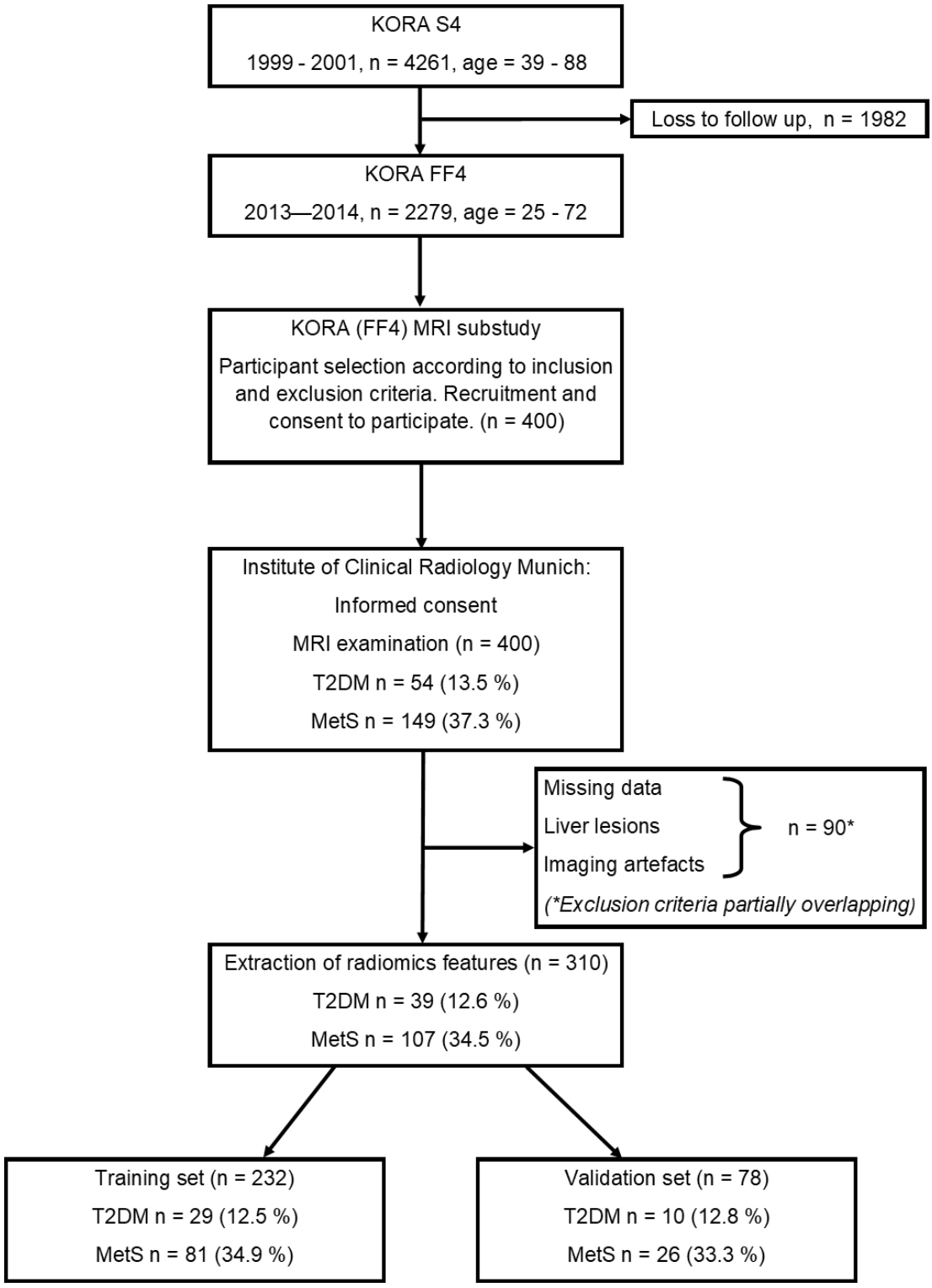

Figure 1 – Structure of the KORA study and participant selection

3.2 Covariate collection

Participants of the KORA MRI substudy were examined in a standardized fashion between June 2013 and September 2014 at the KORA study center. All covariables were collected from standardized interviews, laboratory tests and physical exams during the study visit. Covariate collection has been described in full (Holle et al., 2005, Bamberg et al., 2017) physical exams during the study visit. Covariate collection has been described in full (Holle et al., 2005, Bamberg et al., 2017)

3.3 MRI acquisition parameters

All participants underwent identical imaging protocols on the same 3.0 Tesla (T) whole body scanner (Magnetom Skyra, Siemens Healthcare, Erlangen, Germany). Participants were examined in supine position using an 18-channel body surface coil in conjunction with a table mounted spine matrix coil. Image data was acquired using a three-dimensional parallel accelerated volume interpolated T1-DED prototype sequence (VIBE) (Bashir et al., 2012, Bashir et al., 2013). Imaging parameters constituted TEs of 1.23 ms (out-of-phase), 2.46 ms (in-phase), TR of 4.10 ms, flip angle of 9° and isotropic resolution of 1.7 mm. Image data was acquired during a single breath hold of 15 sec to reduce movement artefacts.

Details of T1-weighted T2*-corrected multi-echo Dixon (T1-MED) and T2-corrected ^{1}H-spectroscopy (MRS) sequences, and analysis of hepatic proton density fat fractions (PDFF) are described elsewhere (Hetterich et al., 2016).

3.4 Segmentation of volumetric liver masks

For this study, a single reader manually delineated volumetric liver masks in the coronal plane of T1-DED water-phase maps employing the Medical Imaging Interaction Toolkit Workbench (MITK; release 2015.5.2) (Wolf et al., 2005). Results were checked by an experienced radiologist with more than 5 years experience in abdominal imaging and corrections were made as required. The resulting reference masks were screened for imaging artefacts by a single reader and images containing artefacts were excluded.

3.5 Image and radiomics data processing.

All necessary computational steps including radiomics features extraction were implemented using the programming language Python v2.7 (Sanner, 1999) and compatible open-source software libraries as described in the following sections.

Firstly, the KORA MRI image data and the volumetric liver reference masks were converted to 3D arrays utilizing the SimpleITK package v.1.0 (Lowekamp et al., 2013). The computation of rfwc maps and all further data manipulation necessary for the generation of artificial test-retest and inter-rater scenarios were conveniently performed on these 3D arrays.

3.5.1 Generation liver volumes of interest

To avoid erroneous inclusion of visceral fat, the initial liver reference masks were eroded by 3 voxels to create the final liver volumes of interest (VOIs) using scipy v1.1.0 (Virtanen et al., 2020).

3.5.2 Computation of rfwc maps

Volumetric rfwc maps were computed from water- and fat-phase T1-DED maps by applying equation 1 to their corresponding image arrays.

3.5.3 Artificial test-retest and inter-rater scenarios

To simulate a retest, copies of the rfwc maps were augmented with synthetic noise ($rfwc_N$) to mimic the typical variance introduced by differences in acquisition and reconstruction parameters. Image noise was approximated by a gaussian distribution with a standard deviation (SD) corresponding to 1.49 % rfwc. This value was chosen based on a meta-analysis that had evaluated the test-retest variance of MRI quantified hepatic fat fractions from 425 individuals published in 11 precision studies (Yokoo et al., 2018).

The effect of variations from manual delineations between raters was assessed by creating computationally deformed VOIs (VOI_D). Three deformed VOI_D were created for each VOI by applying shape transforms implemented by the SimplTK library (Lowekamp et al., 2013). The dice coefficient (mean ± SD) of 0.89 ± 0.08 was chosen based on a literature search of previously published inter-rater variance of abdominal organ segmentation on MRI images (Noel et al., 2014).

After completion of all necessary steps, the 3D arrays representing the rfwc, $rfwc_N$ and $VOI_{(D)}$ were converted to the Nifty file format (.nii) as a suitable input for radiomics feature extraction. The original Dicom header files were used for voxel coordinate mapping. Furthermore, all image data and volumetric masks were interpolated to exact isotropic voxel dimensions of 1.7 x 1.7 x 1.7 mm^3 to account for minimal incongruities.

3.5.4 Extraction of radiomics features

For the radiomics analysis a total of 684 features were determined using the open-source software package PyRadiomics (v1.30) (Van Griethuysen et al., 2017). This total constitutes 74 radiomics features for each of 9 gray level discretization bin widths (gldbw) and 18 first order features that are independent of the gldbw, respectively.

The 74 radiomics features belonged to 5 families of quantitative image descriptors: gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighboring gray tone difference matrix (NGTDM) and gray level dependence matrix (GLDM). A summary is provided in Table 2 a. Gldbw values were selected at intensity bin width of 16, 32, 64, 128, 192, 256, 384, 512 and 640 au (compare Table 2 b). Each textural radiomics feature was calculated in 3D for each voxel including all 26 neighboring voxels in the calculation.

Furthermore, input data was z-score normalized and scaled by a factor of 1000 to attenuate the effect of varying hepatic fat fractions on textural feature computation.

3.5.5 Selection of stable radiomics features

The stability of extracted radiomics features was assessed on the training set by calculating the intraclass correlation coefficients (ICCs) for test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)).

To gauge test-retest reliability the ICC(1,1) with 95 % confidence interval (95 % CI) (Shrout and Fleiss, 1979) was calculated comparing each radiomics feature extracted from both the rfwc and $rfwc_N$ maps using the original VOI as mask input.

Table 2 - Description of radiomics features

"(a) Properties and count of employed radiomics features (b) gldbw and corresponding approximate bin counts. GLCM, gray level co-occurrence matrix; gldbw, gray level discretization bin width; GLDM, gray level dependence matrix; GLSZM, gray level size zone matrix; GLRLM, gray level run length matrix; NGTDM, neighboring gray tone difference." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

a)

Radiomics features	# features	Description
First order	18	global intensity statistics; not affected by gldbw
GLCM	23	histogram statistics of voxel intensity co-occurrence
GLSZM	16	statistics of connected voxels of same intensity
GLRLM	16	statistics of consecutive voxels of same intensity
NGTDM	5	statistics of differences between given voxels and surrounding intensities
GLDM	14	statistics of connected voxels of same intensity dependent on center voxel
Total	92	

b)

gldbw	mean bin count
16	175
32	87
64	44
128	22
192	15
256	11
384	7
512	5
640	4

Similarly, the inter-rater agreement was evaluated in terms of the ICC(3,k) with 95 % CIs. Each VOI_D was used as input mask to extract radiomics features from the rfwc map. The inter-rater agreement was calculated using the previously extracted features from the rfwc-VOI combination as reference.

Radiomics features were considered stable if the lower bounds of the 95 % CI (95 % CI - lb) for both test-retest reliability and intra-rater agreement were ≥ 0.85.

Furthermore, feature stability was assumed to be independent of outcomes (T2DM or MetS).

Unless otherwise stated, all ICC values in this thesis refer to the 95 % CI - lb.

3.5.6 Selection of stable features for predictive models

Since the outcome classes were imbalanced, i.e. numbers of individuals not-affected and affected by either T2DM or MetS were unequal, the training data was rebalanced to potentially improve the ensuing machine learning steps. Thus the training data was processed using the synthetic minority over-sampling technique (SMOTE) (Chawla, 2002) which is part the imbalanced-learn package (https://imbalanced-learn.readthedocs.io). This method generates new “synthetic” data points based on the minority class data by combining similar data points. Therefore, two training sets were created, each re-balanced for either T2DM or MetS, respectively.

The following steps describe the construction of predictive radiomics random forest (RF) models for outcomes T2DM and MetS.

First, stable radiomics features were filtered by univariate RF analysis. On the SMOTE re-balanced training data the classification performance of each

radiomics feature was evaluated individually, relying on 5-fold cross-validation (CV) to optimize bias-variance trade-off.

For each outcome features were assigned a univariate performance index (I_P) in descending order according to their CV-score in terms of the receiver operating characteristic area under the curve (AUROC). To achieve low model complexity and stability only the best performing 10 features were used in the subsequent RF model training and evaluation.

Next, radiomics features were subjected to sequential forward aggregation, thus training 10 RF models for each outcome using features with I_P range (I_N) 1 to 10. The training of each model was repeated 10-times with randomized 5-fold CV, recording the mean and SD of each I_N CV-score (AUROC) to allow the calculation of the 95 % CI. For each outcome the RF models with the best I_N CV-score (RF_{Rad}) were used for further evaluation.

3.5.7 Training of benchmark RF models

In order to benchmark the performance of the radiomics-based predictive models against established risk factors, the T1-MED and MRS quantifications of hepatic PDFF as well as the BMI were chosen to train RF models. Analogous to RF_{Rad} models, these benchmark RF models were trained on the SMOTE re-balanced training sets for MetS and T2DM using 5-fold CV.

3.5.8 Training and validation statistics of random forest models

All RF models were evaluated on the independent validation set. To gain the necessary numeric stability the predictive performance was recorded after each iteration of 10 rounds of randomized 5-fold CV. The AUROC and the

balanced accuracy ($Accuracy_B$) on the validation set are reported as mean and the 95 % CI.

3.6 Statistical analysis

Comparisons between variables of the training and validation sets were assessed using the independent *t*-test or Wilcoxon's *ranksums*-test for continuous variables, depending on the results of the Shapiro-test for normality. Categorical variables were compared using the χ^2-test. The mean hepatic fat fraction estimates quantified by T1-DED, T1-MED and MRS imaging were compared using the paired Wilcoxon's *ranksums*-tests. Significance tests were performed using scipy v1.1.0 (Virtanen et al., 2020) and R v3.6 (R Core team, 2016) .

ICCs were calculated using the R-package psych (https://personality-project.org/r/psych/). The lower boundaries of the 95% confidence interval were used for feature selection and all follow-on analysis.

Correlations between variables were estimated with ordinary least square (OLS) linear regression models using the software package StatsModels (Seabold and Perktold, 2010). Results are reported as β - coefficients with the 95% confidence interval in square brackets, the adjusted-R^2 and *p*-value.

Bland-Altman statistics and charts were prepared using the R-package BlandAltmanLeh (https://cran.r-project.org/web/packages/BlandAltmanLeh/). All other figures were prepared using the packages ggplot2 (Wickham, 2009), Matplotlib (Hunter, 2007) and Bokeh (Van De Ven, 2020).

Training and validation set split, feature selection and random forests were implemented in scikit-learn (Pedregosa et al., 2011).

The performance metrics of RF models, AUROC and $Accuracy_B$ are reported as mean and the boundaries of the 95% confidence interval.

Differences were deemed statistically significant at *p*-values below 0.05. All tests were two-tailed. The term “prediction” is used in its statistical sense to denote outcome classification on the validation data and does not imply temporal sequence or incident outcomes.

4 Results

4.1 General Results

The KORA MRI substudy had primarily enrolled 400 individuals. Image data of 310 individuals (77.5 %) were used in this study (Gutmann et al., 2020). A total of 90 participants were excluded because of insufficient image quality (*n* = 53, 13.3 % [eg. breathing/movement- and swap-artefacts, FOV misalignment]), liver lesions (*n* = *3*; 0.8 %) and either missing or incomplete T1-DED sequences (*n* = 37; 9.3 %). Compare Figure 1.

4.2 Quality management

The hepatic fat fraction of the 310 analyzed participants is presented by box-plots in Figure 2. The mean hepatic fat fraction estimated on whole liver T1-DED rfwc maps was 8.50 ± 7.40, compared to 8.43 ± 8.35 by T1-MED and 8.74 ± 7.93 by MRS. The whole liver rfwc measurements (T1-DED) were strongly correlated with the PDFF quantifications from T1-MED and MRS (*R* = 0.967 and 0.950; both $p < 0.001$) (Fig. 3 a, c). Analogous comparison by Bland-Altman plots shows harmonious agreement between the modalities with only few outliers (Fig. 3 b, d) The results of the correlation and Band-Altman analysis are summarized in Table 3.

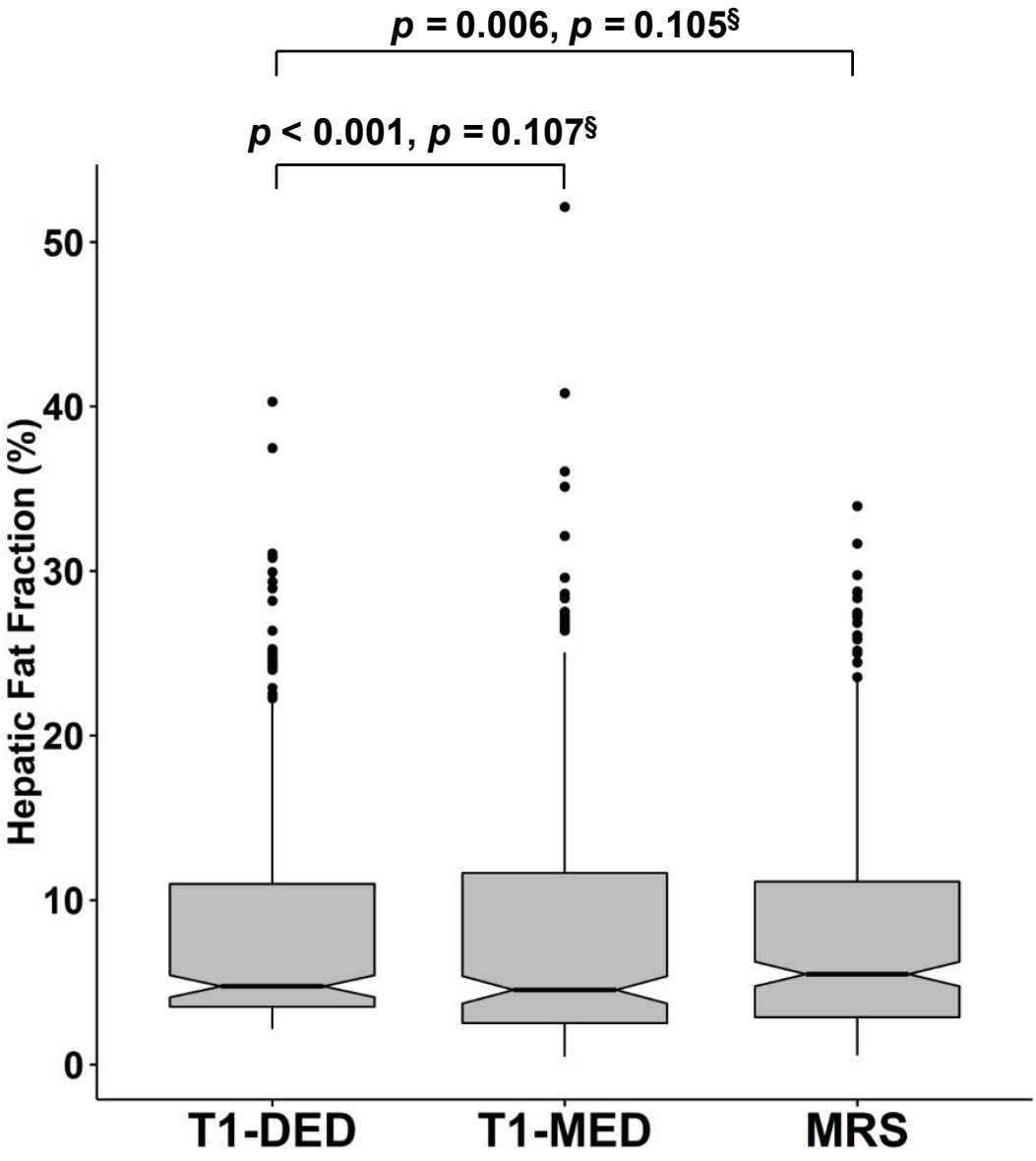

Figure 2 – Hepatic Fat Fraction quantified by T1-DED compared to T1-MED and MRS

Hepatic fat fraction quantifications were performed on volumetric relative fat water maps (T1-weighted dual-echo Dixon, this study) and on manually drawn regions of interest of the liver parenchyme in T1-weighted Multi-echo Dixon and ^{1}H-magnetic resonance spectroscopy proton-density fat fraction maps. § *p* – values after bias calibration. T1-DED, T1-weighted dual-echo Dixon; T1-MED, T1-weighted multi-echo Dixon; MRS, ^{1}H-magnetic resonance spectroscopy.

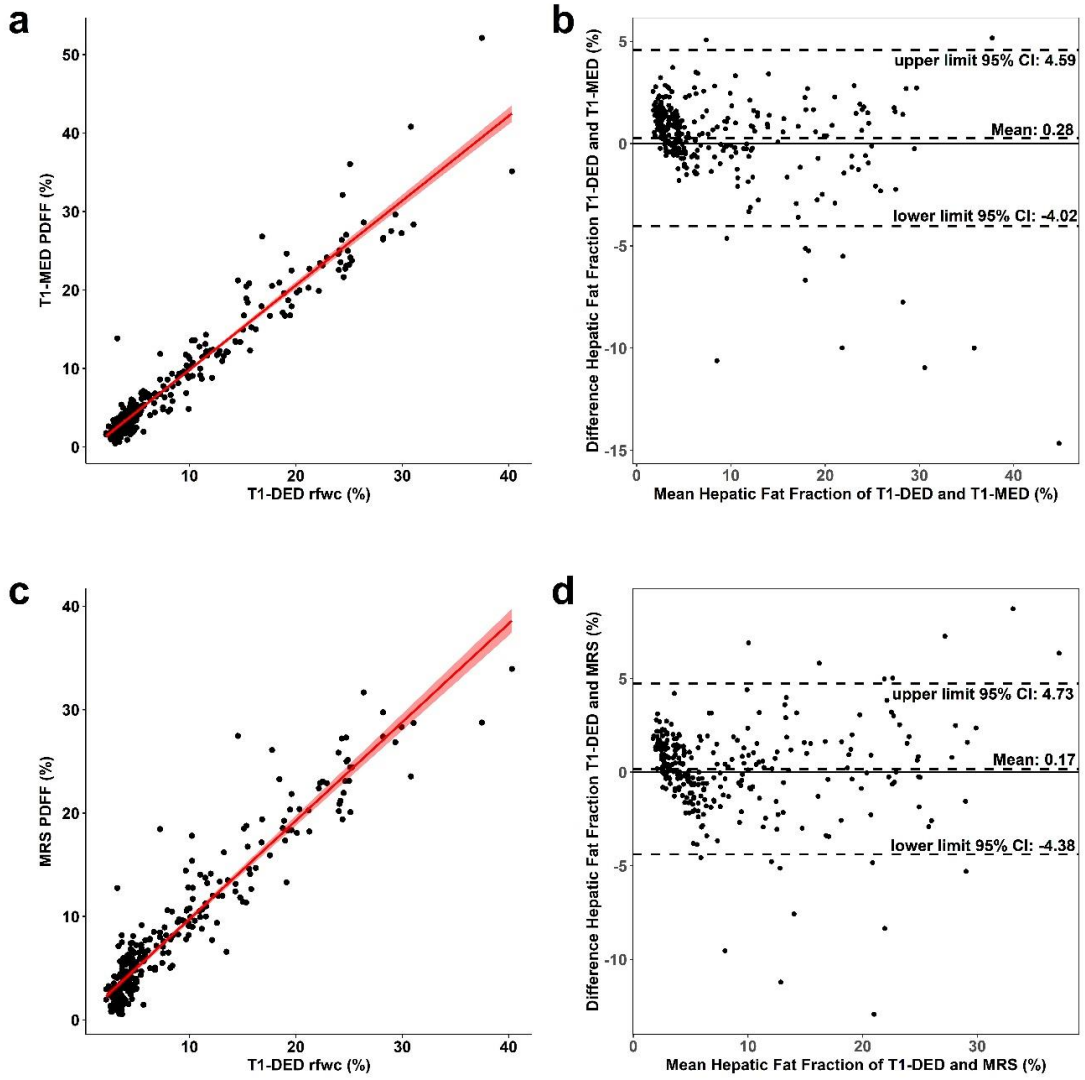

Figure 3 – Correlation and Band-Altman analysis of hepatic fat fraction quantified by T1-DED compared to T1-MED and MRS

T1-weighted dual-echo Dixon and T1-weighted multi-echo Dixon compared by correlation with line of best fit (a) and Bland-Altman plot analysis (b). T1-weighted dual-echo Dixon and ^{1}H-magnetic resonance spectroscopy compared by correlation with line of best fit (c) and Bland-Altman plot analysis (d). T1-DED, T1-weighted dual-echo Dixon; T1-MED, T1-weighted multi-echo Dixon; MRS, ^{1}H-magnetic resonance spectroscopy

Table 3 - Summary of the correlation and Bland-Altman analysis

T1-DED, T1-weighted dual-echo Dixon; T1-MED, T1-weighted multi-echo Dixon; MRS, 1H-magnetic resonance spectroscopy.

Variable	T1-DED / T1-MED	*p*-value	T1-DED / MRS	*p*-value
R	0.967	< 0.001	0.950	< 0.001
Mean difference (95% CI)	0.28 % (-4.02 to 4.59 %)	< 0.001 0.107[$]	0.17 % (-4.38 to 4.73 %)	0.006 0.105[$]

$ *p* - value after bias calibration.

4.3 Epidemiological characteristic of study sample

The study population was predominantly male (57.1 %) and of middle age (mean age 56.1 years). Amongst the participants 12.6 % were diagnosed with T2DM and 34.5% (Gutmann et al., 2020) fulfilled the IDF diagnostic criteria for MetS.

Applied statistical tests revealed no significant difference between the study population and the analyzed subset (p = 0.399 - 0.922) (Gutmann et al., 2020). Furthermore, epidemiological characteristics of training and validation sets were not statistically different (p = 0.443 - 0.995) (Gutmann et al., 2020).

Table 4 provides a complete summary of epidemiological characteristic and statistical results of study population.

4.4 Analysis of selected radiomics features

Analysis of the feature selection process revealed that radiomics features were more robust to inter-rater (567 [82.9 %])(Gutmann et al., 2020) than test-retest variance (181 [26.5 %])(Gutmann et al., 2020). Of initially 684 radiomics features 171 (25.0 %) (Gutmann et al., 2020) exceeded the ICC cut-offs (≥ 0.85) for test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)) in the training set and were thus considered stable (Table 5 a).

Associations of test-retest reliability, inter-rater agreement and the radiomics features' coefficient of variation (CoV) between training and validation sets were investigated with linear regression models. All three parameters were strongly correlated with β - coefficients of 1.027 (Gutmann et al., 2020), 1.072 (Gutmann et al., 2020) and 0.855 (Gutmann et al., 2020) for test-retest reliability (Fig. 4 a), inter-rater agreement (Fig. 4 b) and the CoV (Fig. 5), all with $p < 0.001$ (Gutmann et al., 2020).

The effect of a radiomics features CoV on the stability in the presence of test-retest and inter-rater variance was equally explored. In the training set, the test-retest reliability was positively correlated with the CoV (β 0.277, $p < 0.001$) (Gutmann et al., 2020) (Fig. 6 a), whereas no comparable correlation was found for inter-rater agreement ($p = 0.329$) (Gutmann et al., 2020) (Fig. 6 b). The results of the regression analysis are summarized in Table 6.

Of 171 radiomics features that were stable in the training set and were evaluated for use in predictive RF models, 142 (83.0 %) (Gutmann et al., 2020) remained stable in the validation set (ICC ≥ 0.85). Table 5 b provides a comprehensive summary of the feature stabilities in the validation set.

A box-plot analysis of test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)) of the selected stable features ($n = 171$) in the validation set is provided in Figure 7 a and b, respectively.

Table 4 - Summary of epidemiological characteristics

"Basic characteristics of the KORA-MRI sub-study, analyzed subset, training and validation sets Continuous variables are presented as arithmetic mean ± SD. Categorical variables are presented as counts and percentages. P-values from independent t-tests, Wilcoxon's ranksums-tests or χ2-tests. 95% CI – lb, lower bound of the 95% confidence interval; KORA, Cooperative Health Research in the Region of Augsburg (Kooperative Gesundheitsforschung in der Region Augsburg); MRI, magnetic resonance imaging; PDFF, proton density fat fraction; SD, standard deviation." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

	KORA-MRI	**analyzed subset**	***p* -value (KORA-MRI vs. analyzed subset)**	**training set**	**validation set**	***p*-value (training vs. validation)**
n	400	310		232	78	
gender (% male)	231 (57.8)	177 (57.1)	0.922	133 (57.3)	44 (56.4)	0.992
age (years ± SD)	56.3 ± 9.2	56.1 ± 9.3	0.746	56.1 ± 9.3	56.1 ± 9.2	0.995
weight (kg ± SD)	83.0 ± 16.6	81.9 ± 16.0	0.399	82.2± 16.3	80.9 ± 15.3	0.693
height (cm ± SD)	171.6 ± 9.7	171.5 ± 9.5	0.822	171.5 ± 9.8	171.4 ± 8.8	0.983
body mass index (kg/m^2 ± SD)	28.1 ± 4.9	27.8 ± 4.7	0.451	27.9 ± 4.7	27.5 ± 4.8	0.443
T1-weighted multi-echo Dixon PDFF (% ± SD)	8.43 ± 8.35	8.25 ± 8.29	0.782	8.48 ± 8.70	7.56 ± 6.93	0.802
^{1}H-magnetic resonance spectroscopy PDFF (% ± SD)	8.74 ± 7.93	8.28 ± 7.36	0.582	8.44 ± 7.62	7.78 ± 6.54	0.802
type 2 diabetes mellitus (%)	54 (13.5)	39 (12.6)	0.804	29 (12.5)	10 (12.8)	0.902
metabolic syndrome (%)	149 (37.3)	107 (34.5)	0.501	81 (34.9)	26 (33.3)	0.907

Table 5 - Radiomics features stability selection

"(a) Summary of feature selection according to stability criterion ICC ≥ 0.85 for test-retest reliability (ICC(1,1)), inter-rater agreement (ICC(3,k)) and for both. (b) Stable features selected on training set data that remain stable (ICC ≥ 0.85) in the validation set for test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)) and for both. ICC, intraclass correlation coefficient." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

a) Selection of stable features on the training set

		Stability criteria (ICC ≥ 0.85) fulfilled (training set)		
	All features	**ICC(1,1)**	**ICC(3,k)**	**ICC(1,1) + ICC(3,k)**
n (% all features)	684 (100)	181 (26.5)	567 (82.9)	171 (25.0)

b) Proportion of stable features maintaining stability in the validation set

		Stability criteria (ICC ≥ 0.85) fulfilled (validation set)		
		ICC(1,1)	**ICC(3,k)**	**ICC(1,1) + ICC(3,k)**
n (% stable features per criterion)		150 (82.9)	545 (96.1)	142 (83.0)

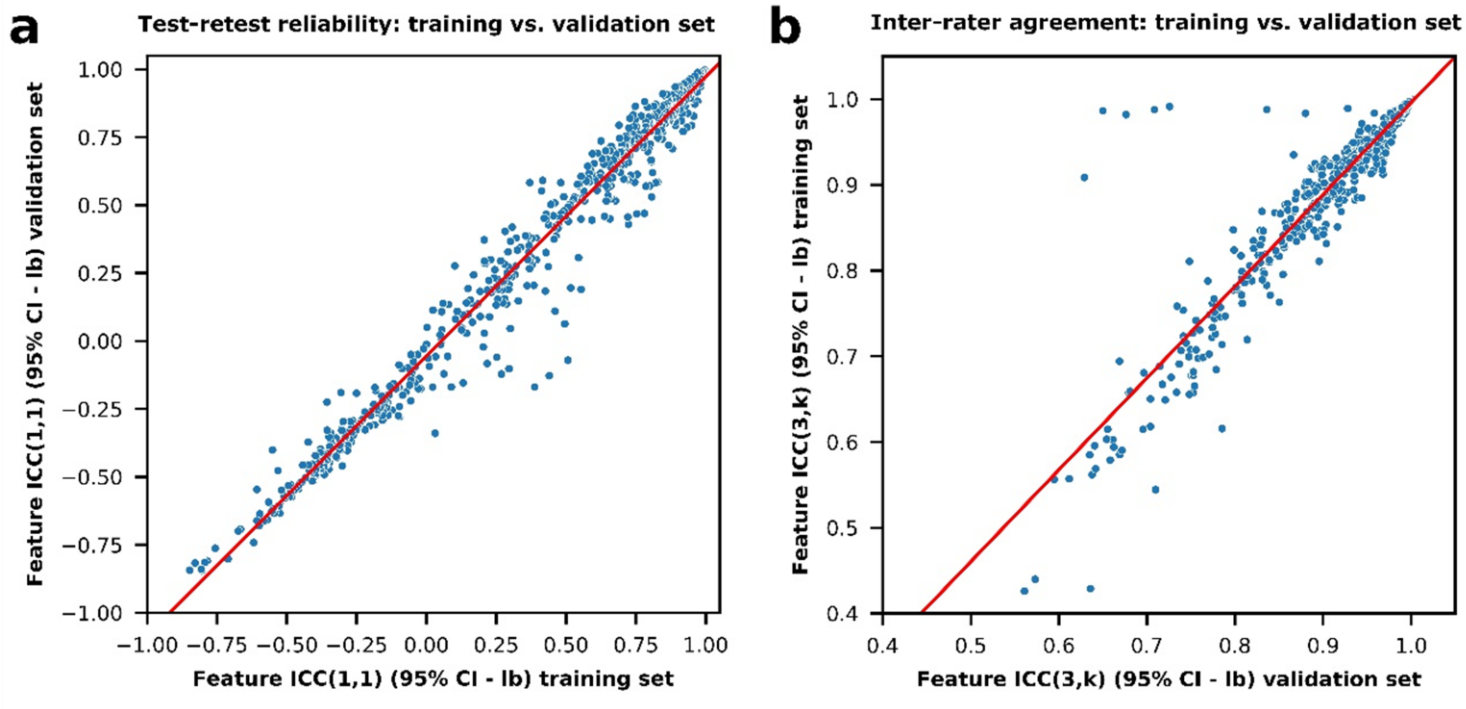

Figure 4 - Feature stability associations in training and validation sets
. *"Scatterplots and linear regression line of best fit of (a) test-retest reliability (ICC(1,1)) and (b) inter-rater agreement (ICC(3,k)) in training and validation sets. 95% CI – lb, lower bound of the 95 % confidence interval; ICC, intraclass correlation coefficient."* This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

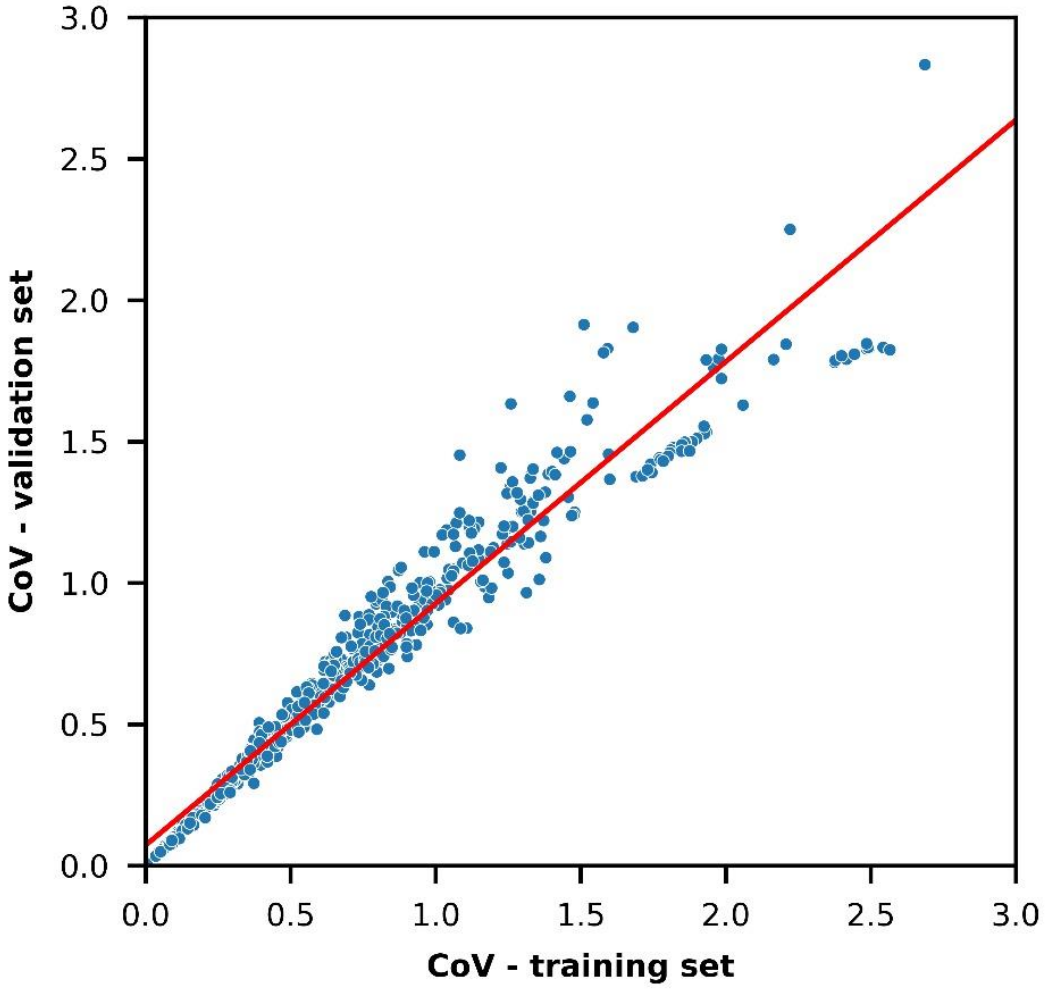

Figure 5 - Association of radiomics feature coefficient of variation in validation and training sets

Scatter plot with line of best fit. CoV, coefficient of variation. This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

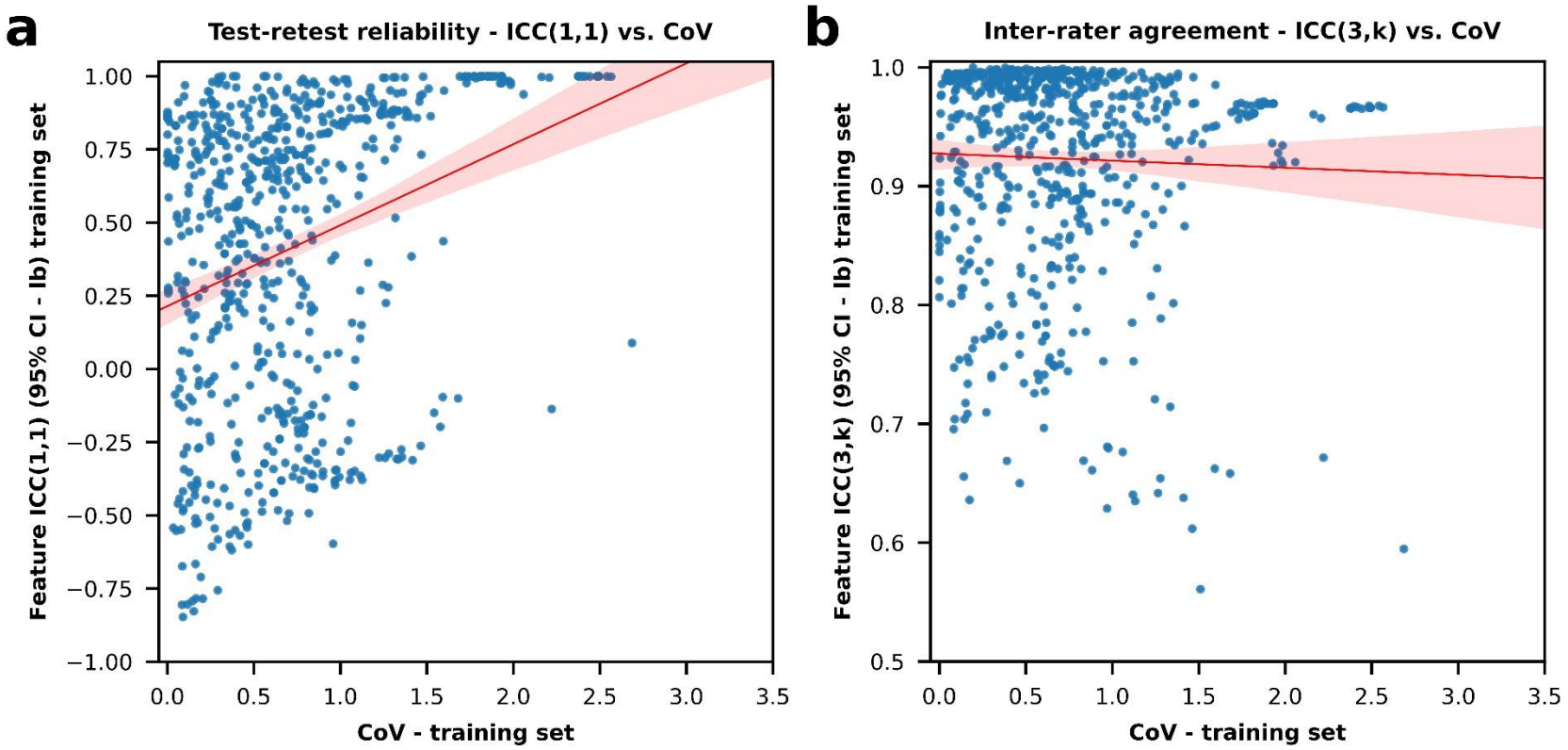

Figure 6 - Association of the coefficient of variation with feature stability
"Association of the feature coefficient of variation with (a) test-retest reliability (ICC(1,1)) and (b) inter-rater agreement (ICC(3,k)) in the training set, presented as scatterplots and linear regression line of best fit with 95 % confidence interval as shaded area. 95% CI – lb, lower bound of the 95% confidence interval; CoV, coefficient of variation; ICC, intraclass correlation coefficient." This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

4.5 Predictive performance of radiomics features and benchmark parameters

In the sequential forward aggregation for outcomes T2DM and MetS, the RF models comprising the top 6 and 8 radiomics features achieved the highest CV-sores, respectively (Fig. 8). These RF_{Rad} models were used for evaluation on the validation set and their constituent radiomics features are summarized in Table 7.

The RF_{RAD} models outperformed all benchmark RF models for both outcomes with respect to AUROC and $Accuracy_B$.

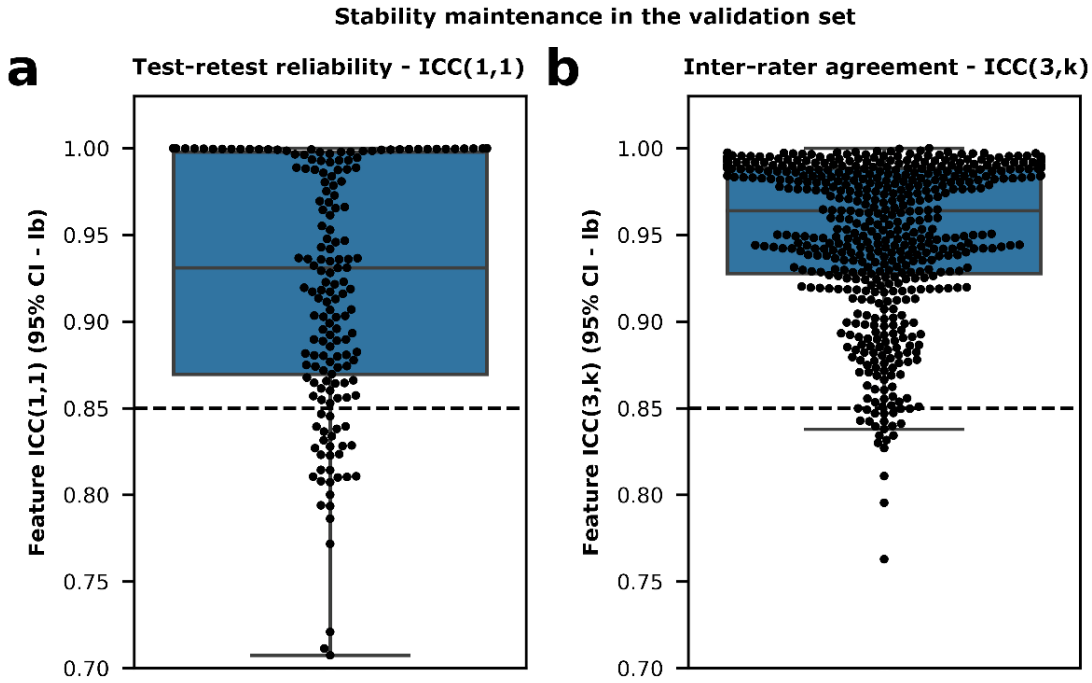

Figure 7 – Box-plots of feature stability in the validation set.
"Box-plots of (a) test-retest reliability (ICC(1,1)) and (b) inter-rater agreement (ICC(3,k)) on the validation set for all features that were stable (ICC ≥ 0.85) in the training set. Individual radiomics features are shown as black dots. The stability threshold (ICC ≥ 0.85) is indicated by dashed horizontal line. 95% CI – lb, lower bound of the 95% confidence interval; CoV, coefficient of variation; ICC, intraclass correlation coefficient." This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

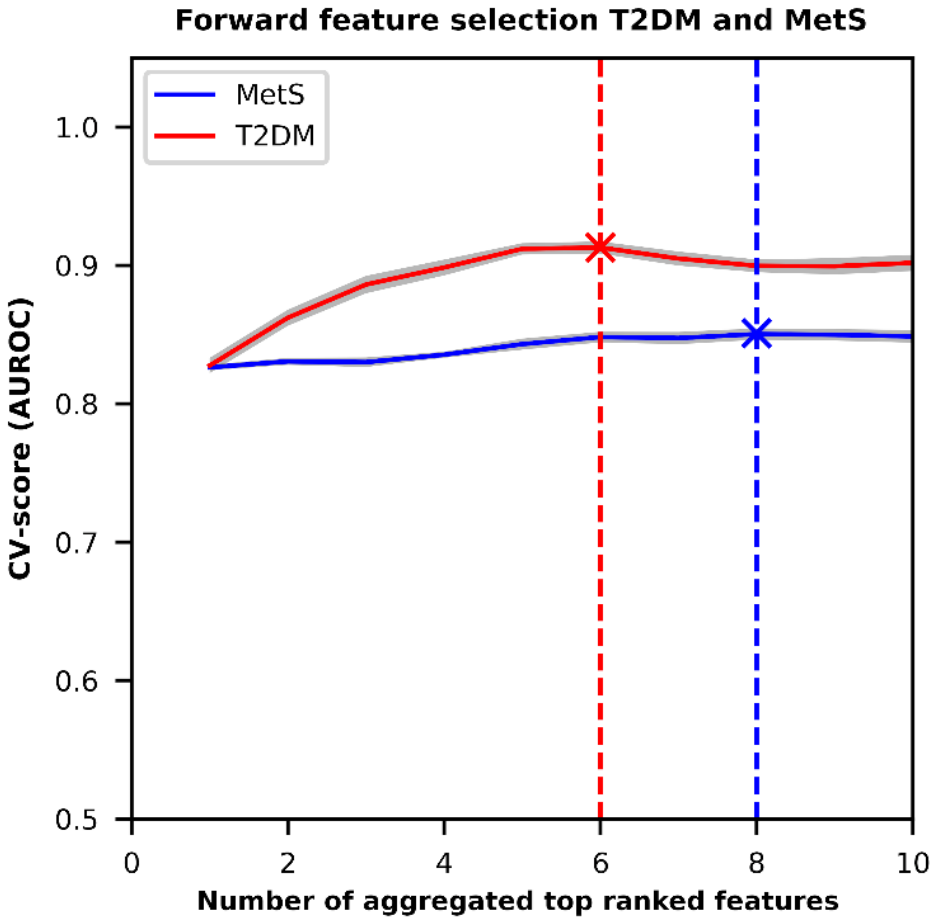

Figure 8 – Feature selection cross-validation scores
AUROC cross-validation scores for forward feature aggregation for the classification of type 2 diabetes mellitus (red) and metabolic syndrome (blue). The 95 % confidence interval of the AUROC is indicated by gray bands. *"Vertical lines mark the number of features with the highest cross-validation scores. AUROC, area under the curve of the receiver operating characteristic; CV, cross-validation; MetS, metabolic syndrome; T2DM, type 2 diabetes mellitus."* This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

The T2DM RF_{RAD} model achieved an AUROC of 0.835 (95 % CI: 0.832 - 0.838) (Gutmann et al., 2020) and $Accuracy_B$ of 0.822 (95 % CI: 0.819 - 0.824) (Gutmann et al., 2020). The benchmark RF models trained on hepatic PDFF (T1-MED: AUROC 0.795 [95 % CI: 0.789 - 0.801]; $Accuracy_B$ 0.747 [95 % CI: 0.747 - 0.747], MRS: AUROC 0.742 [95 % CI: 0.737 - 0.747]; $Accuracy_B$ 0.742 [95 % CI: 0.736 - 0.748]) (Gutmann et al., 2020) performed better than the BMI-based RF model (AUROC 0.716 [95 % CI: 0.714 - 0.718]; $Accuracy_B$ 0.518 [95 % CI: 0.517 - 0.520]) (Gutmann et al., 2020).

MetS was predicted by the RF_{RAD} model with an AUROC of 0.838 (95 % CI: 0.836 - 0.839) (Gutmann et al., 2020) and $Accuracy_B$ of 0.787 (95 % CI: 0.782 - 0.791) (Gutmann et al., 2020). The benchmark RF models for MetS followed a similar trend compared to T2DM with quantifications of hepatic fat (PDFF) outperforming (T1-MED: AUROC 0.824 [95 % CI: 0.822 - 0.826]; $Accuracy_B$ 0.750 [95 % CI: 0.735 - 0.765], MRS: AUROC 0.796 [0.794 - 0.798]; $Accuracy_B$ 0.750 [0.750 - 0.750]) (Gutmann et al., 2020) the BMI (0.780 [0.779 - 0.781]; $Accuracy_B$ 0.725 [0.721 - 0.729)]) (Gutmann et al., 2020) in predictive modelling.

The performance metrics of all RF models are summarized in Table 8. The receiver operating characteristic curves of all RF models are presented in Figure 9 a (T2DM) and b (MetS).

4.6 Stability of radiomics RF model features in the validation set

The stable feature selected for the predictive RF_{RAD} T2DM (n = 6) (Gutmann et al., 2020) and MetS (n = 8) (Gutmann et al., 2020) models were analyzed for both test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)) in the

validation set. For T2DM, 5 radiomics features remained stable (ICC ≥ 0.85) with one feature failing to meet the threshold (ICC ≥ 0.85) for test-retest reliability (ICC(1,1) = 0.79) (Gutmann et al., 2020). Similarly for MetS, one radiomics feature failed to satisfy the required test-retest reliability (ICC(1,1) = 0.82) (Gutmann et al., 2020). A pair plot of test-retest reliabilities (ICC(1,1)) and inter-rater agreements (ICC(3,k)) of RF_{RAD} features in the training and validation sets is provided in Figure 10 (compare Table 7).

Table 6 - Radiomics features regression analysis

"The variable associations were estimated using ordinary least squares regression models. β - coefficients are provided with the boundaries of the 95% CI in parenthesis. 95% CI, 95% confidence interval; CoV, coefficient of variation; ICC, intraclass correlation coefficient." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

Variables	*β*-coeff. (95% CI)	adjusted-R^2	*p*-value
ICC(1,1) training vs. validation	1.027 (1.015 - 1.040)	0.975	< 0.001
ICC(3,k) training vs. validation	1.072 (1.040 - 1.103)	0.869	< 0.001
CoV training vs. validation	0.855 (0.839 - 0.871)	0.943	< 0.001
ICC(1,1) training vs. CoV training	0.277 (0.211 - 0.342)	0.089	< 0.001
ICC(3,k) training vs. CoV training	-0.006 (-0.018 - 0.006)	0	0.329

Table 7 - Radiomics features of predictive random forest models

"Stable and discriminative features of RF_{RAD} models. Features names follow the standard PyRadiomics naming convention prefixed with gldbw, for brevity shortened to "Bw", as well as the prefix "original" indicating extraction without image pre-processing filters. 95% CI – lb, lower bound of the 95% confidence interval; gldbw, gray level discretization bin width; ICC, intraclass correlation coefficient; MetS, metabolic syndrome; RF_{RAD}, radiomics random forest; T2DM, type 2 diabetes mellitus. § Not stable in the validation set with ICC(1,1) 95% CI – lb 0.79. † Not stable in the validation set with ICC(1,1) 95% CI – lb 0.82." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

#	**Radiomics features of RF_{RAD} (T2DM)**
1	Bw_16_original_firstorder_RootMeanSquared
2	Bw_64_original_glrlm_LongRunHighGrayLevelEmphasis
3	Bw_192_original_gldm_LargeDependenceHighGrayLevelEmphasis§
4	Bw_640_original_glcm_JointEntropy
5	Bw_64_original_glszm_SmallAreaHighGrayLevelEmphasis
6	Bw_640_original_glcm_SumEntropy
#	**Radiomics features of RF_{RAD} (MetS)**
1	Bw_64_original_glcm_SumEntropy
2	Bw_32_original_glcm_SumEntropy
3	Bw_16_original_glcm_SumEntropy
4	Bw_16_original_firstorder_InterquartileRange
5	Bw_384_original_glcm_JointEnergy
6	Bw_384_original_glcm_MaximumProbability
7	Bw_384_original_glcm_SumEntropy
8	Bw_64_original_gldm_DependenceEntropy†

.

Table 8 – Random forest model performance for T2DM and MetS

"Random forest performance metrics for outcomes type 2 diabetes mellitus and the metabolic syndrome are reported as AUROC and Accuracy$_B$. All metrics are shown as mean and the boundaries of the 95 % confidence intervals in parenthesis. 95% CI, 95% confidence interval; Accuracy$_B$, balanced accuracy; AUROC, area under the curve of the receiver operating characteristic; MetS, metabolic syndrome; RF, random forest; RF$_{RAD}$, radiomics random forest; T2DM, type 2 diabetes mellitus. $ cross-validation iterations converge on identical classification solutions." This table has previously been published in Academic Radiology (Gutmann, et al., 2020).

	T2DM		MetS	
Predictors	**AUROC (95% CI)**	**Accuracy$_B$ (95% CI)**	**AUROC (95% CI)**	**Accuracy$_B$ (95% CI)**
Radiomic features	0.835 (0.832-0.838)	0.822 (0.819-0.824)	0.838 (0.836-0.839)	0.787 (0.782-0.791)
Benchmarks:				
body mass index	0.716 (0.714-0.718)	0.518 (0.517-0.520)	0.780 (0.779-0.781)	0.725 (0.721-0.729)
T1-weighted multi-echo Dixon (PDFF)	0.795 (0.789-0.801)	0.747 (0.747-0.747)$^{\$}$	0.824 (0.822-0.826)	0.750 (0.735-0.765)
^{1}H-magnetic resonance spectroscopy (PDFF)	0.742 (0.737-0.747)	0.742 (0.736-0.748)	0.796 (0.794-0.798)	0.750 (0.750-0.750)$^{\$}$

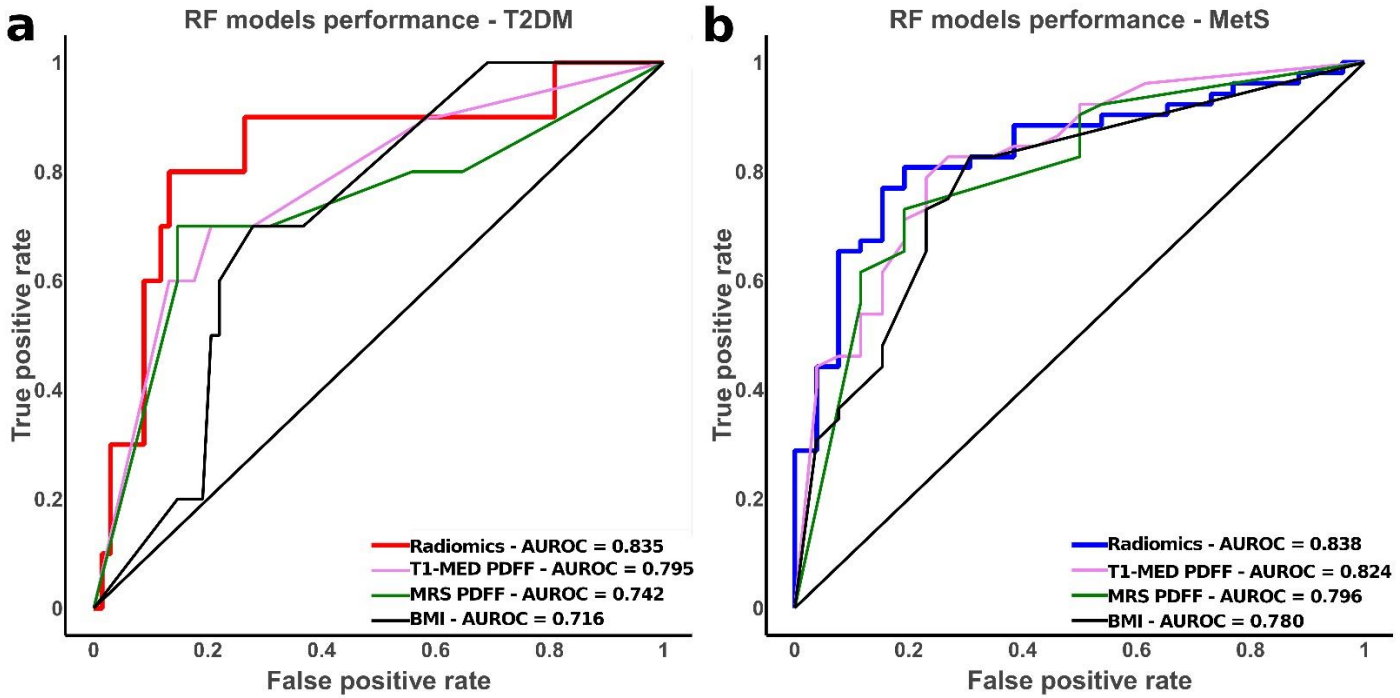

Figure 9 – Random forest receiver operating characteristic curves

Receiver operating characteristic curves of predictive random forest models on the validation set for (a) type 2 diabetes mellitus, (b) the metabolic syndrome by radiomics features and the benchmark parameters. *"BMI, body mass index; T1-MED, T1-weighted multi-echo Dixon; MetS, metabolic syndrome; MRS, ^{1}H-magnetic resonance spectroscopy; PDFF, proton density fat fraction; T2DM, type 2 diabetes mellitus."* This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

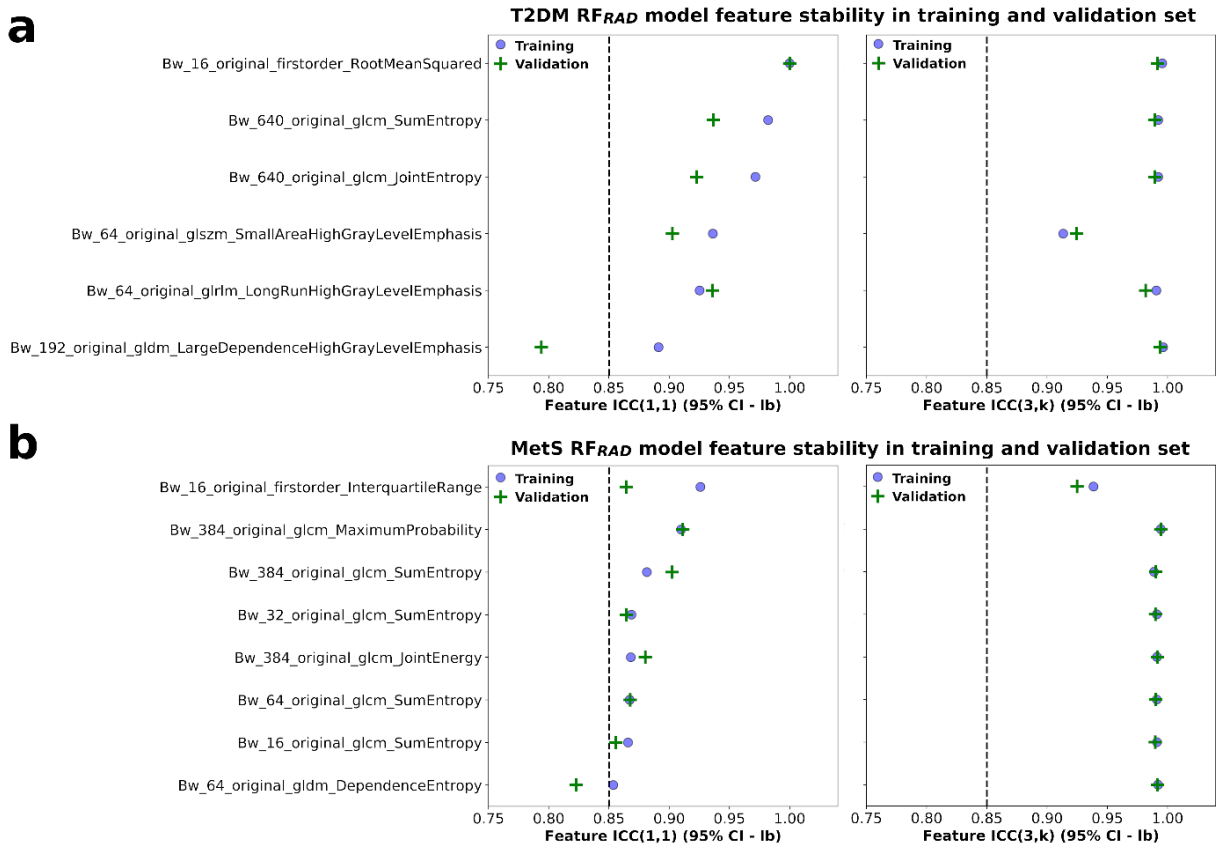

Figure 10 - Pair plots of radiomic feature stability of in training and validation set. *"Stability of RF$_{RAD}$ model features for (a) type 2 diabetes mellitus and (b) the metabolic syndrome in the training and validation sets. The stability threshold (ICC ≥ 0.85) is shown as dashed line. Feature names correspond to standard PyRadiomics naming convention prefixed with the gldbw, for brevity shortened to "Bw", as well as the prefix "original" indicating extraction without image pre-processing filters. 95% CI – lb, lower bound of the 95% confidence interval; gldbw, grey level discretization bin width; ICC, intraclass correlation coefficient; MetS, metabolic syndrome; RF$_{RAD}$, radiomics random forest; T2DM, type 2 diabetes mellitus."* This figure has previously been published in Academic Radiology (Gutmann, et al. 2020).

5 Discussion

This study set out to explore the prospect of utilizing radiomics features, extracted from manually segmented T1-DED volumetric liver rfwc maps, to predict T2DM and MetS in individuals from a representative cross-sectional cohort. In outline, radiomics features were evaluated for stability by exploiting computationally simulated test-retest and inter-rater scenarios. Analysis revealed that radiomics features were more susceptible to test-retest than inter-rater variance (Gutmann et al., 2020). In general, stability observations on the training and validation set were strongly correlated (Gutmann et al., 2020). Features that were stable in the training set (ICC(1,1) and ICC(3,k) ≥ 0.85) were further down-sampled and used to train predictive RF_{RAD} models for both T2DM and MetS (Gutmann et al., 2020). The RF_{RAD} models predicted T2DM and MetS with higher AUROC and $Accuracy_B$ than T1-MED and MRS derived hepatic PDFF as well as the BMI (Gutmann et al., 2020).

Participants of the KORA MRI substudy underwent an extensive imaging protocol that included three types of MRI imaging, namely the aforementioned T1-DED, T1-MED but also a T2-weighted half-Fourier singleshot turbo spin-echo (T2-HASTE) sequence that all provided a FOV with the desired full anatomical coverage of the liver (Bamberg et al., 2017). However, in order to compute rotationally invariant features, a prerequisite for reproducibility and comparison between datasets, image voxels dimensions need to be isotropic (Zwanenburg et al., 2016). Among the available data only the T1-DED images fulfilled this requirement without need for further processing steps. Since there is currently no consensus recommendation for an interpolation algorithm to be employed by radiomics studies (Zwanenburg et al., 2016), the anisotropic T1-MED and T2-HASTE sequence data was forgone in favor of T1-DED images.

Whilst the harnessed T1-DED images (Bashir et al., 2012, Hetterich et al., 2016) provide the required isotropic voxel dimension, the sequence lacks the necessary multiple echo trains to correct for T2 bias as well as T2* decay and is thus susceptible to hepatic iron overload. Furthermore, the manually contoured

liver masks also enclose small blood vessels that may confound the hepatic fat fraction quantified as rfwc. Therefore, as a quality control step for the analyzed subset (n = 310), the rfwc values from T1-DED images were compared with previous quantifications from T1-MED and MRS sequence data that adjusted for T2 bias as well as T2* decay, (Hetterich et al., 2016, Bamberg et al., 2017) and were evaluated in regions of interest within the liver parenchyme that were drawn avoiding blood vessels (Hetterich et al., 2016) to detect random differences. The hepatic fat fraction values (rfwc) were found to be highly correlated with T1-MED (R = 0.967) and MRS (R = 0.950) PDFF. T1-DED values were higher (both comparisons $p < 0.001$) and in accordance the Band-Altman analysis detected a small positive bias in the quantification, which agrees with the preceding in-depth investigation of a smaller KORA MRI subset (n = 215), that included T1-DED rfwc quantifications from vessel free parenchyme (Hetterich et al., 2016). Importantly, T1-DED quantifications were not-significantly different from T1-MED (p = 0.107) and MRS (p = 0.105) after linear – i.e. non-random - calibration. In summary, there is no evidence that hepatic iron overload or the inclusion of small blood vessels did have an appreciable random effect on the volumetric liver rfwc maps used for the ensuing radiomics analysis.

Next, a defined set of 684 radiomics features was computed from the volumetric liver rfwc maps of each participant (n = 310) and subjected to stability analysis. Of those, 171 features (25.0 %) were found to be stable in both simulated test-retest and inter-rater scenarios (ICC(1,1) and ICC(3,k) ≥ 0.85), a proportion that is comparable to reports from similar CT (Zwanenburg et al., 2019) and MRI based (Bologna et al., 2018, Schwier et al., 2019) studies.

A basic requirement of any feature selection method, including the proposed approach, is the generalization of observed feature stabilities from the training to the independent validation set. In an analogous diagnostic setting, stable features could therefore be expected to report consistently on disease state in hitherto unseen patients. Accordingly, the stability selection was carried out on the training set, thus allowing for the desired independent evaluation on the validation set and comparison of observed feature stabilities in between these two sets. Regression models show strong, non-random ($p < 0.001$) associations between

feature stabilities in the training and validation sets near unity (*β* 1.0) and with excellent goodness-of-fit for both test-retest reliabilities (*β* 1.027 [1.015 – 1.040], adjusted-R^2 0.975) (Gutmann et al., 2020) and inter-rater agreements (*β* 1.072 [1.040 – 1.103], adjusted-R^2 0.869) (Gutmann et al., 2020). Of 171 features that were stable in the training set and used for subsequent predictive modelling, 142 (83.0 %) (Gutmann et al., 2020) also exceeded the ICC threshold (≥ 0.85) for test-retest reliability and inter-rater agreement in the validation set. The remaining 29 (17.0 %) features still achieved stabilities of ICC > 0.70 (Gutmann et al., 2020), which is considered a moderate reliability by clinical research standards (Portney and Watkins, 2009). As observed for stabilities, the COV of radiomics features from training and validation sets displayed a strong association with excellent goodness-of-fit (*β* 0.943, adjusted-R^2 0.943) (Gutmann et al., 2020).

This evidence outlined above supports the basic assumption that stabilities observed in the training set as well as the COV, a measure of dispersion and potential information content generalize to the validation set.

Further analysis on training set data revealed that the COV was positively associated with the test-retest reliability, albeit weakly (*β* 0.277, adjusted-R^2 0.089, $p < 0.001$) (Gutmann et al., 2020), whereas the COV and the inter-rater agreement were not associated ($p < 0.329$) (Gutmann et al., 2020). At the same time, the breakdown of the selection process demonstrated that feature stability was far more susceptible to test-retest (26.5 % stable with ICC(1,1) ≥ 0.85) (Gutmann et al., 2020) than inter-rater variance (82.9 % stable with ICC(3,k) ≥ 0.85) (Gutmann et al., 2020). A possible explanation is that image noise comprising test-retest variance can subtly affect textures across the entire liver parenchyme resulting in more noticeable changes in the calculated features. The variance in the volumetric liver segmentation between readers that was approximated by deformed VOIs with mean dice coefficient of 0.89 ± 0.08 (Gutmann et al., 2020) leaves most liver parenchyme unaltered for feature calculation. A recent study dedicated to quantifying the inter-rater effect on radiomics features stability in several large public CT tumor segmentation datasets (dice coefficients 0.85 to 0.87) found that between 84 % and 88 % of

features were stable at ICC > 0.8 (Haarburger et al., 2020), which is in good overall agreement with this present study.

The RF_{RAD} model presented for the T2DM outcome prediction performed better than the RF model trained on the benchmark parameters obtained from the same study sample. Likewise, the RF_{RAD} model was more effective than previously published models based on the MRI quantifications of hepatic PDFF in a Chinese cohort (Wang et al., 2019), VAT volume (Wang et al., 2019, Linge et al., 2019) or various combinations of these parameters including skeletal muscle fat infiltration (Linge et al., 2019). For the prediction of T2DM status, these studies reported AUROCs based on hepatic PDFF by T1-MED imaging of 0.79 (Wang et al., 2019) or composites thereof including volumetric adipose tissue parameters of 0.78 (Linge et al., 2019). Models trained on the BMI predicted T2DM status with AUROC of 0.73 (Linge et al., 2019). Since Linge et al. (2019) only reported the BMI based predictions for the genders separately, the value is the mean reported for male and female study participants. The evaluation of RF benchmark models in the current investigation also concluded that T1-MED quantified hepatic PDFF was superior to the BMI for T2DM prediction with AUROCs of 0.795 (Gutmann et al., 2020) and 0.716 (Gutmann et al., 2020), respectively. The agreement on relative utility and similarity in performance with these prior studies may indicate that models based on radiomics signatures may equally perform better in other cohorts. However, the participants in these studies were either somewhat younger 51.5 ± 8.6 years (Wang et al., 2019) or older 62.60 ± 7.51 years (Linge et al., 2019) than those analyzed here with mean age of 56.1 ± 9.3 years (Gutmann et al., 2020). Further differences relating to participant selection, epidemiological parameters but also predictive modelling could confound this conclusion. The study by Linge et al. (2019) used a random split of participants for training and validations without ensuring representative stratification of epidemiological parameters. Since Wang et al. (2019) eschewed validation on independent data, the AUROC on independent data is likely to be lower.

In the case of MetS, previous studies that investigated the relationships with hepatic fat fraction in terms of MRI quantified PDFF found statistically

significant associations (Ducluzeau et al., 2013) or increased risk for said outcome (Chen et al., 2020b). Unfortunately, the literature search could not identify any publications evaluating this relationship on cross-sectional cohorts in terms of predictive models for direct comparison of performance metrics. However, both the homeostasis model assessment of insulin resistance (HOMA-IR), a blood test used in the clinical diagnosis of insulin resistance and pancreatic beta cell function, as well as the BMI collected from participants of two large cross-sectional cohorts were evaluated in predictive models. In a subset of a Spanish general population study (EPIRCE) comprising 2459 adult participants the BMI model predicted MetS with an AUROC of 0.69. In addition, RF_{Rad} models were superior to the anthropometric benchmark parameter BMI in the prediction of T2DM and MetS (Gutmann et al., 2020).

There are some limitations to this investigation. Firstly, it is uncertain whether the amount of noise introduced to simulate the test-retest scenario was appropriate. The studies that were used in the meta-analysis of test-retest reliabilities did not adhere to identical protocols to quantify the hepatic fat fraction from image data (Yokoo et al., 2018). Differences in the liver regions and volume analyzed adds “inter-protocol” variance to the estimated noise (Yokoo et al., 2018). It should also be considered that some of the included data was collected before 2011 (Yokoo et al., 2018) and may be confounded by outdated image processing and reconstruction methods, not relevant to current studies. Future studies should take these effects into consideration to select optimal parameters since the stability of radiomic features was far more sensitive to test-retest than inter-rater variance to avoid either excessive or insufficient feature elimination.

The radiomic extraction process was designed in adherence to the Image Biomarker Standardization Initiative (IBSI) guidelines (Zwanenburg et al., 2016) for MRI image data. There is a clear and justified recommendation to compute radiomics features from isotropic and normalized image data, which was followed (Zwanenburg et al., 2016). In contrast, there is no guidance on the choice of optimal discretization bin width. An optimization of this parameter is necessary for each type of MRI sequence (Carre et al., 2020) and potentially also the tissue analyzed, although the latter point is conjecture. Thus, the optimal bin width is

usually selected empirically and/or under consideration of related work. In the present case, no previous study provided a precedent for MRI-derived hepatic fat fraction maps. Therefore, radiomics features were extracted over a wide range of discretization bin widths (gldbws) to increase the likelihood of obtaining a sufficient number of stable and informative features. Yet other methods that have shown to increase the number of stable features such as smoothing (Zwanenburg et al., 2019) and Wavelet (Schwier et al., 2019) filters have not been included in this study. Due to the combinatorial process of parameter search, the enormous number of resulting features over the wide range of discretizations would have been prohibitive. Future work could investigate the utility of these filters on a subset of features to improve their performance in radiomics models.

This investigation is further limited by the ethnic composition and by the nature of its study design. The cohort consists of Western-European individuals only, therefore it is uncertain whether the presented radiomics models are equally applicable to other ethnicities. Large cross-sectional studies have already shown clear evidence that age, gender and ethnicity confound the predictive power of anthropometric parameters for metabolic outcomes T2DM (Yoon et al., 2016, Hartwig et al., 2016) and MetS (Cheong et al., 2015). The influence of age, gender and ethnicity on radiomics models for these outcomes should be vigorously investigated using cohorts with a sufficiently large number of participants to draw conclusions with the necessary statistical confidence. A complete validation would also require a multi-center design to assess “non-simulated” versus the proposed simulated test-retest and inter-rater setup. Although such computational approaches for stability selection have been tested previously (Gevaert et al., 2014, Bologna et al., 2018, Zwanenburg et al., 2019), multi-center data would allow for potential adjustment of dice-coefficients and image noise in future applications of the proposed feature selection methodology. Furthermore, in order to establish radiomics features as a biomarker of cardiometabolic risk, an analysis and evaluation on studies with follow-up design would be required.

In summary, the proposed feature selection methodology has proven to be effective, at least in the context of imaging data of a single center study to select

stable radiomics features. In predictive models, stable radiomics features were effective biomarkers for the related CVD risk factors T2DM and MetS achieving higher AUROC and $Accuracy_B$ than MR-based quantifications of the hepatic fraction and the anthropometric obesity surrogate BMI. Therefore, radiomics features of hepatic fat fraction maps deserve consideration to be evaluated on larger and more diverse population cross-sections as well as longitudinal data to gauge their value in diagnostic and prognostic clinical settings. Eventually they could easily be combined with MRI whole body composition scores to develop personalized cardiometabolic risk models.

6 Summary

Objective: To assess MRI derived radiomics features of liver fat collected from a cohort of individuals without prior cardiovascular events as imaging biomarkers of type 2 diabetes mellitus (T2DM) and metabolic syndrome (MetS).

Material and methods: 400 participants of the KORA MRI substudy underwent comprehensive whole body MRI imaging protocols including T1-weighted dual-echo Dixon (T1-DED), T1-weighted multi-echo Dixon (T1-MED) and magnetic resonance spectroscopy (MRS) sequences. A total of 684 radiomics features were extracted on T1-DED relative fat water content (rfwc) maps of 310 artefact free manually contoured liver volumes of interest (VOI). The corresponding individuals (n = 310, T2DM 12.6 %, MetS 34.5 %) were assigned to stratified training (n = 232, 75 %) and validation (n = 78, 25 %) sets. To assess feature stability, test-retest and inter-rater variance was approximated by generating noise augmented rfwc maps and computationally deformed VOIs, respectively. Feature stability was assessed in terms of the intraclass correlation coefficient (ICC) for test-retest reliability (ICC(1,1)) and inter-rater agreement (ICC(3,k)) on training set data. Stable features (ICC ≥ 0.85) were assessed as imaging biomarkers of T2DM and MetS in random forest (RF) models. For benchmarking, RF models were trained on the participants' hepatic proton density fat fraction (PDFF) quantified previously on the T1-MED and MRS images as well as the body mass index (BMI). All RF models were evaluated on the validation set using the area under the curve of the receiver operating characteristic (AUROC) and the balanced accuracy ($Accuracy_B$) as performance metrics.

Results: The epidemiological characteristics in training and validation sets were not statistically significantly different (p < 0.001). Furthermore, training and validation sets showed strong associations for both test-retest reliability (β 1.027 [1.015 – 1.040]) and inter-rater agreement (β 1.072 [1.040 – 1.103]). A total of 171 features (25.0 %) met the stability threshold (ICC ≥ 0.85). In subsequent RF modelling radiomics features predicted T2DM with AUROC 0.835 and $Accuracy_B$

0.822 and MetS with AUROC 0.838 and Accuracy$_B$ 0.787, thereby outperforming all benchmark RF models in both metric categories.

Conclusion: In this single-center study radiomics features of MRI-derived hepatic fat were superior biomarkers of T2DM and MetS than hepatic PDFF and the BMI. In the future, hepatic radiomics features deserve further exploration and development as potential biomarkers in metabolic disease.

Deutsche Zusammenfassung

Ziel: Diese Arbeit untersucht radiomische Merkmale von Leberfett aus MRT Bilddaten einer Kohorte ohne vorherige kardiovaskuläre Ereignisse als bildgebende Biomarker für Diabetes mellitus Typ 2 (T2DM) und das metabolische Syndrom (MetS).

Material und Methoden: Die 400 Probanden der KORA-MRT Studie wurden einer umfassenden Ganzkörper-MRT Untersuchung, einschließlich T1-gewichtetem Doppelecho-Dixon (T1-DED), T1-gewichtetem Multiecho-Dixon (T1-MED) und einer Magnetresonanzspektroskopie (MRS) unterzogen. Insgesamt wurden jeweils 684 radiomische Merkmale aus T1-DED relativen Fettwassergehalts (rfwc)-Karten von 310 artefaktfreien manuell konturierten Lebervolumina (VOI) extrahiert. Die entsprechenden Probanden (n = 310, T2DM 12,6 %, MetS 34,5 %) wurden stratifizierten Trainings- (n = 232, 75 %) und Validierungsdatensätzen (n = 78, 25 %) zugeordnet. Um die radiomische Merkmalsstabilität zu beurteilen, wurde die Test-Retest- und Interrater-Varianz durch Generieren von rauschverstärkten rfwc-Karten bzw. künstlich deformierten VOIs angenähert. Die Merkmalsstabilität wurde als Intraclass-Korrelationskoeffizient (ICC) für die Test-Retest-Zuverlässigkeit (ICC (1,1)) und die Interrater-Reliabilität (ICC (3, k)) anhand des Trainingsdatensatzes bewertet. Stabile Merkmale (ICC ≥ 0.85) wurden als bildgebende Biomarker für T2DM und

MetS mittels Random Forest (RF) Modellen beurteilt. Für ein Benchmarking wurden zudem RF Modelle basierend auf der hepatischen Protonendichte-Fettfraktion (PDFF), die zuvor in T1-MED- und MRS-Bilddaten der Probanden quantifiziert worden war, sowie ihrem Body Mass Index (BMI) erstellt. Alle RF Modelle wurden am Validierungsdatensatz unter Verwendung des Integrals der Empfänger-Opertationscharakteristik (AUROC) sowie der gewichteten Genauigkeit ($Accuracy_B$) als Leistungsmetrik evaluiert.

Ergebnisse: Die epidemiologischen Eigenschaften in Trainings- und Validierungsdatensätzen waren statistisch nicht signifikant unterschiedlich (p < 0,001). Darüber hinaus zeigten Trainings- und Testdaten starke Assoziationen sowohl für die Test-Retest-Zuverlässigkeit (β 1,027 [1,015 – 1,040]) als auch für die Interrater-Reliabilität (β 1,072 [1,040 – 1,103]). Insgesamt 171 Merkmale (25,0 %) erfüllten die Stabilitätsschwelle (ICC ≥ 0,85). In der abschließenden radiomischen RF-Modellierung wurden T2DM mit einem AUROC von 0,835 und $Accuracy_B$ von 0,822 sowie MetS mit einem AUROC von 0,838 und $Accuracy_B$ von 0,787 vorhergesagt, welche damit alle Benchmark RF Modelle in den jeweiligen Metrikkategorien übertraf.

Schlussfolgerung: In dieser Single-Center-Studie waren radiomische Merkmale von MRT-Bilddaten des Leberfettes als Biomarker für T2DM und MetS sowohl dem hepatischen PDFF als auch dem BMI überlegen. In Zukunft verdienen radiomische Merkmale der Leber eine weitere Erforschung und Entwicklung als potenzielle Biomarker bei Stoffwechselerkrankungen.

7 Bibliography

Acar, B, Ozeke, O, Karakurt, M, Ozen, Y, Ozbay, MB, Unal, S, Karanfil, M, Yayla, C, Cay, S, Maden, O, Topaloglu, S, Aras, D, et al. 2019. Association of Prediabetes With Higher Coronary Atherosclerotic Burden Among Patients With First Diagnosed Acute Coronary Syndrome. *Angiology,* 70**,** 174-180.

Aerts, HJ, Velazquez, ER, Leijenaar, RT, Parmar, C, Grossmann, P, Carvalho, S, Bussink, J, Monshouwer, R, Haibe-Kains, B, Rietveld, D, Hoebers, F, Rietbergen, MM, et al. 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun,* 5**,** 4006.

Aguilar, M, Bhuket, T, Torres, S, Liu, B & Wong, RJ 2015. Prevalence of the Metabolic Syndrome in the United States, 2003-2012. *JAMA,* 313**,** 1973–1974.

Alberti, KG, Zimmet, P & Shaw, J 2006. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med,* 23**,** 469-80.

Alberti, KG, Zimmet, P, Shaw, J & Group, IDFETFC 2005. The metabolic syndrome--a new worldwide definition. *Lancet,* 366**,** 1059-62.

ALLHAT Collaborative Research Group. The Antihypertensive Lipid-Lowering Treatment to Prevent Heart Attack trial 2002. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA,* 288**,** 2981-97.

Anavi, S, Madar, Z & Tirosh, O 2017. Non-alcoholic fatty liver disease, to struggle with the strangle: Oxygen availability in fatty livers. *Redox Biol,* 13**,** 386-392.

American Diabetes Association 2020. 3. Prevention or Delay of Type 2 Diabetes: Standards of Medical Care in Diabetes—2020. *Diabetes Care,* 43**,** S32-S36.

Ballestri, S, Zona, S, Targher, G, Romagnoli, D, Baldelli, E, Nascimbeni, F, Roverato, A, Guaraldi, G & Lonardo, A 2016. Nonalcoholic fatty liver disease is associated with an almost twofold increased risk of incident type 2 diabetes and metabolic syndrome. Evidence from a systematic review and meta-analysis. *J Gastroenterol Hepatol,* 31**,** 936-44.

Bamberg, F, Hetterich, H, Rospleszcz, S, Lorbeer, R, Auweter, SD, Schlett, CL, Schafnitzel, A, Bayerl, C, Schindler, A, Saam, T, Muller-Peltzer, K, Sommer, W, et al. 2017. Subclinical Disease Burden as Assessed by Whole-Body MRI in Subjects With Prediabetes, Subjects With Diabetes, and Normal Control Subjects From the General Population: The KORA-MRI Study. *Diabetes,* 66**,** 158-169.

Barter, P & Genest, J 2019. HDL cholesterol and ASCVD risk stratification: A debate. *Atherosclerosis,* 283**,** 7-12.

Bashir, MR, Merkle, EM, Smith, AD & Boll, DT 2012. Hepatic MR imaging for in vivo differentiation of steatosis, iron deposition and combined storage

disorder: single-ratio in/opposed phase analysis vs. dual-ratio Dixon discrimination. *Eur J Radiol,* 81**,** e101-9.

Bashir, MR, Zhong, X, Dale, BM, Gupta, RT, Boll, DT & Merkle, EM 2013. Automated patient-tailored screening of the liver for diffuse steatosis and iron overload using MRI. *AJR Am J Roentgenol,* 201**,** 583-8.

Boden-Albala, B, Sacco, RL, Lee, HS, Grahame-Clarke, C, Rundek, T, Elkind, MV, Wright, C, Giardina, EG, Ditullio, MR, Homma, S & Paik, MC 2008. Metabolic syndrome and ischemic stroke risk: Northern Manhattan Study. *Stroke,* 39**,** 30-5.

Bohte, AE, Van Werven, JR, Bipat, S & Stoker, J 2011. The diagnostic accuracy of US, CT, MRI and 1H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. *Eur Radiol,* 21**,** 87-97.

Bologna, M, Corino, VDA, Montin, E, Messina, A, Calareso, G, Greco, FG, Sdao, S & Mainardi, LT 2018. Assessment of Stability and Discrimination Capacity of Radiomic Features on Apparent Diffusion Coefficient Images. *J Digit Imaging,* 31**,** 879-894.

Bonekamp, S, Tang, A, Mashhood, A, Wolfson, T, Changchien, C, Middleton, MS, Clark, L, Gamst, A, Loomba, R & Sirlin, CB 2014. Spatial distribution of MRI-Determined hepatic proton density fat fraction in adults with nonalcoholic fatty liver disease. *J Magn Reson Imaging,* 39**,** 1525-32.

Bulut, A & Avci, B 2019. Carotid intima-media thickness values are significantly higher in patients with prediabetes compared to normal glucose metabolism. *Medicine (Baltimore),* 98**,** e17805.

Bydder, M, Yokoo, T, Hamilton, G, Middleton, MS, Chavez, AD, Schwimmer, JB, Lavine, JE & Sirlin, CB 2008. Relaxation effects in the quantification of fat using gradient echo imaging. *Magn Reson Imaging,* 26**,** 347-59.

Cai, J, Zhang, XJ, Ji, YX, Zhang, P, She, ZG & Li, H 2020. Nonalcoholic Fatty Liver Disease Pandemic Fuels the Upsurge in Cardiovascular Diseases. *Circ Res,* 126**,** 679-704.

Cameron, AJ, Magliano, DJ, Zimmet, PZ, Welborn, TA, Colagiuri, S, Tonkin, AM & Shaw, JE 2008. The metabolic syndrome as a tool for predicting future diabetes: the AusDiab study. *J Intern Med,* 264**,** 177-86.

Cameron, AJ, Shaw, JE & Zimmet, PZ 2004. The metabolic syndrome: prevalence in worldwide populations. *Endocrinol. Metab. Clin. North Am.,* 33**,** 351–375.

Carre, A, Klausner, G, Edjlali, M, Lerousseau, M, Briend-Diop, J, Sun, R, Ammari, S, Reuze, S, Alvarez Andres, E, Estienne, T, Niyoteka, S, Battistella, E, et al. 2020. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep,* 10**,** 12340.

Castelli, WP 1992. Epidemiology of triglycerides: a view from Framingham. *Am J Cardiol,* 70**,** 3H-9H.

Chalasani, N, Younossi, Z, Lavine, JE, Charlton, M, Cusi, K, Rinella, M, Harrison, SA, Brunt, EM & Sanyal, AJ 2018. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology,* 67**,** 328-357.

Chawla, NB, K. W.; Hall L. O.; Kegelmeyer W. P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res***,** 321-357.

Chen, H, Zheng, X, Zong, X, Li, Z, Li, N, Hur, J, Fritz, CD, Chapman, W, Jr., Nickel, KB, Tipping, A, Colditz, GA, Giovannucci, EL, et al. 2020a. Metabolic syndrome, metabolic comorbid conditions and risk of early-onset colorectal cancer. *Gut* [Online]. Available: 10.1136/gutjnl-2020-321661

Chen, J, Duan, S, Ma, J, Wang, R, Chen, J, Liu, X, Xue, L, Xie, S & Yao, S 2020b. MRI-determined liver fat correlates with risk of metabolic syndrome in patients with nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol,* 32**,** 754-761.

Chen, J, Gu, D, Chen, CS, Wu, X, Hamm, LL, Muntner, P, Batuman, V, Lee, CH, Whelton, PK & He, J 2007. Association between the metabolic syndrome and chronic kidney disease in Chinese adults. *Nephrol Dial Transplant,* 22**,** 1100-6.

Cheong, KC, Ghazali, SM, Hock, LK, Subenthiran, S, Huey, TC, Kuay, LK, Mustapha, FI, Yusoff, AF & Mustafa, AN 2015. The discriminative ability of waist circumference, body mass index and waist-to-hip ratio in identifying metabolic syndrome: Variations by age, sex and race. *Diabetes Metab Syndr,* 9**,** 74-8.

Collins, R, Reith, C, Emberson, J, Armitage, J, Baigent, C, Blackwell, L, Blumenthal, R, Danesh, J, Smith, GD, Demets, D, Evans, S, Law, M, et al. 2016. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet,* 388**,** 2532-2561.

Davies, MJ, D'alessio, DA, Fradkin, J, Kernan, WN, Mathieu, C, Mingrone, G, Rossing, P, Tsapas, A, Wexler, DJ & Buse, JB 2018. Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia,* 61**,** 2461-2498.

Dibaba, DT, Braithwaite, D & Akinyemiju, T 2018. Metabolic Syndrome and the Risk of Breast Cancer and Subtypes by Race, Menopause and BMI. *Cancers (Basel),* 10.

Dixon, WT 1984. Simple proton spectroscopic imaging. *Radiology,* 153**,** 189-94.

Dobbs, R, Sawers, C, Thompson, F, Manyika, J, Woetzel, J, Child, P, Mckenna, S & Spatharou, A. 2014. Overcoming obesity: An initial economic analysis. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Economic%20Studies%20TEMP/Our%20Insights/How%20the%20world%20could%20better%20fight%20obesity/MGI_Overcoming_obesity_Executive_summary.pdf [Accessed 01/10/2020]

Ducluzeau, PH, Boursier, J, Bertrais, S, Dubois, S, Gauthier, A, Rohmer, V, Gagnadoux, F, Leftheriotis, G, Cales, P, Andriantsitohaina, R, Roullier, V & Aube, C 2013. MRI measurement of liver fat content predicts the metabolic syndrome. *Diabetes Metab,* 39**,** 314-21.

Eckel, RH, Grundy, SM & Zimmet, PZ 2005. The metabolic syndrome. *Lancet,* 365**,** 1415-28.

Einarson, TR, Acs, A, Ludwig, C & Panton, UH 2018. Economic Burden of Cardiovascular Disease in Type 2 Diabetes: A Systematic Review. *Value Health,* 21**,** 881-890.

Esposito, K, Chiodini, P, Colao, A, Lenzi, A & Giugliano, D 2012. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. *Diabetes Care,* 35**,** 2402-11.

Expert Panel on Detection, E & Adults, TOHBCI 2001. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA,* 285**,** 2486-2497.

Feng, Y, Hong, X, Li, Z, Zhang, W, Jin, D, Liu, X, Zhang, Y, Hu, FB, Wei, L-J, Zang, T, Xu, X & Xu, X 2006. Prevalence of metabolic syndrome and its relation to body composition in a Chinese rural population. *Obesity (Silver Spring).* 14**,** 2089–2098.

Fernández-Bergés, D, Cabrera De León, A, Sanz, H, Elosua, R, Guembe, MJ, Alzamora, M, Vega-Alonso, T, Félix-Redondo, FJ, Ortiz-Marrón, H, Rigo, F, Lama, C, Gavrila, D, et al. 2012. Metabolic Syndrome in Spain: Prevalence and Coronary Risk Associated With Harmonized Definition and WHO Proposal. DARIOS Study. *Rev. Esp. Cardiol.,* 65**,** 241–248.

Fishbein, MH, Gardner, KG, Potter, CJ, Schmalbrock, P & Smith, MA 1997. Introduction of fast MR imaging in the assessment of hepatic steatosis. *Magn Reson Imaging,* 15**,** 287-93.

Ford, ES 2005a. Prevalence of the metabolic syndrome defined by the International Diabetes Federation among adults in the U.S. *Diabetes Care,* 28**,** 2745-9.

Ford, ES 2005b. Risks for all-cause mortality, cardiovascular disease, and diabetes associated with the metabolic syndrome: a summary of the evidence. *Diabetes Care,* 28**,** 1769-78.

Forstermann, U & Sessa, WC 2012. Nitric oxide synthases: regulation and function. *Eur Heart J,* 33**,** 829-37.

Fried, M, Yumuk, V, Oppert, JM, Scopinaro, N, Torres, AJ, Weiner, R, Yashkov, Y, Fruhbeck, G, European Association for the Study Of, O & International Federation for the Surgery of Obesity - European, C 2013. Interdisciplinary European Guidelines on metabolic and bariatric surgery. *Obes Facts,* 6**,** 449-68.

Fruchart, JC, Nierman, MC, Stroes, ES, Kastelein, JJ & Duriez, P 2004. New risk factors for atherosclerosis and patient risk assessment. *Circulation,* 109**,** III15-9.

Fukuda, T, Hamaguchi, M, Kojima, T, Mitsuhashi, K, Hashimoto, Y, Ohbora, A, Kato, T, Nakamura, N & Fukui, M 2016. Transient remission of nonalcoholic fatty liver disease decreases the risk of incident type 2 diabetes mellitus in Japanese men. *Eur J Gastroenterol Hepatol,* 28**,** 1443-1449.

Gatenby, RA, Grove, O & Gillies, RJ 2013. Quantitative imaging in cancer evolution and ecology. *Radiology,* 269**,** 8-15.

Gerich, JE 1993. Control of glycaemia. *Baillieres Clin Endocrinol Metab,* 7**,** 551-86.

Gevaert, O, Mitchell, LA, Achrol, AS, Xu, J, Echegaray, S, Steinberg, GK, Cheshier, SH, Napel, S, Zaharchuk, G & Plevritis, SK 2014. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology,* 273**,** 168-74.

Gillies, RJ, Anderson, AR, Gatenby, RA & Morse, DL 2010. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol,* 65**,** 517-21.

Ginsberg, HN, Zhang, YL & Hernandez-Ono, A 2005. Regulation of plasma triglycerides in insulin resistance and diabetes. *Arch Med Res,* 36**,** 232-40.

Goit, LN & Yang, S 2019. Treatment of Hypertension: A Review. *Yangtze Medicine,* 3**,** 101–123.

Gray, A, Raikou, M, Mcguire, A, Fenn, P, Stevens, R, Cull, C, Stratton, I, Adler, A, Holman, R & Turner, R 2000. Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside randomised controlled trial (UKPDS 41). United Kingdom Prospective Diabetes Study Group. *BMJ,* 320**,** 1373-8.

Grundy, SM 2008. Metabolic Syndrome Pandemic. *Arterioscler Thromb Vasc Biol,* 28**,** 629-36.

Gutmann, DAP, Rospleszcz, S, Rathmann, W, Schlett, CL, Peters, A, Wachinger, C, Gatidis, S & Bamberg, F. 2020. MRI-Derived Radiomics Features of Hepatic Fat Predict Metabolic States in Individuals without Cardiovascular Disease. *Acad Radiol* [Online]. Available: 10.1016/j.acra.2020.06.030

Haarburger, C, Muller-Franzes, G, Weninger, L, Kuhl, C, Truhn, D & Merhof, D 2020. Radiomics feature reproducibility under inter-rater variability in segmentations of CT images. *Sci Rep,* 10**,** 12688.

Haller H. & M., H 1975. Synoptische Betrachtung metabolischer Risikofaktoren. *In:* Haller, H, Hanefeld, M & Jaross, W (eds.) *Lipidstoffwechselstörungen.* Jena: Gustav Fischer Verlag.

Hanefeld, M & Leonhardt, W 1981. Das metabolische Syndrom. *Dtsch Gesundheitwes,* 36**,** 545-51.

Hartwig, S, Kluttig, A, Tiller, D, Fricke, J, Muller, G, Schipf, S, Volzke, H, Schunk, M, Meisinger, C, Schienkiewitz, A, Heidemann, C, Moebus, S, et al. 2016. Anthropometric markers and their association with incident type 2 diabetes mellitus: which marker is best for prediction? Pooled analysis of four German population-based cohort studies and comparison with a nationwide cohort study. *BMJ Open* [Online], 6. Available: 10.1136/bmjopen-2015-009266

Haslam, DW & James, WP 2005. Obesity. *Lancet,* 366**,** 1197-209.

Hegab, Z, Gibbons, S, Neyses, L & Mamas, MA 2012. Role of advanced glycation end products in cardiovascular disease. *World J Cardiol,* 4**,** 90-102.

Hetterich, H, Bayerl, C, Peters, A, Heier, M, Linkohr, B, Meisinger, C, Auweter, S, Kannengiesser, SA, Kramer, H, Ertl-Wagner, B & Bamberg, F 2016. Feasibility of a three-step magnetic resonance imaging approach for the assessment of hepatic steatosis in an asymptomatic study population. *Eur Radiol,* 26**,** 1895-904.

Hirode, G & Wong, RJ 2020. Trends in the Prevalence of Metabolic Syndrome in the United States, 2011-2016. *JAMA,* 323**,** 2526–2528.

Ho, JS, Cannaday, JJ, Barlow, CE, Mitchell, TL, Cooper, KH & Fitzgerald, SJ 2008. Relation of the number of metabolic syndrome risk factors with all-cause and cardiovascular mortality. *Am J Cardiol,* 102**,** 689-92.

Hokanson, JE & Austin, MA 1996. Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol

level: a meta-analysis of population-based prospective studies. *J Cardiovasc Risk,* 3**,** 213-9.

Holle, R, Happich, M, Lowel, H, Wichmann, HE & Group, MKS 2005. KORA--a research platform for population based health research. *Gesundheitswesen,* 67 Suppl 1**,** S19-25.

Hood, L & Friend, SH 2011. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol,* 8**,** 184-7.

Huang, J, Huang, JL, Withers, M, Chien, K-L, Trihandini, I, Elcarte, E, Chung, V & Wong, MC 2018. Prevalence of metabolic syndrome in Chinese women and men: a systematic review and meta-analysis of data from 734 511 individuals. *Lancet,* 392**,** S14.

Hunter, JD 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering,* 9**,** 90–95.

Ishizaka, N, Ishizaka, Y, Toda, E-I, Hashimoto, H, Nagai, R & Yamakado, M 2005. Hypertension Is the Most Common Component of Metabolic Syndrome and the Greatest Contributor to Carotid Arteriosclerosis in Apparently Healthy Japanese Individuals. *Hypertens. Res.,* 28**,** 27–34.

Kaddoura, R, Orabi, B & Salam, AM 2020. Efficacy and safety of PCSK9 monoclonal antibodies: an evidence-based review and update. *J Drug Assess,* 9**,** 129-144.

Kaess, BM, Pedley, A, Massaro, JM, Murabito, J, Hoffmann, U & Fox, CS 2012. The ratio of visceral to subcutaneous fat, a metric of body fat distribution, is a unique correlate of cardiometabolic risk. *Diabetologia,* 55**,** 2622-2630.

Katsimardou, A, Imprialos, K, Stavropoulos, K, Sachinidis, A, Doumas, M & Athyros, V 2020. Hypertension in Metabolic Syndrome: Novel Insights. *Curr Hypertens Rev,* 16**,** 12-18.

Khan, MAB, Hashim, MJ, King, JK, Govender, RD, Mustafa, H & Al Kaabi, J 2020. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J Epidemiol Glob Health,* 10**,** 107-111.

Kim, H, Taksali, SE, Dufour, S, Befroy, D, Goodman, TR, Petersen, KF, Shulman, GI, Caprio, S & Constable, RT 2008. Comparative MR study of hepatic fat quantification using single-voxel proton spectroscopy, two-point dixon and three-point IDEAL. *Magn Reson Med,* 59**,** 521-7.

Konig, M, Bulik, S & Holzhutter, HG. 2012. Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism. *PLoS Comput Biol* [Online], 8. Available: 10.1371/journal.pcbi.1002577

Kumar, V, Gu, Y, Basu, S, Berglund, A, Eschrich, SA, Schabath, MB, Forster, K, Aerts, HJ, Dekker, A, Fenstermacher, D, Goldgof, DB, Hall, LO, et al. 2012. Radiomics: the process and the challenges. *Magn Reson Imaging,* 30**,** 1234-48.

Kwon, H, Kim, D & Kim, JS 2017. Body Fat Distribution and the Risk of Incident Metabolic Syndrome: A Longitudinal Cohort Study. *Sci Rep,* 7**,** 10955.

Kylin, E 1923. Studien Hypertonie-Hyperglykämie-Hyperurikämiesyndrome. *Zentralbl f innere Med Leipz,* 81**,** 105-127.

Lambin, P, Leijenaar, RTH, Deist, TM, Peerlings, J, De Jong, EEC, Van Timmeren, J, Sanduleanu, S, Larue, RTHM, Even, AJG, Jochems, A, Van Wijk, Y, Woodruff, H, et al. 2017. Radiomics: the bridge between medical

imaging and personalized medicine. *Nature Reviews Clinical Oncology,* 14**,** 749.

Lambin, P, Rios-Velazquez, E, Leijenaar, R, Carvalho, S, Van Stiphout, RG, Granton, P, Zegers, CM, Gillies, R, Boellard, R, Dekker, A & Aerts, HJ 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer,* 48**,** 441-6.

Lee, M-K, Han, K, Kim, MK, Koh, ES, Kim, ES, Nam, GE, Hong, O-K, Kim, B & Kwon, H-S 2020. Combinations of metabolic syndrome components and the risk of type 2 diabetes mellitus: A nationwide cohort study. *Diabetes Res. Clin. Pract.,* 165**,** 108237.

Lee, SS, Lee, Y, Kim, N, Kim, SW, Byun, JH, Park, SH, Lee, MG & Ha, HK 2011. Hepatic fat quantification using chemical shift MR imaging and MR spectroscopy in the presence of hepatic iron deposition: validation in phantoms and in patients with chronic liver disease. *J Magn Reson Imaging,* 33**,** 1390-8.

Lewington, S, Clarke, R, Qizilbash, N, Peto, R & Collins, R 2002. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet,* 360**,** 1903–1913.

Lim, S, Shin, H, Song, JH, Kwak, SH, Kang, SM, Yoon, JW, Choi, SH, Cho, SI, Park, KS, Lee, HK, Jang, HC & Koh, KK 2011. Increasing Prevalence of Metabolic Syndrome in Korea: The Korean National Health and Nutrition Examination Survey for 1998–2007. *Diabetes Care,* 34**,** 1323–1328.

Linge, J, Whitcher, B, Borga, M & Dahlqvist Leinhard, O 2019. Sub-phenotyping Metabolic Disorders Using Body Composition: An Individualized, Nonparametric Approach Utilizing Large Data Sets. *Obesity (Silver Spring),* 27**,** 1190-1199.

Liu, J, Fox, CS, Hickson, DA, May, WD, Hairston, KG, Carr, JJ & Taylor, HA 2010. Impact of abdominal visceral and subcutaneous adipose tissue on cardiometabolic risk factors: the Jackson Heart Study. *J Clin Endocrinol Metab,* 95**,** 5419-26.

Lohsoonthorn, V, Dhanamun, B & Williams, MA 2006. Prevalence of metabolic syndrome and its relationship to white blood cell count in a population of Thai men and women receiving routine health examinations. *Am. J. Hypertens.,* 19**,** 339–345.

Lowekamp, BC, Chen, DT, Ibanez, L & Blezek, D 2013. The Design of SimpleITK. *Front. Neuroinform.,* 7.

Lund Haheim, L, Wisloff, TF, Holme, I & Nafstad, P 2006. Metabolic syndrome predicts prostate cancer in a cohort of middle-aged Norwegian men followed for 27 years. *Am J Epidemiol,* 164**,** 769-74.

Mach, F, Baigent, C, Catapano, AL, Koskinas, KC, Casula, M, Badimon, L, Chapman, MJ, De Backer, GG, Delgado, V, Ference, BA, Graham, IM, Halliday, A, et al. 2020. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J,* 41**,** 111-188.

Mcqueen, RB, Ghushchyan, V, Olufade, T, Sheehan, JJ, Nair, KV & Saseen, JJ 2016. Incremental increases in economic burden parallels cardiometabolic risk factors in the US. *Diabetes Metab Syndr Obes,* 9**,** 233-41.

Meisamy, S, Hines, CD, Hamilton, G, Sirlin, CB, Mckenzie, CA, Yu, H, Brittain, JH & Reeder, SB 2011. Quantification of hepatic steatosis with T1-independent, T2-corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. *Radiology,* 258**,** 767-75.

Mendizabal, Y, Llorens, S & Nava, E 2013. Hypertension in metabolic syndrome: vascular pathophysiology. *Int J Hypertens,* 2013**,** 230868.

Moebus, S, Hanisch, J, Bramlage, P, Lösch, C, Hauner, H, Wasem, J & Jöckel, K-H 2008. Regional Differences in the Prevalence of the Metabolic Syndrome in Primary Care Practices in Germany. *Dtsch. Ärztebl. Int.,* 105**,** 207.

Nasr, P, Fredrikson, M, Ekstedt, M & Kechagias, S 2020. The amount of liver fat predicts mortality and development of type 2 diabetes in non-alcoholic fatty liver disease. *Liver Int,* 40**,** 1069–1078.

NCD Risk Factor Collaboration 2017. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet,* 390**,** 2627-2642.

Ng, M, Fleming, T, Robinson, M, Thomson, B, Graetz, N, Margono, C, Mullany, EC, Biryukov, S, Abbafati, C, Abera, SF, Abraham, JP, Abu-Rmeileh, NM, et al. 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet,* 384**,** 766-81.

Nguyen, P, Leray, V, Diez, M, Serisier, S, Le Bloc'h, J, Siliart, B & Dumon, H 2008. Liver lipid metabolism. *J Anim Physiol Anim Nutr (Berl),* 92**,** 272-83.

Nichols, GA & Moler, EJ 2011. Metabolic syndrome components are associated with future medical costs independent of cardiovascular hospitalization and incident diabetes. *Metab Syndr Relat Disord,* 9**,** 127-33.

Noel, CE, Zhu, F, Lee, AY, Yanle, H & Parikh, PJ 2014. Segmentation precision of abdominal anatomy for MRI-based radiotherapy. *Med Dosim,* 39**,** 212-7.

Nordestgaard, BG 2016. Triglyceride-Rich Lipoproteins and Atherosclerotic Cardiovascular Disease: New Insights From Epidemiology, Genetics, and Biology. *Circ Res,* 118**,** 547-63.

Nougaret, S, Tibermacine, H, Tardieu, M & Sala, E 2019. Radiomics: an Introductory Guide to What It May Foretell. *Curr Oncol Rep,* 21**,** 70.

Noureddin, M, Lam, J, Peterson, MR, Middleton, M, Hamilton, G, Le, TA, Bettencourt, R, Changchien, C, Brenner, DA, Sirlin, C & Loomba, R 2013. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology,* 58**,** 1930-40.

Palmer, MK & Toth, PP 2019. Trends in Lipids, Obesity, Metabolic Syndrome, and Diabetes Mellitus in the United States: An NHANES Analysis (2003-2004 to 2013-2014). *Obesity (Silver Spring),* 27**,** 309-314.

Park, HJ, Lee, SS, Park, B, Yun, J, Sung, YS, Shim, WH, Shin, YM, Kim, SY, Lee, SJ & Lee, MG 2019. Radiomics Analysis of Gadoxetic Acid-enhanced MRI for Staging Liver Fibrosis. *Radiology,* 290**,** 380-387.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R, Dubourg, V, Vanderplas, J, Passos,

A, et al. 2011. Scikit-learn: Machine Learning in Python. *JMLR*, 2825-2830.

Perreault, L, Pan, Q, Mather, KJ, Watson, KE, Hamman, RF & Kahn, SE 2012. Effect of regression from prediabetes to normal glucose regulation on long-term reduction in diabetes risk: results from the Diabetes Prevention Program Outcomes Study. *The Lancet,* 379**,** 2243-2251.

Pihlajamaki, J, Gylling, H, Miettinen, TA & Laakso, M 2004. Insulin resistance is associated with increased cholesterol synthesis and decreased cholesterol absorption in normoglycemic men. *J Lipid Res,* 45**,** 507-12.

Portney, LG & Watkins, MP 2009. *Foundations of clinical research: applications to practice*, Pearson/Prentice Hall Upper Saddle River, NJ.

Rado, SD, Lorbeer, R, Gatidis, S, Machann, J, Storz, C, Nikolaou, K, Rathmann, W, Hoffmann, U, Peters, A, Bamberg, F & Schlett, CL 2019. MRI-based assessment and characterization of epicardial and paracardial fat depots in the context of impaired glucose metabolism and subclinical left-ventricular alterations. *Br J Radiol,* 92**,** 20180562.

R Development Core Team. 2010. *R: A language and environment for statistical computing* [Online]. Vienna, Austria. http://www.R-project.org [Accessed 09/07/2019].

Reaven, GM 1988. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes,* 37**,** 1595–1607.

Reaven, GM & Chen, YD 1988. Role of insulin in regulation of lipoprotein metabolism in diabetes. *Diabetes Metab Rev,* 4**,** 639-52.

Reaven, GM, Lerner, RL, Stern, MP & Farquhar, JW 1967. Role of insulin in endogenous hypertriglyceridemia. *J Clin Invest,* 46**,** 1756-67.

Reeder, SB, Cruite, I, Hamilton, G & Sirlin, CB 2011. Quantitative Assessment of Liver Fat with Magnetic Resonance Imaging and Spectroscopy. *J Magn Reson Imaging,* 34**,** 729-749.

Rhee, SY & Kim, YS 2018. The Role of Advanced Glycation End Products in Diabetic Vascular Complications. *Diabetes Metab J,* 42**,** 188-195.

Rosito, GA, Massaro, JM, Hoffmann, U, Ruberg, FL, Mahabadi, AA, Vasan, RS, O'donnell, CJ & Fox, CS 2008. Pericardial fat, visceral abdominal fat, cardiovascular disease risk factors, and vascular calcification in a community-based sample: the Framingham Heart Study. *Circulation,* 117**,** 605-13.

Saklayen, MG 2018. The Global Epidemic of the Metabolic Syndrome. *Curr Hypertens Rep,* 20**,** 12.

Sanchez, E, Betriu, A, Lopez-Cano, C, Hernandez, M, Fernandez, E, Purroy, F, Bermudez-Lopez, M, Farras-Salles, C, Barril, S, Pamplona, R, Rius, F, Hernandez, C, et al. 2019. Characteristics of atheromatosis in the prediabetes stage: a cross-sectional investigation of the ILERVAS project. *Cardiovasc Diabetol,* 18**,** 154.

Sanner, MF 1999. Python: a programming language for software integration and development. *J Mol Graph Model,* 17**,** 57-61.

Schillaci, G, Pirro, M, Vaudo, G, Gemelli, F, Marchesi, S, Porcellati, C & Mannarino, E 2004. Prognostic value of the metabolic syndrome in essential hypertension. *J Am Coll Cardiol,* 43**,** 1817-22.

Schwier, M, Van Griethuysen, J, Vangel, MG, Pieper, S, Peled, S, Tempany, C, Aerts, H, Kikinis, R, Fennessy, FM & Fedorov, A 2019. Repeatability of Multiparametric Prostate MRI Radiomics Features. *Sci Rep,* 9**,** 9441.

Scicali, R, Giral, P, Gallo, A, Di Pino, A, Rabuazzo, AM, Purrello, F, Cluzel, P, Redheuil, A, Bruckert, E & Rosenbaum, D 2016. HbA1c increase is associated with higher coronary and peripheral atherosclerotic burden in non diabetic patients. *Atherosclerosis,* 255**,** 102-108.

Scuteri, A, Laurent, S, Cucca, F, Cockcroft, J, Cunha, PG, Mañas, LR, Raso, FUM, Muiesan, ML, Ryliškytė, L, Rietzschel, E, Strait, J, Vlachopoulos, C, et al. 2015. THE METABOLIC SYNDROME ACROSS EUROPE – DIFFERENT CLUSTERS OF RISK FACTORS. *European journal of preventive cardiology,* 22**,** 486.

Seabold, S & Perktold, J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference, 2010. Austin, TX, 61.

Sharma, P, Martin, DR, Pineda, N, Xu, Q, Vos, M, Anania, F & Hu, X 2009. Quantitative analysis of T2-correction in single-voxel magnetic resonance spectroscopy of hepatic lipid fraction. *J Magn Reson Imaging,* 29**,** 629-35.

Shi, TH, Wang, B & Natarajan, S 2020. The Influence of Metabolic Syndrome in Predicting Mortality Risk Among US Adults: Importance of Metabolic Syndrome Even in Adults With Normal Weight. *Preventing chronic disease,* 17**,** E36-E36.

Shrout, PE & Fleiss, JL 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull,* 86**,** 420-8.

Singh, VP, Bali, A, Singh, N & Jaggi, AS 2014. Advanced glycation end products and diabetic complications. *Korean J Physiol Pharmacol,* 18**,** 1-14.

Smith, GI, Shankaran, M, Yoshino, M, Schweitzer, GG, Chondronikola, M, Beals, JW, Okunade, AL, Patterson, BW, Nyangau, E, Field, T, Sirlin, CB, Talukdar, S, et al. 2020. Insulin resistance drives hepatic de novo lipogenesis in nonalcoholic fatty liver disease. *J Clin Invest,* 130**,** 1453-1460.

Sookoian, S, Castano, GO, Burgueno, AL, Rosselli, MS, Gianotti, TF, Mallardi, P, Martino, JS & Pirola, CJ 2010. Circulating levels and hepatic expression of molecular mediators of atherosclerosis in nonalcoholic fatty liver disease. *Atherosclerosis,* 209**,** 585-91.

Stefansson, VTN, Schei, J, Solbu, MD, Jenssen, TG, Melsom, T & Eriksen, BO 2018. Metabolic syndrome but not obesity measures are risk factors for accelerated age-related glomerular filtration rate decline in the general population. *Kidney Int,* 93**,** 1183-1190.

Sullivan, PW, Ghushchyan, V, Wyatt, HR & Hill, JO 2007. The medical cost of cardiometabolic risk factor clusters in the United States. *Obesity (Silver Spring),* 15**,** 3150-8.

Targher, G, Byrne, CD, Lonardo, A, Zoppini, G & Barbui, C 2016. Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. *J Hepatol,* 65**,** 589-600.

Targher, G, Chonchol, M, Miele, L, Zoppini, G, Pichiri, I & Muggeo, M 2009. Nonalcoholic fatty liver disease as a contributor to hypercoagulation and

thrombophilia in the metabolic syndrome. *Semin Thromb Hemost,* 35**,** 277-87.

Thomsen, C, Becker, U, Winkler, K, Christoffersen, P, Jensen, M & Henriksen, O 1994. Quantification of liver fat using magnetic resonance spectroscopy. *Magn Reson Imaging,* 12**,** 487-95.

Tocci, G, Ferrucci, A, Bruno, G, Mannarino, E, Nati, G, Trimarco, B & Volpe, M 2015. Prevalence of metabolic syndrome in the clinical practice of general medicine in Italy. *Cardiovasc Diagn Ther,* 5**,** 271-9.

Toplak, H, Woodward, E, Yumuk, V, Oppert, JM, Halford, JC & Fruhbeck, G 2015. 2014 EASO Position Statement on the Use of Anti-Obesity Drugs. *Obes Facts,* 8**,** 166-74.

Trefts, E, Gannon, M & Wasserman, DH 2017. The liver. *Curr Biol,* 27**,** R1147-R1151.

Trinh, L, Lind, E, Peterson, P, Svensson, J, Olsson, LE & Mansson, S 2017. High-Resolution MR Imaging of Muscular Fat Fraction-Comparison of Three T2-Based Methods and Chemical Shift-Encoded Imaging. *Tomography,* 3**,** 153-162.

Ulbrich, EJ, Fischer, MA, Manoliu, A, Marcon, M, Luechinger, R, Nanz, D & Reiner, CS. 2015. Age- and Gender Dependent Liver Fat Content in a Healthy Normal BMI Population as Quantified by Fat-Water Separating DIXON MR Imaging. *PLoS One* [Online], 10. Available: 10.1371/journal.pone.0141691

Unger, T, Borghi, C, Charchar, F, Khan, NA, Poulter, NR, Prabhakaran, D, Ramirez, A, Schlaich, M, Stergiou, GS, Tomaszewski, M, Wainford, RD, Williams, B, et al. 2020. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. *Hypertension,* 75**,** 1334-1357.

Vague, J 1947. La differenciation sexuelle, facteur determinant des formes de l'obesite. *Presse Med,* 30**,** 339-40.

Van De Ven, B. 2020. Bokeh [Online]. https://bokeh.org/ [Accessed 06/08/2020]

Van Griethuysen, JJM, Fedorov, A, Parmar, C, Hosny, A, Aucoin, N, Narayan, V, Beets-Tan, RGH, Fillion-Robin, JC, Pieper, S & Aerts, H 2017. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res,* 77**,** e104-e107.

Vavlukis, M & Vavlukis, A 2018. Adding ezetimibe to statin therapy: latest evidence and clinical implications. *Drugs Context,* 7**,** 212534.

Virtanen, P, Gommers, R, Oliphant, TE, Haberland, M, Reddy, T, Cournapeau, D, Burovski, E, Peterson, P, Weckesser, W, Bright, J, Van Der Walt, SJ, Brett, M, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods,* 17**,** 261-272.

Wang, JJ, Li, HB, Kinnunen, L, Hu, G, Jarvinen, TM, Miettinen, ME, Yuan, S & Tuomilehto, J 2007. How well does the metabolic syndrome defined by five definitions predict incident diabetes and incident coronary heart disease in a Chinese population? *Atherosclerosis,* 192**,** 161-8.

Wang, M, Luo, Y, Cai, H, Xu, L, Huang, M, Li, C, Dong, Z, Li, ZP & Feng, ST 2019. Prediction of type 2 diabetes mellitus using noninvasive MRI quantitation of visceral abdominal adiposity tissue volume. *Quant Imaging Med Surg,* 9**,** 1076-1086.

Wickham, H 2009. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Williams, B, Mancia, G, Spiering, W, Agabiti Rosei, E, Azizi, M, Burnier, M, Clement, DL, Coca, A, De Simone, G, Dominiczak, A, Kahan, T, Mahfoud, F, et al. 2018. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J,* 39**,** 3021-3104.

Wilson, PW, D'agostino, RB, Parise, H, Sullivan, L & Meigs, JB 2005. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation,* 112**,** 3066-72.

Wolf, I, Vetter, M, Wegner, I, Bottger, T, Nolden, M, Schobinger, M, Hastenteufel, M, Kunert, T & Meinzer, HP 2005. The medical imaging interaction toolkit. *Med Image Anal,* 9**,** 594-604.

World Health Organization & International Diabetes Federation 2006. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia : report of a WHO/IDF consultation. Geneva: World Health Organization.

World Health Organization. 1999. *Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus* [Online]. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/66040 [Accessed 26/09/2020].

World Health Organization. 2020. *Obesity and overweight* [Online]. https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [Accessed 2020/10/26].

Wu, H & Ballantyne, CM 2020. Metabolic Inflammation and Insulin Resistance in Obesity. *Circ Res,* 126**,** 1549-1564.

Yamazaki, H, Tsuboya, T, Tsuji, K, Dohke, M & Maguchi, H 2015. Independent Association Between Improvement of Nonalcoholic Fatty Liver Disease and Reduced Incidence of Type 2 Diabetes. *Diabetes Care,* 38**,** 1673-9.

Yang, J, Dou, G, Tesche, C, De Cecco, CN, Jacobs, BE, Schoepf, UJ & Chen, Y 2019. Progression of coronary atherosclerotic plaque burden and relationship with adverse cardiovascular event in asymptomatic diabetic patients. *BMC Cardiovasc Disord,* 19**,** 39.

Yaribeygi, H, Farrokhi, FR, Butler, AE & Sahebkar, A 2019. Insulin resistance: Review of the underlying molecular mechanisms. *J. Cell. Physiol.,* 234**,** 8152–8161.

Yarnell, JW, Patterson, CC, Sweetnam, PM, Thomas, HF, Bainton, D, Elwood, PC, Bolton, CH & Miller, NE 2001. Do total and high density lipoprotein cholesterol and triglycerides act independently in the prediction of ischemic heart disease? Ten-year follow-up of Caerphilly and Speedwell Cohorts. *Arterioscler Thromb Vasc Biol,* 21**,** 1340-5.

Yki-Jarvinen, H 2014. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol,* 2**,** 901-10.

Yokoo, T, Serai, SD, Pirasteh, A, Bashir, MR, Hamilton, G, Hernando, D, Hu, HH, Hetterich, H, Kuhn, JP, Kukuk, GM, Loomba, R, Middleton, MS, et al. 2018. Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis. *Radiology,* 286**,** 486-498.

Yoon, YS, Choi, HS, Kim, JK, Kim, YI & Oh, SW 2016. Differences in the associations of anthropometric measures with insulin resistance and type 2 diabetes mellitus between Korean and US populations: Comparisons of representative nationwide sample data. *Obes Res Clin Pract,* 10**,** 642-651.

Younossi, Z, Anstee, QM, Marietti, M, Hardy, T, Henry, L, Eslam, M, George, J & Bugianesi, E 2018. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol,* 15**,** 11-20.

Younossi, ZM, Otgonsuren, M, Henry, L, Venkatesan, C, Mishra, A, Erario, M & Hunt, S 2015. Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009. *Hepatology,* 62**,** 1723-30.

Yumuk, V, Fruhbeck, G, Oppert, JM, Woodward, E & Toplak, H 2014. An EASO position statement on multidisciplinary obesity management in adults. *Obes Facts,* 7**,** 96-101.

Zelber-Sagi, S, Lotan, R, Shibolet, O, Webb, M, Buch, A, Nitzan-Kaluski, D, Halpern, Z, Santo, E & Oren, R 2013. Non-alcoholic fatty liver disease independently predicts prediabetes during a 7-year prospective follow-up. *Liver Int,* 33**,** 1406-12.

Zhou, YY, Zhou, XD, Wu, SJ, Fan, DH, Van Poucke, S, Chen, YP, Fu, SW & Zheng, MH 2018. Nonalcoholic fatty liver disease contributes to subclinical atherosclerosis: A systematic review and meta-analysis. *Hepatol Commun,* 2**,** 376-392.

Zwanenburg, A, Leger, S, Agolli, L, Pilz, K, Troost, EGC, Richter, C & Lock, S 2019. Assessing robustness of radiomic features by image perturbation. *Sci Rep,* 9**,** 614.

Zwanenburg, A, Leger, S, Vallières, M & Löck, S 2016. Image biomarker standardisation initiative. [Online] Available: http://arxiv.org/abs/1612.07003

8 Erklärung zum Eigenanteil

Diese Arbeit wurde am Universitätsklinikum Tübingen in der Abteilung für Diagnostische und Interventionelle Radiologie unter Betreuung von Prof. Dr. Fabian Bamberg durchgeführt.

Die T1-DED Datensätze sowie sämtliche Kovariaten einschließlich der Leberfettmessungen aus T1-MED und MRS Bilddaten der KORA MRT Studie wurden mir, Daniel A. P. Gutmann, auf Antrag von dem Helmholtz Zentrum München zur Verfügung gestellt.

Die Grundidee einer Heterogenitätsanalyse stammt von Prof. Dr. med. Fabian Bamberg. Die Wahl einer radiomischen Merkmalsanalyse stammt von Prof. Dr. med. Sergios Gatidis und mir. Das Konzept die radiomische Merkmalsstabilität anhand von simulierten Test-Retest- und Interrater-Szenarien zu bewerten stammt von mir.

Die manuelle Segmentierung der Lebervolumina an T1-DED Bilddaten wurde von mir durchgeführt. Eine Überprüfung der Ergebnisse erfolgte durch Prof. Dr. med. Sergios Gatidis.

Hinsichtlich der Erfassung radiomischer Merkmale wurden die folgenden Schritte, einschließlich der Entwicklung notwendiger Python Skripte von mir durchgeführt: (i) Berechnung von rfwc-Karten aus T1-DED Bilddaten sowie sämtliches Postprocessing, (ii) Generierung rauschverstärkter rfwc Karten, (iii) Erstellung räumlich deformierter Lebervolumina sowie die (iv) Extraktion der radiomischen Merkmale.

Die Aufteilung der Probanden in Trainings- und Validierungsdatensätze sowie die statistische Auswertung der epidemiologischen Parameter wurden von mir durchgeführt und von Dr. Susanne Rospleszcz überprüft. Die Details des statistischen Analysekonzepts der Merkmalstabilität wurden von Dr. Susanne Rospleszcz und von mir erarbeitet. Dr. Susanne Rospleszcz berechnete die Test-Retest-Zuverlässigkeit. Die Interrater-Reliabilität wurde von mir bestimmt. Die

Analyse des Selektionsprozesses und der Merkmalsstabilität erfolgte unter Anregung von Dr. Susanne Rospleszcz sowie durch mich.

Die Evaluierung der radiomischen Merkmale als Bildgebende Marker für T2DM und MetS mittels Random Forest Algorithmen wurde von mir durchgeführt.

Alle Tabellen und Graphiken der Publikation wurden, auch unter Berücksichtigung von Vorschlägen von Dr. Susanne Rospleszcz, von mir erstellt.

Die aus den Daten hervorgegangene wissenschaftliche Publikation habe ich selbstständig mit Korrekturen und Anregungen von Dr. Susanne Rospleszcz und ferner Prof. Dr. med. Fabian Bamberg verfasst.

Die Literaturrecherche für die vorliegende Arbeit und die Publikation habe ich in vollem Umfang selbständig durchgeführt.

Ich versichere das hier vorliegende Manuskript selbstständig verfasst und keine weiteren als die angegebenen Quellen verwendet zu haben.

Sigmaringen, den 21.11.2020

Daniel A. P. Gutmann

9 List of publications

First author:

Gutmann, DAP, Rospleszcz, S, Rathmann, W, Schlett, CL, Peters, A, Wachinger, C, Gatidis, S & Bamberg, F. 2020. MRI-Derived Radiomics Features of Hepatic Fat Predict Metabolic States in Individuals without Cardiovascular Disease. *Acad Radiol* [Online] Available: 10.1016/j.acra.2020.06.030

Co-author on conference papers:

These papers used the manually segmented liver masks prepared for this study as input for Deep-learning networks. I had no part in data analysis or preparation of the manuscripts.

Senapati, J, Roy, AG, Pölsterl, S, Gutmann, DAP, Gatidis, S, Schlett, CL, Peters, A, Bamberg, F & Wachinger, C. 2020b. Bayesian Neural Networks for Uncertainty Estimation of Imaging Biomarkers. [Online] Available: 10.1007/978-3-030-59861-7_28

Gutiérrez-Becker, B, Gatidis, S, Gutmann, DAP, Peters, A, Schlett, CL, Bamberg, F & Wachinger, C 2018. Deep Shape Analysis on Abdominal Organs for Diabetes Prediction. *CoRR,* abs/1808.01946.

10 Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Dr. med. Fabian Bamberg for his continuous support of my dissertation and related scientific inquiries. He was always patient, motivating and freely shared his immense knowledge. His guidance was paramount to my work - I could not have imagined having had better advisor and mentor for my dissertation.

My sincere thanks also go to Prof. Dr. med. Sergios Gatidis, Dr. Susanne Rospleszcz and the staff of the Helmholtz center in Munich. Without their precious support it would not have been possible to conduct this research.