



Hochschule für angewandte Wissenschaften Coburg
Fakultät Elektrotechnik und Informatik

Studiengang: Bachelor Informatik

Einsatz von maschinellem Lernen zur
Echtzeit-Überwachung der psychischen
Gesundheit durch Analyse physiologischer
Marker mittels alltäglicher tragbarer Sensoren

Keck, Paul

Abgabe der Arbeit: 28.05.2024

Betreut durch:

Prof. Dr. Stephan Streuber, Hochschule Coburg

Inhaltsverzeichnis

Abbildungsverzeichnis	3
Tabellenverzeichnis	4
1 Einführung	5
2 Theoretischer Rahmen Stress	6
2.1 Moderne Interpretation	6
2.2 Physiologische Marker für Stress	7
3 Wearables	8
3.1 Technologische Grundlagen Wearables	8
3.1.1 Photoplethysmographie	9
3.1.2 Elektrokardiogramm	9
3.1.3 Weiter relevante Sensorik	9
3.2 Limitationen von Wearables	10
4 Validierung Wesad Dataset	11
4.1 Begebenheiten des Datensatzes	11
4.2 Auswertung der EKG-Daten	11
4.3 Bewertung und Diskussion	13
5 Maschinelles Lernen zur Klassifizierung von Stress	14
5.1 Random-Forest	14
5.2 Fallstudie RF	16
5.3 Unsupervised	16
6 Proposal	17
7 Fazit	18
Literaturverzeichnis	19
Ehrenwörtliche Erklärung	22

Abbildungsverzeichnis

Abb. 1:	Erkannte R-Peaks im 12-Sekunden-Ausschnitt des gefilterten EKG-Signals .	12
Abb. 2:	Beispielhafte Darstellung Desicion Tree	15

Tabellenverzeichnis

Tab. 1:	Ergebnisse WESAD-Exploration	13
---------	--	----

1 Einführung

In unserer modernen, schnelllebigen Welt ist Stress zu einer allgegenwärtigen Erscheinung geworden. Angesichts der wachsenden Anforderungen im beruflichen und privaten Leben, einem Rückgang des Wirtschaftswachstums in der westlichen Welt und angespannten geopolitischen Verhältnissen erleben viele Menschen regelmäßig hohe Stresslevel, die zunehmend mit langfristigen gesundheitlichen Problemen, insbesondere Depressionen, in Verbindung gebracht werden [WLS08] [WP01].

Die Bürde psychischer Störungen war bereits vor der Covid-19-Pandemie ein prägendes Thema in der Gesellschaft, doch durch die Pandemie nahm das Ausmaß dieser Probleme noch weiter zu. Die Auswirkungen der globalen COVID-19-Pandemie haben weltweit zu einer signifikanten Zunahme von Depressionen und Angststörungen geführt. Laut einer Studie wurden während der Pandemie erhebliche Einflüsse von täglichen SARS-CoV-2-Infektionsraten auf die Prävalenz von schweren depressiven Störungen und Angststörungen beobachtet. Global gesehen stieg die Anzahl der Fälle von schweren depressiven Störungen um zusätzliche 53,2 Millionen (eine Zunahme von 27,6%), was einer Gesamtprävalenz von 3152,9 Fällen pro 100.000 Einwohner entspricht, und die Fälle von Angststörungen stiegen um 76,2 Millionen (eine Zunahme von 25,6%), was einer Gesamtprävalenz von 4802,4 Fällen pro 100.000 Einwohner entspricht [SMHS⁺21].

In dieser Zeit wachsender Herausforderungen für die psychische Gesundheit sind innovative Ansätze zur Überwachung und Früherkennung von Stresssymptomen gefragter denn je. Die grundlegenden Arbeiten von Wissenschaftlern wie Walter B. Cannon und Hans Selye haben gezeigt, dass Stress nicht nur durch subjektive Erfahrungen gekennzeichnet ist, sondern sich auch durch messbare physiologische Veränderungen auszeichnet. Cannon beschrieb die „fight-or-flight“-Reaktion, während Selye die Stressreaktion und das General Adaptation Syndrome entwickelte, die erklären, wie Stress die körperliche Gesundheit beeinflusst [Can15] [Sel36].

Prognosen zufolge sollen im Jahr 2028 allein über sechshundert Millionen Wearables, wie Smartwatches und Fitness-Tracker, ausgeliefert werden [LU24]. Diese Geräte sind bereits heute in der Lage, entscheidende physiologische Marker wie Herzfrequenz, Schlafmuster und Aktivitätslevel zu erfassen, die Schlüsselindikatoren für die Erkennung von Stress sind. Trotz der Fähigkeit, wichtige physiologische Daten zu sammeln, konzentriert sich die Forschung bisher hauptsächlich auf klinische Sensoren. Die alltäglichen Wearables, die einen kontinuierlichen Zugang zu physiologischen Daten bieten, sind jedoch in der breiten Bevölkerung bereits weit verbreitet und könnten eine Schlüsselrolle bei der Früherkennung und dem Management von Stress spielen.

In dieser Arbeit wird untersucht, wie maschinelles Lernen eingesetzt werden kann, um physiologische Daten aus der in alltäglichen Wearables verbauten Sensorik zur Klassifizierung von

Stresszuständen zu nutzen. Dazu wird zuerst ein Überblick über die Definitionen von Stress und insbesondere dessen messbare physiologische Auswirkungen.... TOOOODDDDDDOOOOO!!!

2 Theoretischer Rahmen Stress

"The most significant feature of these bodily reactions in pain and in the presence of emotion provoking objects is that they are of the nature of reflexes" [Can15]. Dieses Erkenntnis bildet das Fundament von Walter B. Cannons Forschungen, die zeigen, dass der Körper in als stressvoll wahrgenommenen Situationen, wie bei Schmerz und Hunger, Adrenalin ausschüttet. Cannon prägte den Begriff der "Homöostase", der den Prozess beschreibt, durch den der Körper ein inneres Gleichgewicht trotz externer Störungen aufrechterhält. Seine Arbeiten legen den Schwerpunkt auf die physiologischen Reflexe, bekannt als die "Fight-or-Flight"-Reaktion, die eine zentrale Rolle in der Aufrechterhaltung dieser Homöostase spielen.

Hans Selye, der in den 1930er Jahren den Begriff "Stress" wissenschaftlich prägte, erweiterte dieses Konzept durch die Einführung des General Adaptation Syndroms (GAS). Er definierte Stress als eine unspezifische Antwort des Körpers auf jegliche Art von Reizen, ob positiv oder negativ. Das GAS beschreibt, wie der Körper in drei Phasen auf Stressoren reagiert:

Alarmreaktion: Der Körper reagiert sofort auf einen Stressor, indem er Adrenalin freisetzt und den von Cannon beschriebenen "Fight-or-Flight"-Reflex auslöst, um die Homöostase schnell wiederherzustellen. Resistenz-Phase: Bei anhaltendem Stress versucht der Körper, sich anzupassen und die Homöostase durch Regulation physiologischer und psychologischer Systeme aufrechtzuerhalten. Erschöpfungs-Phase: Gelingt es dem Körper nicht, sich anzupassen, führt dies zu einer Erschöpfung, die gesundheitliche Schäden bis hin zum Tod zur Folge haben kann [Sel36].

2.1 Moderne Interpretation

Aufbauend auf diesen grundlegenden Konzepten entwickelt sich die Definition von Stress aber durchaus mit wachsendem Forschungsstand weiter. In neueren Konzepten, wie von David S. Goldstein und anderen Wissenschaftlern entwickelt, wird Stress als eine bewusste oder unbewusste Bedrohung der Homöostase angesehen. Dieser Ansatz erkennt an, dass die Reaktion des Körpers auf Stress durchaus spezifisch sein kann, je nachdem, welche Herausforderung der Homöostase vorliegt, wie das Individuum den Stressor wahrnimmt und wie gut es sich in der Lage sieht, mit ihm umzugehen.

Goldstein hebt hervor, dass neben Cannons "sympathoadrenerges System", welches sich Allgemein auf das Zusammenspiel des Sympathischen Nervensystems und der Nebenniere zur

wiederherstellung der Homöostase durch hormonelle Regulierung bezieht, zur Aufrechterhaltung der Homöostase auch das Hypothalamus-Hypophysen-Nebennierenrinden-System (HPA-Achse) eine zentrale Rolle in der Stressreaktion spielt. Beide Systeme sind aktiv an der Regulation und Anpassung bei Stress beteiligt und tragen dazu bei, eine neue Stabilität zu erreichen, die als "Allostase" bezeichnet wird.

Goldstein benutzt eine Analogie zu einem Heizkreislauf, um Allostase zu erklären: Ein Mensch stellt in einem Thermostat eine Soll-Temperatur ein; dieses Thermostat erfasst die aktuelle Raumtemperatur und aktiviert die Heizung, wenn die Temperatur unter den Sollwert fällt. Ähnlich reguliert der Körper durch Allostase die "innere Temperatur" oder physiologische Sollwerte, indem er nach Bedarf Anpassungen vornimmt, um Stabilität durch Veränderung zu erreichen.

Allostase beschreibt die Fähigkeit, Stabilität durch Veränderung zu erhalten, wobei "allostatische Last" die Kosten dieser Anpassungen misst. Diese Last kann, besonders bei langfristigem oder intensivem Stress, zu negativen gesundheitlichen Auswirkungen führen. Goldstein führt aus, dass zum Beispiel eine chronische Erhöhung des Blutdrucks, die zunächst dazu dient, die Durchblutung des Gehirns sicherzustellen, langfristig zu Gefäßschädigungen und daraus resultierenden Krankheiten wie Schlaganfall oder Herzinfarkt führen kann. [GK07]

Das moderne Verständnis von Stress betont somit nicht nur die Akutreaktion des Körpers, sondern auch die langfristigen Auswirkungen der Stressbewältigung und die Bedeutung der Anpassungsmechanismen, die über die unmittelbare "Fight-or-Flight"-Reaktion hinausgehen.

2.2 Physiologische Marker für Stress

Wie bereits erleuchtet wurde, versucht der Körper verallgemeinert über zwei Systeme mit Stress umzugehen; der HPA-Achse und dem sympathoadrenergen System. Das sympathoadrenerge System bewirkt eine Ausschüttung von Adrenalin sowie Noradrenalin, welche in kurzer Zeit physiologische Anpassung wie einer Erhöhung des Blutdrucks bewirken, wohingegen die HPA-Achse etwas langsamer über Anregung von Cortisolproduktion fungiert. [KJ23]

Da sich diese Arbeit darauf fokussiert, Stress anhand messbarer physiologischer Merkmale zu klassifizieren, widmen wir einen etwas ausführlicheren Blick auf folgendes Paper von Ernst:

Die Studie umfasst einen multimodalen Ansatz um physiologische Marker in akuten Stresssituationen aufzuzeichnen und auszuwerten. In fünfundsechzig gesunden Teilnehmern wurden mithilfe des Mannheim Multicomponent Stress Test (MMST), einen verschiedenen Komponenten umfassenden Test, darunter kognitive Aufgaben und sensorische Stimulation, eine kontrollierte, akute Stressreaktion erzeugt. Während des Test wurden mit Hilfe einer Kombination von Biosensorik wie Elektrokardiogramm (EKG) und Photoplethysmographie (PPG) und

Selbstevaluierungsbögen sechzig verschiedene physiologische Marker aufgezeichnet, darunter Herzratenvariabilität, Hautleitfähigkeit, Atmungsrate und mehr. Weiterhin wurden über unter anderem Speichelpuben die Konzentrationen von Cortisol gemessen, um eine Aktivierung der HPA-Achse nachweisen zu können.

Diese wurden einer statistischen Untersuchung unterzogen, um die Differenzen zwischen diesem Markern im Ruhezustand im Vergleich zum Stresszustand aufzuzeigen. Demzufolge wurde unter anderem eine Reduktion der Herzratenvariabilität; herausgerechnet als Root Mean Square of Successive Differences (RMSSD) aus den EKG Messungen, von im Schnitt 46.0 ± 18.2 Millisekunden auf 37.3 ± 16.1 Millisekunden, was einer Differenz des Durchschnittes von 19 Prozent bedeutet, festgestellt. [ESP⁺23]

Insgesamt deckten sich die Ergebnisse weitgehend mit Erkenntnissen anderer Literatur sowie mit den Standardisierten Selbstevaluierungsbögen, womit ein wissenschaftlich zuverlässiger Rahmen an bedeutenden physiologischen Markern zur Erkennung von Stress geschaffen wurde, an dem wir uns in dieser Arbeit noch weitergehend bedienen werden.

3 Wearables

Wie eingangs bereits erwähnt ist der Markt für Wearables gigantisch und soll im Jahr 2028 eine Auslieferungszahl von sechshundert Millionen Stück erreichen. Insbesondere Smartwatches und Fitnesstracker erfreuen sich in Europa einer steigenden Popularität, wonach laut einer Statista Umfrage in Deutschland im Jahre 2023 bereits fünfunddreißig Prozent ein solches nutzen (Vergleich 2021: neunzehn Prozent). [Boc23]

Insbesondere diese eröffnen breite Möglichkeiten zur Erfassung von Stress, da viele dieser Wearables nicht invasive Biosensorik verbaut haben, und sowieso schon alltäglich getragen werden. Um eingrenzend tiefer auf die technologischen Merkmale eingehen zu können, fokussieren wir uns auf die Apple Watch, welche nach geschätzten Absatzzahlen aus dem Jahr 2020 mit über 33 Millionen Stück eine große Marktdominanz zeigte. [Boc21]

3.1 Technologische Grundlagen Wearables

Es wird kurz die für die Erfassung von physiologischen Stressmarker wichtigste Sensorik der Apple Watch beleuchtet. Dabei werden speziell auf die PPG und das EKG eingegangen, da diese genutzt werden um die obig erwähnte wichtige Herzratenvariabilität (HRV) zu berechnen.

3.1.1 Photoplethysmographie

PPG ist eine kostengünstige Methode um zahlreiche Marker des Herz-Kreislaufes zu erfassen, indem ein Lichtsignal, meist im Infrarot-Bereich, entsendet wird und mittels Photodioden gemessen wird, wie viel dieses Lichts vom mikrovaskulären Blut absorbiert wurde. Da sich das Blutvolumen im Laufe des Herzschlagszyklus ändert, kann bei wiederholten Messung die Herzfrequenz sowie andere daraus resultierende Werte errechnet werden. [All07].

Ein solcher Sensor ist in jeder Apple Watch verbaut, welche Stand jetzt auf den Markt kam. Die Apple Watch benutzt Grüne-LEDs, welches mehrer hundert male pro Sekunde leuchten, wobei das reflektierte Licht von Photodioden gemessen wird. Weiterhin besitzt der PPG Sensor der Apple Watch einen Hintergrundoperationsmodus, welcher für durchgängige Messungen im Hintergrund mit Infrarot fungiert.

Tatsächlich benutzt jede Apple Watch die PPG Messungen zur Ermittlung der HRV, sowie zur Erfassung der Herzfrequenz und der Blutsauerstoffsättigung.[App23]

3.1.2 Elektrokardiogramm

Das Herz fungiert auf der Basis von rhythmischen An- und Entspannen der Vorhöfe und Herzkammern. Dieser Vorgang wird durch elektrische Potentiale angestoßen, welche mit Hilfe eines EKGs gemessen werden können.

Seit der Einführung der Apple Watch Series 4 ist ein EKG Bestandteil der Sensorausstattung. Dieses erfasst die elektrischen Signale der Herzmuskeln durch spezielle Sensoren an der Rückseite der Uhr und der digitalen Krone, welche der Nutzer mit der nicht dominanten Hand für circa 30 Sekunden berühren muss und so den Stromkreis schließt.

3.1.3 Weiter relevante Sensorik

Die Apple Watch verfügt noch über weiter Sensorik, welche zur Klassifizierung von Stress nützliche Daten liefern kann. Ein Mikrofon kann benutzt werden um Änderungen an der Stimmlage festzustellen, wohingegen das verbaute 3-Achsen Gyroskop und der Beschleunigungssensor benutzt werden können um Körperbewegungen wahrzunehmen. Óscar Martínez Mozos et al. benutzen diese Metriken in Kombination mit EKG und PPG Daten klinischer Sensoren um mittels des Klassifizierungsalgorithmus Support Vector Machine (SVM) und dem Metalgorithmus Adaptive Boosting (AdaBoost) binär zwischen nicht-Stress und Stress zu klassifizieren. [MSA⁺17]

Auch wenn diese Sensorik keine Daten zu den physiologischen Markern ermittelt, auf welche sich in dieser Arbeit fokussiert wird, behalten wird diese für supplementäre Features im Hinterkopf.

3.2 Limitationen von Wearables

Eine Limitation der Apple Watch könnte die Messgenauigkeit der verbauten Sensorik sein. Hernando et al. validierten in einer Studie die Genauigkeit der aus den RR-Intervalle, welche die Apple Watch mithilfe der PPG aufzeichnet, herausgerechneten Herzfrequenz sowie HRV. Zwanzig Teilnehmer trugen ein Validierungsgerät und eine Apple Watch 4 während einer 5-minütigen Ruhephase und einer 5-minütigen Phase, in welcher mittels des Stroop Test leichter Stress induziert wurde.

Die Studie ergab nach Synchronisation der Signale und statistischer Auswertung eine sehr gute Übereinstimmung jeweils in den rohen RR-Intervallen, als auch in der Herzfrequenz und der HRV. Gleichzeitig wurde aber auch festgestellt, dass die Apple Watch vermehrt RR-Intervalle nicht identifizieren konnte, was der Autor auf eine Nichterkennung des PPG-Pulses zurückführt, was bei nicht richtigem Sitz beziehungsweise vor allem auch viel Bewegung passieren kann.

Dies beeinträchtigt die Messergebnisse nicht erwähnenswert, gilt aber im Auge behalten zu werden, da verinzelt fehlende RR-Intervalle insbesondere die HRV, welche beispielsweise aus der Standardabweichung der Abstände zwischen RR-Intervallen berechnet werden kann, beeinflussen. [HRS⁺18]

Ähnliche Erkenntnisse berichten auch B. Bent et al., welche vier Konsumer-Wearables, Apple Watch 4 inkludiert, und zwei Wearables in Forschungsqualität gegen ein EKG als Referenzstandard gebenchmarked haben. Dabei wurden dreißig Teilnehmer in jeweils drei Runden, umfassend einer Ruhephase im Sitzen, tiefen Atmens, körperlicher Aktivität und einer folgenden Ruhephase, mit Geräten ausgestattet, um pro Teilnehmer eine Messung der Herzfrequenz jedes Wearables zu haben. Sie berichten von einer höheren Genauigkeit der Herzfrequenz der Konsumer-Wearables im Ruhezustand als die der Forschungsgeräte, bestätigen aber Hernandos et al. Vermutung, dass die Messungen bei körperlicher Aktivität niedrigere Qualität aufweisen. [BGKD20]

Bei dieser Studie wurden die Teilnehmer auch extra nach diversen Hautfarbtönen ausgewählt, um zu untersuchen ob die Hautfarbe einen Einfluss auf die Qualität der Herzfrequenzmessung hat, was Aufgrund der Absorptionsfähigkeit von Melanin eine oft vermutete Limitation von PPG-Messungen ist. B. Brent et al. konnten keine Korrelation zwischen dem Hautfarbton und der Messungsgenauigkeit aufzeigen, was aber kritisch zu betrachten ist, da dies nur bedingt dem Konsens anderer Studien folgt.

Die Genauigkeit von alltäglichen Wearables scheint vielversprechend zu sein, es gilt zu beachten, dass dies während körperlicher Aktivität abweichen kann und im Raum steht, dass PPG einen Racial-Bias beinhaltet, welcher physisch bedingt ist, aber nichtsdestotrotz vorallem bei prediktiven medizinischen Verfahren so gut es geht eliminiert werden muss.

4 Validierung Wesad Dataset

Um die Erkenntnisse von H. Ernst et al. zu reproduzieren und im anbeacht der obig genannten Limitation, welche sich hauptsächlich auf PPG beziehen, möchte ich einen eigenen kleinen empirischen Anteil anhand des WESAD-Datensatzes [SRD⁺18] beitragen. Es wird versucht aus dem im Datensatz gegebenem EKG-Signal die Änderung der Herzfrequenz sowie der Herzratenvariabilität von der Baseline zum Stresszustand zu errechnen.

4.1 Begebenheiten des Datensatzes

Der Datensatz umfasst physiologische und Bewegungsdaten, die von 17 Teilnehmern (zwei Teilnehmer sind aufgrund von Komplikationen der Sensorik ausgeschieden) während einer Laborsitzung mit einem an der Brust und einem am Handgelenk getragenen Gerät aufgezeichnet wurden. Das an der Brust befestigte Gerät zeichnete ein 3-Elektroden EKG mit einer Abtastrate von 700Hz auf, welches genauer untersucht wird.

Die Studie wurde mit einem definierten Protokoll durchgeführt, das darauf abzielte, verschiedene emotionale Zustände zu induzieren:

Baseline-Zustand: Die Teilnehmer ruhten sich aus, während sie neutrales Lesematerial zur Verfügung hatten.

Amusement-Zustand: Die Teilnehmer sahen sich eine Reihe von lustigen Videoclips an.

Stress-Zustand: Dies wurde durch den Trier Social Stress Test (TSST) induziert, der öffentliches Sprechen und mentale Arithmetikaufgaben umfasste.

Meditationsperioden: Nach den Stress- und Amusement-Bedingungen führten die Teilnehmer eine geleitete Meditation durch, um sie zu beruhigen.

Ausserdem wurde die Studie multimodal ausgelegt, wobei die Teilnehmer nach dem Durchlauf des Protokolls mehrere Fragebögen zur Selbstevaluierung des Empfindes während des Protokolls ausfüllten.

4.2 Auswertung der EKG-Daten

Der Datensatz enthält eine Python-Pickle pro Teilnehmer mit den synchronisierten Sensor-Daten sowie den jeweiligen Labels, wobei 0 die Baseline-Messungen und 2 die Stress-Messungen darstellt. Es sind noch Labels für 'Amusement' und anderweitige Zustände verfügbar, diese wurden aber nicht berücksichtigt.

Weiterhin beinhaltet jeder Teilnehmer-Orderner einer Textdatei, in welcher unter anderem angemerkt wurde, ob ein Teilnehmer beispielsweise vor der Studie raucht, Koffein konsumierte oder Sport betrieb. Es werden sich nur Teilnehmer angeschaut, welche all diese Anmerkungen mit 'Nein' beantworteten.

Die Pickle Dateien wurde in ein Pandas DataFrame geladen. Darauf wurden pro Teilnehmer Teilsegmente von jeweils 30-Sekunden bestimmt, was bei einer Abtastrate von 700Hz genau 21.000 Datenpunkten entspricht. Segmente welche nicht die Anzahl dieser Datenpunkte erfüllen konnten (die letzten einer Messreihe eines Teilnehmers) wurden nicht beachtet. Das Label welches für ein Segment galt, wurde über ein Mehrheitsvotum der Label der einzelnen Datenpunkte bestimmt.

Mithilfe des Packages neurokit2, welches speziell für EKG-Signale parametrisierte Filter anwendet, wurden pro Segment die R-Peaks bestimmt aus welchen die RR-Intervalle berechnet wurde. Siehe Abbildung 1.

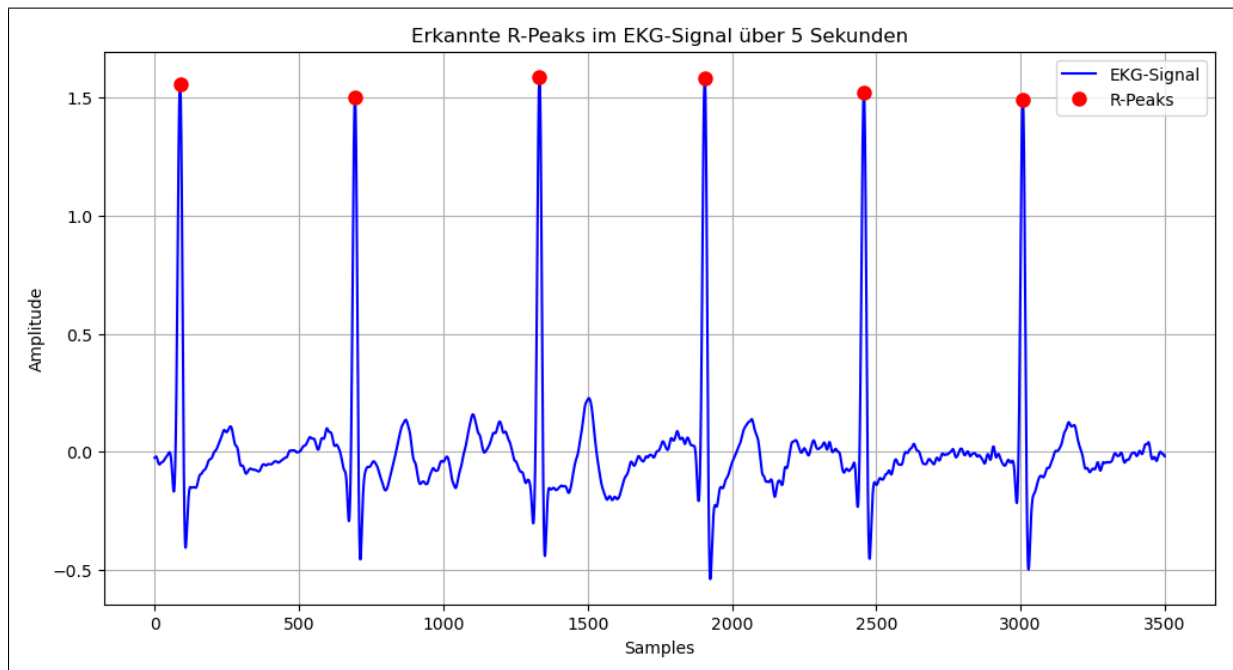


Abb. 1: Erkannte R-Peaks im 12-Sekunden-Ausschnitt des gefilterten EKG-Signals

Mit diesen konnte pro Segment die Durchschnittliche Herzfrequenz sowie die HRV mittels RMSSD wie folgt:

$$\text{RMSSD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}$$

berechnet werden.

Daraufhin wurden alle Segmente entfernt, welche nach dem Majority-Voting nicht mit Baseline oder Stress gelabelt wurden.

4.3 Bewertung und Diskussion

Im Tabelle 1 werden die durchschnittliche Herzfrequenz und HRV pro Teilnehmer pro Label aufgezeigt, als auch die prozentuale Veränderung von Baseline zu Stress.

Über alle berücksichtigten Teilnehmer veränderte sich die Herzfrequenz um **28.874% ± 17.287** und die HRV_{RMSSD} um **-46.676% ± 22.230**.

Teilnehmer	Kein Stress		Stress		Delta	
	HR_Mean [BPM]	HRV_Mean [ms]	HR_Mean [BPM]	HRV_Mean [ms]	HR_Delta [%]	HRV_Delta [%]
S2	72.99	96.38	76.91	43.07	5.36	-55.31
S3	60.99	130.06	86.29	55.63	41.48	-57.23
S4	64.75	80.48	79.62	42.05	22.96	-47.76
S10	92.09	33.37	107.13	29.90	16.33	-10.40
S13	84.92	35.80	108.32	19.30	27.56	-46.08
S14	88.51	29.21	126.90	11.22	43.36	-61.58
S15	80.01	46.45	84.94	42.33	6.16	-8.86
S16	86.99	36.48	128.14	11.21	47.31	-69.26
S17	75.88	85.89	113.32	31.25	49.33	-18.74

Tab. 1: Ergebnisse WESAD-Exploration

Betrachtet man die Ergebnisse aus der Studie von Ernst et al. [ESP⁺23] zum Vergleich, liegen dort die Veränderungen der Herzfrequenz von Baseline zu Stress im Durchschnitt bei **8.361%** (herausgerechnet aus den RR_{mean} [ms] Angaben) sowie bei **-19%** für die HRV_{RMSSD} . Es gibt mehrere Vermutungen, wieso die Veränderungen im Schnitt bei Auswertung des WESAD-Datensatzes präsenter hervortreten. Einerseits ist die geringere gültige Sample-Size zu beachten (WESAD : 9 | Ernst et al. 60), was bei der Betrachtung der von Teilnehmer zu Teilnehmer unterschiedlichen Reaktionen in Tabelle 1 vermutlich der Hauptfaktor ist. Weiterhin unterschieden sich die verwendeten standardisierten Test (TSST; MMST). Insgesamt ist festzuhalten, dass grundsätzlich in Stressreaktionen durch Aktivierung der HPA-Achse und erhöhter Aktivität des sympathischen Nervensystems eine Reduktion der HRV_{RMSSD} und der Herzfrequenz [BPM] zu erwarten ist. Diese Erkenntnis kann genutzt werden, um Klassifizierungsmodelle anhand dieser Features zu trainieren und letztendlich einen Stresszustand zu klassifizieren.

Diese kleine Exkursion hat gezeigt, dass grundsätzlich das 30-sekündige EKG einer Apple Watch 4 und höher dafür geeignet ist, physiologische Werte aufzuzeichnen, aus denen aussagekräftige Features extrahiert werden können.

In Siirtolas Paper wurde auf genau diesen Datensatz unter anderem ein Random Forest (RF) Model trainiert, welches bei binärer Stress/nicht-Stress Klassifizierung eine Balanced-Accuracy von 73.2% bei alleiniger Verwendung der Herzfrequenz als Feature erreichte. [Sii19] RF wird im nächsten Kapitel dieser Arbeit noch genauer vorgestellt. Dies stellt einen guten Wert, welcher sich sogar im Mittelfeld von State-of-the-Art Lösungen aufhält, dar. Es gilt aber zu beachten, dass ein personalisiertes Model pro Teilnehmer trainiert und der Durchschnitt über alle Balanced-Accuracies gebildet wurde, was methodisch etwas fraglich ist, da ein Baseline Herzfrequenz des einen Teilnehmers durchaus auch der Stressbereich eines anderen sein könnte, aber auch die Möglichkeiten für personalisierte Stress-Modelle in den Raum stellt.

5 Maschinelles Lernen zur Klassifizierung von Stress

In diesem Abschnitt wird untersucht, welche Algorithmen zur Klassifizierung von Stress anhand physiologischer Marker gemessen durch Wearable Biosensors typischerweise zum Einsatz kommen. Gideon et al. führten dazu ein ausführliches Literatur-Review durch, aus welchem ich zwei Erkenntnisse hervorbringen möchte.

Bei Papern welche sich mit maschinellem Lernen zur Klassifizierung von Stress befassen trifft man immer wieder auf die gleichen öffentlichen Datensätze welche als Grundlage dienen, darunter auch der WESAD-Datensatz.

Unter den am häufigst vorkommend und gleichzeitig präzisesten Algorithmen befindet sich RF, weshalb dieser im folgendem näher untersucht wird. [VTSR23]

5.1 Random-Forest

RF ist ein Supervised-Learning Ensemble Verfahren, also eine Kombination von mehreren Modellen, in diesem Fall Decision Trees, welches lernen kann gelabelte Daten zu klassifizieren. Es wurde 2001 von Leo Breimann in dem Paper 'Random Forrests' erstmalig vorgestellt und hat aufgrund beispielsweise der Resistenz gegen Overfitting einen bleibenden Eindruck in der Welt des maschinellen Lernens hinterlassen.

Um dieses Ensemble Verfahren zu verstehen, benötigt es Wissen über die Funktionsweise von Decision Trees, dem grundlegenden Modell. Diese sind an sich bereits ein Modell, welches für Klassifizierungs- oder Regressionsaufgaben genutzt werden kann. Sie funktionieren, in dem sie Daten in einer Baumartigen Struktur organisieren, wobei jeder Knoten im Baum eine

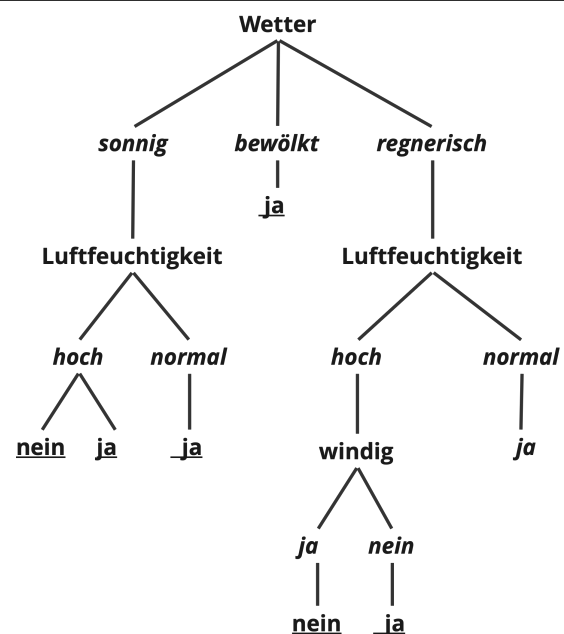
Entscheidung auf Basis eines Attributes trifft und jeder Ast zu einer möglichen Entscheidung oder einem Endergebnis (dem Blattknoten) führt. Die Aufteilung in Knoten und Äste basiert auf folgendem Prinzipien:

Ein Feature wird als Wurzelknoten gewählt, in dem der Information-Gain errechnet wird, einfach ausgedrückt: Welches Feature teilt den Datensatz am besten. Es werden pro Ausprägung des gewählten Features weiter Knoten erstellt, und für jeden dieser wiederum das Feature mit dem meisten Information-Gain als nächster Knoten gewählt.

Dies wird wiederholt, bis entweder alle Features aufgebraucht sind, alle Datenpunkte im aktuellen Subset zum gleichen Label gehören oder eine vorgegebene Tiefe erreicht wurde. Am Blattknoten eines jeden Astes befindet sich dadurch ein Knoten, welcher einer Ausprägung des Labels entspricht. So baut sich der Baum rekursiv auf, was aber beispielsweise ein mögliches Overfitting durch eine hohe Baumtiefe zur Folge haben kann.

Wetter	Luftfeuchtigkeit	Windig	Sport
sonnig	hoch	nein	Nein
sonnig	hoch	ja	Nein
bewölkt	hoch	nein	Ja
regnerisch	hoch	nein	Ja
regnerisch	normal	nein	Ja
regnerisch	normal	ja	Nein
bewölkt	normal	ja	Ja
sonnig	hoch	nein	Nein
sonnig	normal	nein	Ja
regnerisch	normal	nein	Ja
sonnig	normal	ja	Ja
bewölkt	hoch	ja	Ja
bewölkt	normal	nein	Ja
regnerisch	hoch	ja	Nein

(a) Beispieldatensatz mit Label = Sport



(b) Beispielbild des resultierenden Decision Trees

Abb. 2: Beispielhafte Darstellung Decision Tree

Abbildung 2 stellt dies simplifiziert und beispielhaft da. Weiterhin wird der ID3-Algorithmus, welcher zur Errechnung des Information-Gains durch Entropie benutzt wird offen gelassen und es wird auf die Quelle für diesen Abschnitt, ein grundlegendes Paper von J. Ross Quinlan, verwiesen. [Qui86]

Das Random-Forest Verfahren nutzt also diese Decision Trees und wendet weitere Techniken an, um deren Schwächen zu negieren, diese lauten unter anderem wie folgt:

- **Bootstrapping:** Aus dem ursprünglichen Datensatz werden durch Ziehen mit Zurücklegen mehrere neue Datensätze (Bootstrap-Samples) erstellt. Jeder dieser neuen Datensätze dient als Trainingsdatensatz für einen der Entscheidungsbäume.
- **Random Feature Selection:** Bei der Aufspaltung eines Knotens in jedem Entscheidungsbaum wird nicht der gesamte Satz von Features betrachtet, sondern nur eine zufällige Auswahl davon. Dies erhöht die Diversität der Bäume und verhindert, dass einzelne dominante Features alle Bäume gleich beeinflussen.
- **Training:** Jeder Entscheidungsbaum wird auf seinem jeweiligen Bootstrap-Sample und der zufälligen Auswahl von Features trainiert.
- **Aggregation:** Nachdem alle Entscheidungsbäume trainiert wurden, werden ihre Vorhersagen kombiniert. Bei Klassifikationsaufgaben wird in der Regel die Mehrheitsabstimmung verwendet, während bei Regressionsaufgaben der Durchschnitt der Vorhersagen genommen wird.

Dadurch erlangt das Verfahren einige Vorteile gegenüber der einzelnen Decision Tree Variante, wie der Resistenz gegen Overfitting, einer niedrigeren Anfälligkeit für Ausreißer und einer meist höheren Präzision als anderen herkömmlichen Klassifikatoren.

5.2 Fallstudie RF

Neben Siirtola et al. [Sii19] erzeugt RF auch im Paper von Perna et al. [GSDI21] die besten Ergebnisse. Es wird in diesem Abschnitt einmal Beispielsweise auf die Methodik dieses Papers eingegangen.

Verwendet wurde abermals der WESAD-Datensatz [SRD⁺18]. Aus diesem wurden nur Sensordaten des RespiBAN für den weiteren Verlauf verwendet, ergo unter anderem EKG-Signale.

5.3 Unsupervised

Ein im Vergleich relativ unerforschter Bereich ist die Anwendung von Unsupervised-Learning zur Klassifizierung von Stress, welches aber vorallem im Bezug auf Alltags-Wearables ausdrücklich attraktiv wirkt.

Kurz zusammengefasst stellen Unsupervised-Learning-Algorithmen eine Gruppe der Algorithmen des maschinellen Lernens ad, bei der ohne gelabelte Daten gearbeitet wird, um versteckte Muster oder Strukturen in den Eingabedaten zu entdecken, indem sie Ähnlichkeiten oder Unterschiede zwischen den Datenpunkten identifizieren und diese in Cluster, Assoziationen oder Hauptkomponenten einteilen.

Wie Iqbal et al. im für diesen Abschnitt benutzten Paper [IEWS22] bereits schrieb, ist die geringe Verfügbarkeit von hochwertigen gelabelten Datensätzen eine sehr große Herausforderung für die Entwicklung der Klassifizierungsmodelle. Dies spiegelt sich auch darin wieder, dass man immer auf die gleichen verwendeten Datensätze stößt. Weiterhin ist die Methodik der Erfassung der Labels, sei es durch Formulare im Nachgang oder durch beispielsweise TSST ein immernoch recht unentschlüsselter Punkt, da man immer auf Abweichungen in der Methodik der bereits verfügbaren Datensätze stößt.

Der Gedankengang keine gelabelten Daten zu benötigen, sondern stattdessen auf Unsupervised-Learning zu setzen ist also aufgrund dieses Punktes bereits einer näheren Untersuchung wert. Auf den Bezug zu Alltags-Wearables wird in Sektion 7 ausdrücklicher eingegangen.

Diesen Vorteilen muss man allerdings einem eventuellen Performance-Verlust gegenüber klassischen Supervised-Learning Modellen entgegenstellen, was durch Iqbal et al. untersucht wurde.

Dieser zog zwei bereits erwähnte öffentlich verfügbare Datensätze heran (SWELL-KW & Stress Recognition in Automobile Drivers) welche beide mehrere physiologische Marker beinhalten und gelabelt wurden. Es wurden für je Kategorie Unsupervised-/Supervised-Learning sieben Algorithmen trainiert und evaluiert, je mit Angabe der Hyperparameter. Es wurde jeweils auf Herzrate und Atmung einzeln und in Kombination trainiert.

Bei der Feature Auswahl ist anzumängeln, dass keine Begründung für die Wahl von nur einem beziehungsweise zweier Features angegeben wurde. Weitergehend ist auch der Prozess des Feature-Engineerings nicht nachvollziehbar, es wurde nur erwähnt, dass die Daten von Ausreißern bereinigt wurden, jedoch nicht ob die Rohen EKG Signale oder rausgerechnete Features verwendet wurden.

Das Resultat der Auswertung zeigte aber auf, dass Unsupervised-Learning Algorithmen durchaus im Schnitt der Accuracy schlechter performten, dies aber nur marginal und durchaus auch Algorithmen dieser Kategorie besser performen konnten als einige des Supervised-Learning.

Das Ergebnis stützt die mögliche Relevanz solcher Algorithmen für die Stress-Klassifizierung, weshalb in [IEWS22] wie auch hier aufgefordert wird, mehr in diese Richtung zu forschen. Dies wird ich auch in der nächsten Sektion widerspiegeln.

6 Proposal

Möchte auf Cortisol Detection im Schweiß (State-of-the-Art) aufmerksam machen, dass es sehhhhrrr schwierig ist anhand physiologischer Marker alleine ein generalisiertes Stress Model aufzubauen

ERGO -> Unsupervised Modelle zum antrainieren personalisierter Modell + Multisensorik + Biosensor

In dem Abschnitt auch Challenges aufzeigen.

7 Fazit

Literaturverzeichnis

- [All07] ALLEN, John: Photoplethysmography and its application in clinical physiological measurement. In: *Physiological Measurement* 28 (2007), 04, S. R1–39. <http://dx.doi.org/10.1088/0967-3334/28/3/R01>. – DOI 10.1088/0967–3334/28/3/R01
- [App23] Apple: *Your heart rate. What it means, and where on Apple Watch you'll find it.* <https://support.apple.com/en-us/HT204666>, 2023. – Zugriff am 20. April 2024
- [BGKD20] BENT, Brinnae ; GOLDSTEIN, Benjamin A. ; KIBBE, Warren A. ; DUNN, Jessilyn P.: Investigating sources of inaccuracy in wearable optical heart rate sensors. In: *NPJ Digital Medicine* 3 (2020), 18. <http://dx.doi.org/10.1038/s41746-020-0226-6>. – DOI 10.1038/s41746–020–0226–6
- [Boc21] BOCKSCH, R.: *Apple Watch - mit Abstand Nr. 1.* Digitales Bild. <https://de.statista.com/infografik/15047/geschaetzter-weltweiter-smartwatch-absatz/>. Version: 3 2021. – Zugriff am 20. April 2024
- [Boc23] BOCKSCH, R.: *Nutzung von Wearables in Europa deutlich gestiegen.* Digitales Bild. <https://de.statista.com/infografik/31378/nutzung-von-wearables-in-europa/>. Version: 12 2023. – Zugriff am 20. April 2024
- [Can15] CANNON, Walter B.: *Bodily Changes in Pain, Hunger, Fear and Rage: An Account of Recent Researches into the Function of Emotional Excitement.* New York : D. Appleton and Company, 1915
- [ESP⁺23] ERNST, Hannes ; SCHERPF, Matthieu ; PANNASCH, Sebastian ; HELMERT, Jens R. ; MALBERG, Hagen ; SCHMIDT, Martin: Assessment of the human response to acute mental stress-An overview and a multimodal study. In: *PLoS ONE* 18 (2023), Nr. 11, e0294069. <http://dx.doi.org/10.1371/journal.pone.0294069>. – DOI 10.1371/journal.pone.0294069. – ISSN 1932–6203
- [GK07] GOLDSTEIN, David S. ; KOPIN, Irwin J.: Evolution of concepts of stress. In: *Stress* 10 (2007), Nr. 2, 109–120. <http://dx.doi.org/10.1080/10253890701288935>. – DOI 10.1080/10253890701288935. – PMID: 17514579
- [GSDI21] GARG, Perna ; SANTHOSH, Jayasankar ; DENGEL, Andreas ; ISHIMARU, Shoya: Stress Detection by Machine Learning and Wearable Sensors. (2021), 43–45. <http://dx.doi.org/10.1145/3397482.3450732>. – DOI 10.1145/3397482.3450732. ISBN 9781450380188

- [HRS⁺18] HERNANDO, David ; ROCA, Surya ; SANCHO, Jorge ; ALESANCO, Álvaro ; BAILÓN, Raquel: Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects. In: *Sensors* 18 (2018), Nr. 8. <http://dx.doi.org/10.3390/s18082619>. – DOI 10.3390/s18082619. – ISSN 1424–8220

- [IEWS22] IQBAL, Talha ; ELAHI, Adnan ; WIJNS, William ; SHAHZAD, Atif: Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection. In: *Frontiers in Medical Technology* 4 (2022). <http://dx.doi.org/10.3389/fmedt.2022.782756>. – DOI 10.3389/fmedt.2022.782756. – ISSN 2673–3129

- [KJ23] KAISER, Marianne ; JAILLARDON, Laetitia: Pathogenesis of the crosstalk between reproductive function and stress in animals-part 1: Hypothalamo-pituitary-adrenal axis, sympatho-adrenomedullary system and kisspeptin. In: *Reproduction in Domestic Animals* 58 Suppl 2 (2023), 176-183. <http://dx.doi.org/10.1111/rda.14444>. – DOI 10.1111/rda.14444. – ISSN 1439–0531

- [LU24] LLAMAS, Ramon T. ; UBRANI, Jitesh: *Wearable Devices Market Insights*. International Data Corporation (IDC), 2024. – Accessed: 13.04.2024

- [MSA⁺17] MOZOS, Óscar M. ; SANDULESCU, Virginia ; ANDREWS, Sally ; ELLIS, David ; BELLOTTO, Nicola ; DOBRESCU, Radu ; FERRÁNDEZ, J. M.: Stress Detection Using Wearable Physiological and Sociometric Sensors. In: *International journal of neural systems* 27 2 (2017), 1650041. <https://api.semanticscholar.org/CorpusID:2934562>

- [Qui86] QUINLAN, J. R.: Induction of Decision Trees. In: *Machine Learning* 1 (1986), Nr. 1, 81–106. <http://dx.doi.org/10.1007/BF00116251>. – DOI 10.1007/BF00116251. – ISSN 1573–0565

- [Sel36] SELYE, Hans: A Syndrome produced by Diverse Nocuous Agents. In: *Nature* 138 (1936), 7, Nr. 3479, 32. <http://dx.doi.org/10.1038/138032a0>. – DOI 10.1038/138032a0

- [Sii19] SIIRTOLA, Pekka: Continuous stress detection using the sensors of commercial smartwatch. In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. New York, NY, USA : Association for Computing Machinery, 2019 (UbiComp/ISWC '19 Adjunct). – ISBN 9781450368698, 1198–1201

- [SMHS⁺21] SANTOMAURO, Damian F. ; MANTILLA HERRERA, Ana M. ; SHADID, Jamileh ; ZHENG, Peng ; ASHBAUGH, Charlie ; PIGOTT, David M. ; ABBAFATI, Cristiana ; ADOLPH, Christopher ; AMLAG, Joanne O. ; ARAVKIN, Aleksandr Y. u. a.: Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. In: *The Lancet* 398 (2021), Nr. 10312, 1700–1712. [http://dx.doi.org/10.1016/S0140-6736\(21\)02143-7](http://dx.doi.org/10.1016/S0140-6736(21)02143-7). – DOI 10.1016/S0140–6736(21)02143–7
- [SRD⁺18] SCHMIDT, Philip ; REISS, Attila ; DUERICHEN, Robert ; MARBERGER, Claus ; LAERHOVEN, Kristof V.: Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, 400–408
- [VTSR23] VOS, Gideon ; TRINH, Kelly ; SARNYAI, Zoltan ; RAHIMI AZGHADI, Mostafa: Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. In: *International Journal of Medical Informatics* 173 (2023), 105026. <http://dx.doi.org/https://doi.org/10.1016/j.ijmedinf.2023.105026>. – DOI <https://doi.org/10.1016/j.ijmedinf.2023.105026>. – ISSN 1386–5056
- [WLSD08] WANG, Jianli ; LESAGE, Alain ; SCHMITZ, N ; DRAPEAU, Aline: The relationship between work stress and mental disorders in men and women: Findings from a population-based study. In: *Journal of epidemiology and community health* 62 (2008), 02, S. 42–7. <http://dx.doi.org/10.1136/jech.2006.050591>. – DOI 10.1136/jech.2006.050591
- [WP01] WANG, Jianli ; PATTEN, Scott: Perceived work stress and major depression in the Canadian employed population, 20–49 years old. In: *Journal of occupational health psychology* 6 (2001), 11, S. 283–9. <http://dx.doi.org/10.1037//1076-8998.6.4.283>. – DOI 10.1037//1076–8998.6.4.283

Persönliche Angaben / Personal details

Keck, Paul

Familienname, Vorname / Surnames, given names

06.05.2002

Geburtsdatum / Date of birth

Bachelor Informatik

Studiengang / Course of study

00767721

Matrikelnummer / Student registration number

Eigenständigkeitserklärung***Declaration***

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und noch nicht anderweitig für Prüfungszwecke vorgelegt habe. Ich habe keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt. Die Arbeit wurde weder in Gänze noch in Teilen von einer Künstlichen Intelligenz (KI) erstellt, es sei denn, die zur Erstellung genutzte KI wurde von der zuständigen Prüfungskommission oder der bzw. dem zuständigen Prüfenden ausdrücklich zugelassen. Wörtliche oder sinngemäße Zitate habe ich als solche gekennzeichnet.

Es ist mir bekannt, dass im Rahmen der Beurteilung meiner Arbeit Plagiatserkennungssoftware zum Einsatz kommen kann.

Es ist mir bewusst, dass Verstöße gegen Prüfungsvorschriften zur Bewertung meiner Arbeit mit „nicht ausreichend“ und in schweren Fällen auch zum Verlust sämtlicher Wiederholungsversuche führen können.

I hereby certify that I have written this thesis independently and have not submitted it elsewhere for examination purposes. I have not used any sources or aids other than those indicated. The work has not been created in whole or in part by an artificial intelligence (AI), unless the AI used to create the work has been expressly approved by the responsible examination board or examiner. I have marked verbatim quotations or quotations in the spirit of the text as such.

I am aware that plagiarism detection software may be used in the assessment of my work.

I am aware that violations of examination regulations can lead to my work being graded as "unsatisfactory" and, in serious cases, to the loss of all repeat attempts.

Unterschrift Studierende/Studierender / Signature student

<ORT>, den 28.05.2024

Ort, Datum / Place, date