

Vorhabenbezeichnung: Cell-o: Ein smarter Stress Coach zur Krankheitsprävention
 Laufzeit des Vorhabens: 01.02.2021 - 31.06.2024 (nach Verlängerung)
 FKZ: 16SV8590, 16SV8591, 16SV8592, 16SV8593

Abschlussbericht

Verbundvorhaben:

Cell-o: Ein smarter Stress Coach zur Krankheitsprävention

Förderkennzeichen	Zuwendungsempfänger und Projektpartner:innen	Laufzeit
16SV8590	Universitätsklinikum Jena, Klinik für Psychiatrie und Psychotherapie, Philosophenweg 3, 07743 Jena, Prof. Dr. Martin Walter <i>sowie</i> Universitätsklinikum Tübingen, Eberhard Karls Universität Tübingen, Medizinische Fakultät, AG Innovative hirnfunktionelle Verfahren, Allgemeine Psychiatrie und Psychotherapie mit Poliklinik, Calwerstraße 14, 72076 Tübingen Prof. Dr. Birgit Derntl	01.02.2021 - 31.07.2024
16SV8591	Universität des Saarlandes, AG Probabilistic Learning, Saarland Informatics Campus, 66123 Saarbrücken, Prof. Dr. Isabel Valera	01.02.2021 - 31.03.2024
16SV8592	Fraunhofer Institut für Graphische Datenverarbeitung IGD, Joachim-Jungius-Straße 11 18059 Rostock, Mario Aehnelt	01.02.2021 - 31.01.2024
16SV8593	Thryve, mHealth Pioneers GmbH (Körtestr. 10, 10967 Berlin), Paul Burggraf	01.02.2021 - 31.07.2024

Kontaktperson Standortkoordinator: Prof. Dr. Martin Walter
 Name: Prof. Dr. Martin Walter
 Straße: Philosophenweg 3
 PLZ Ort: 07743 Jena
 E-Mail: martin.walter@med.uni-jena.de
 Telefon: +49 9390 101

Inhalt

Teil I – Kurzbericht	S. 3
1. ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischer Stand, an den angeknüpft wurde	S. 3
2. Ablauf des Vorhabens	S. 3
3. wesentliche Ergebnisse sowie ggf. Zusammenarbeit mit anderen Forschungseinrichtungen	S. 4
Teil II – Eingehende Darstellung im Vergleich zur ursprünglichen Vorhabenbeschreibung	S. 5
1. wichtigste Positionen des zahlenmäßigen Nachweises	S. 29
2. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten	S. 29
3. voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses – auch konkrete Planungen für die nähere Zukunft - im Sinne des fortgeschriebenen Verwertungsplans	S. 29
4. während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen	S. 30
5. erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF	S. 30

Teil I – Kurzbericht

1. ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Stress ist das zweithäufigste arbeitsbedingte Gesundheitsproblem, in Deutschland und EU-weit. Stressassoziierte Folgeerkrankungen sind bspw. Depressionen, Angst- und Panikzustände oder Burnout. Für ein besseres Stressmanagement sollte eine software- und sensorbasierte Gesundheitsanwendung entwickelt werden, welche mögliche Stressauslöser identifizieren und individuelle Selbsthilfe bei Stresserfahrung sowie personalisierte Gesundheitsvorhersagen mittels Sensoren kommerziell verfügbarer Wearables und Smartphones ermöglichen soll.

Wirksame Präventionskonzepte und auch deren Verbesserung erfordern eine wissenschaftliche Charakterisierung neuronaler, physiologischer und verhaltensbezogener Signaturen von Stress sowie wissenschaftliche Belege für die konkrete Wirksamkeit derartiger Präventionsprogramme. Unter dem großen Hauptziel der BMBF-Fördermaßnahme der Entwicklung von Systemen der Augmented Intelligence, um Menschen mithilfe von Methoden der Künstlichen Intelligenz (KI) bei Problemlösungen optimal zu assistieren [„Adaptive Technologien für die Gesellschaft – Intelligentes Zusammenwirken von Mensch und Künstlicher Intelligenz“ (MeKi)], war das Ziel des Cell-o-Projektes u.a. die Erfassung von individuellem Stress und die entsprechende präventive Verbesserung des Wohlbefindens und der mentalen Gesundheit via App-Smartwatch-System, um stress-bedingte psychische Gesundheitsprobleme vorzubeugen und zu reduzieren. Kommerzielle Smartwatches und die zu entwickelnde App sollten Informationen über die eigene Wahrnehmung, den körperlichen Zustand und das Verhalten erfassen und u.a. über entsprechende Feedback-Verfahren Rückmeldung an die Nutzer:innen geben. Durch die konkrete Mensch-Technik-Interaktion sollte auf diese Weise eine höhere Achtsamkeit erreicht werden, die das eigene Wohlbefinden positiv beeinflusst und Stress auf Dauer reduziert sowie verhindern kann. Smartphone-basierte Apps bergen ein großes Potenzial, Menschen in ihrer individuellen Stressbewältigung zu unterstützen und die Ursachenfindung von Stress und auch von Resilienz zu ermöglichen. Smartwatches und entsprechende Apps stehen jederzeit zur Verfügung und können in allen Situationen getragen und genutzt werden und so zur Charakterisierung neuronaler, physiologischer und verhaltensbezogener Signaturen von Stress genutzt werden. Darüber hinaus können auf diesem Weg wiederum Auswirkungen von Interventionen, wie z.B. Biofeedback-Training, in verschiedenen Altersgruppen und mit unterschiedlichen Lebensstilen erfasst und analysiert werden. Die zum Zeitpunkt der Antragstellung verfügbaren Stimmungs- und Stresserkennungssysteme wiesen jedoch für eine derartige Nutzung von App-Smartwatch-Systemen mit Genauigkeiten von 55-76% eine noch zu niedrige Präzision auf. Zudem wurde für die Mehrheit der verfügbaren Apps bis dato keine systematische Wirksamkeitsanalyse durchgeführt - nur zwei Apps lieferten Evidenz für deren methodische Wirksamkeit. Systeme oder Modelle, die sich auf Gesundheitsprognosen von Stress beziehen, gab es bis dahin ebenfalls nicht. Es sollten im Projekt Methoden und Technologien des maschinellen Lernens bzw. KI entwickelt und erprobt werden, mit denen die gesammelten (Vital)-Daten der Nutzer:innen u.a. auf Anomalien bzw. eine akute Belastungssituation analysiert werden können.

Ziel des Projektes war daher ein Demonstrator einer softwarebasierten Gesundheitsanwendung, die handelsübliche Wearables und Smartphones nutzt und mit folgenden Funktionen ausgestattet sein soll: Identifikator möglicher Stressauslöser, personalisiertes Feedback im Falle von Stress und personalisierte Stressvorhersagen.

2. Ablauf des Vorhabens

Die Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Jena (UKJ) fungierte als Projektkoordinator des Verbundprojektes. Die zugehörige Forschungsgruppe führte die entwickelten Softwarelösungen und Algorithmen zu einem App-System zusammen und führte entsprechende Experimente zur Intervention und Feldversuche mit der entwickelten App durch. Auch der Technologietransfer oblag den Kolleg:innen des UKJ. Im Verlauf des Projektes wurde die AG Innovative hirnfunktionelle Verfahren der Universität Tübingen (UKT) als Projekt-Partner in den Kooperationsvertrag aufgenommen. Das UKT hat an der Konzeption des

zugrundeliegenden BMBF-Antrags mitgewirkt und eine Feld-Studie zur Erhebung von Stresszuständen in Alltagssituation durchgeführt. Die AG Probabilistic Learning (PLG) unter der Leitung von Frau Prof. Dr. Isabel Valera von der Universität des Saarlandes entwickelte die für das Projekt notwendigen maschinellen Lernalgorithmen zur Stresserkennung. Das Fraunhofer IGD entwickelte und trainierte vorhandene Verfahren des maschinellen Lernens für die Erkennung von Vitalparametern und den individuellen Belastungszustand weiter und waren für die Vorverarbeitung der Daten, die Integration verschiedener Signaltypen und den Entwurf von Klassifikatoren verantwortlich. Die Kolleg:innen der mHealth Pioneers GmbH (Thryve) erfassten die Datenlage zur weiteren Analyse und integrierte die Daten in einen Software-Demonstrator. Thryve war darüber hinaus für den Funktionsumfang der Wearable-Programmierschnittstelle, den Funktionsumfang der Smartwatch-Erkennung und den Funktionsumfang des Back-Ends verantwortlich. Im Verlauf des Projektes ergaben sich Veränderungen, Anpassungen des Arbeitsplans und Verzögerungen (vgl. II „Eingehende Darstellung“), die jedoch aufgeholt oder anderweitig gelöst werden konnten. Regelmäßige Treffen, sowohl in Präsenz, als auch aufgrund damaliger Beschränkungen in digitaler Form mit allen Partnern wurden weitestgehend planmäßig organisiert und durchgeführt. Es erfolgten verschiedene große sowie auch kleinere Labor- und Feldstudien zu unterschiedlichen Schwerpunkten des Projektes, u.a. zur Charakterisierung von Stress oder zur Bereitstellung von Datensätzen für die Modellierung der mittelfristigen Stressvorhersage. Letzteres stand zu Projektende noch aus (s.u.), aber die Daten waren sehr wichtig für die Entwicklung, das Trainieren und die Testung von ML-Algorithmen für Stresserkennung.

Als Innovation im Projekt standen neben der Integration von erweiterter künstlicher Intelligenz (Augmented Intelligence, AI) und der Kooperative Mensch-Maschine-Interaktionen die geräte- und plattformunabhängige Stresserkennung, das Benutzer-Feedback, die personalisierte Stresserkennung und -vorhersage und die wissenschaftliche Belegbarkeit im Vordergrund.

3. wesentliche Ergebnisse sowie ggf. Zusammenarbeit mit anderen Forschungseinrichtungen

Eine erste funktionierende Version des Cell-o-Systems, der Algorithmen zur Stresserkennung und der App wurde erfolgreich entwickelt und ist funktionsfähig. Die Infrastruktur aus App, Server und Geräten erfüllt bereits die Datenschutz- und Datensicherheitsanforderungen und ermöglicht die Vernetzung mit Datenplattformen der großen Massenmarkthersteller. Sie unterstützt die Messung u.a. von EKG-, Aktivitäts-, Puls- oder Schlafdaten sowie deren Übertragung an den Server. Zudem können mit der App auch subjektive Daten, wie diverse EMA-Assessments (EMA: *engl. ecological momentary assessments*; Momentanbewertungen in realer Alltagsumgebung) zu flexiblen Zeiten an Endnutzer:innen gesendet werden.

Es wurde vollständige Echtzeit-ML-Pipeline für die Vorhersage von Stress unter Verwendung kommerzieller Smartwatch-Daten fertiggestellt, die die Ausführung von Algorithmen in Echtzeit, EMA-Anpassungen, Datenfilterung und Imputation sowie Benutzerfeedback umfasst. Darüber gehen aus den Projektarbeiten reichhaltige, qualitativ hochwertige, multimodale Datensätze hervor, die sowohl Labor- als auch Felddaten enthalten und die für die weitere Stressforschung und zur Analyse weiterer Fragestellungen zur Verfügung stehen. Es erfolgt bereits die Übertragung von Erkenntnissen in die Bearbeitung von laufenden oder geplanten Projekten (siehe II „Eingehende Darstellung“). Es wurde ein 3-stufiges Stresserkennungssystem entwickelt, das Smartwatches und maschinelles Lernen für die akute Stresserkennung kombiniert. Ziel ist es, Nutzer:innen über Unterschiede zwischen ihrem physiologischen Stresszustand und der subjektiven Wahrnehmung zu informieren. Die App nutzt EMA zur Erhebung subjektiver Daten.

Zukünftige Projekte planen die Integration der akuten Stresserkennung in die Vorhersage von chronischem Stress. Gespräche mit Krankenkassen wie TK, AOK und Barmer laufen, da Interesse an einem genauen Stresserkennungssystem besteht. Tests der Algorithmen zeigten, dass die geforderten Genauigkeiten bereits erreicht werden, allerdings nicht mit Smartwatches (s.u.). Weitere Tests und Evaluierungen zur Integration der Methoden in Gesundheitsangebote laufen. Internationale Forschungsgruppen sind ebenfalls an einer Zusammenarbeit bzgl. der Cell-o-Ergebnisse interessiert.

Teil II – Eingehende Darstellung (im Vergleich zur ursprünglichen Vorhabenbeschreibung).

Zentraler Innovationsansatz des Projektes war ein System, das menschliches Handeln und maschinelles Lernen für eine kooperative Interaktion zwischen Mensch und KI kombiniert. Die dadurch verbesserte Voraussage gesundheitlicher Entwicklungen und daraus resultierenden gezielten Interventionen sollten es Nutzenden ermöglichen, eine personalisierte Unterstützung im Umgang mit Stress zu erhalten. Eine erste funktionierende Version dieses Cell-o-Systems, eine erste Version der Algorithmen der Stresserkennung und eine erste Version der s.g. Pulsatio-App (Abb. 1) sind integriert und konzertiert. Im Ergebnis steht somit planmäßig die Infrastruktur aus App, Server und verbundener Geräte in einem funktionsfähigen Demonstrator (Abb. 2) bereit. Diese Infrastruktur erfüllt bereits jetzt die zentralen Anforderungen möglicher zukünftiger Verwertungspartner hinsichtlich des Datenschutzes und der Datensicherheit. Die Infrastruktur ermöglicht die Vernetzung zu den Datenplattformen der dominierenden Massenmarkthersteller, die zuverlässige Messung von EKG-, Aktivitäts-, Puls- und Schlafdaten sowie deren Übertragung an den Server. Es wurden Kalibrierungsansätze mit Ähnlichkeitsabgleich, Erkennung von Anomalien und Feinabstimmung zur Verbesserung der Personalisierung und Verringerung der Bereichsverschiebung vorgeschlagen und u.a. durch die PLG eine Forschungsarbeit über den Multi-level Stress Predictor (MuStP) verfasst, in der die verbesserte Leistung durch Personalisierung und Domänenanpassung nachgewiesen wurde (s.u.). Alle entwickelten Codes und Anleitungsdateien wurden mit allen Projektpartnern geteilt.

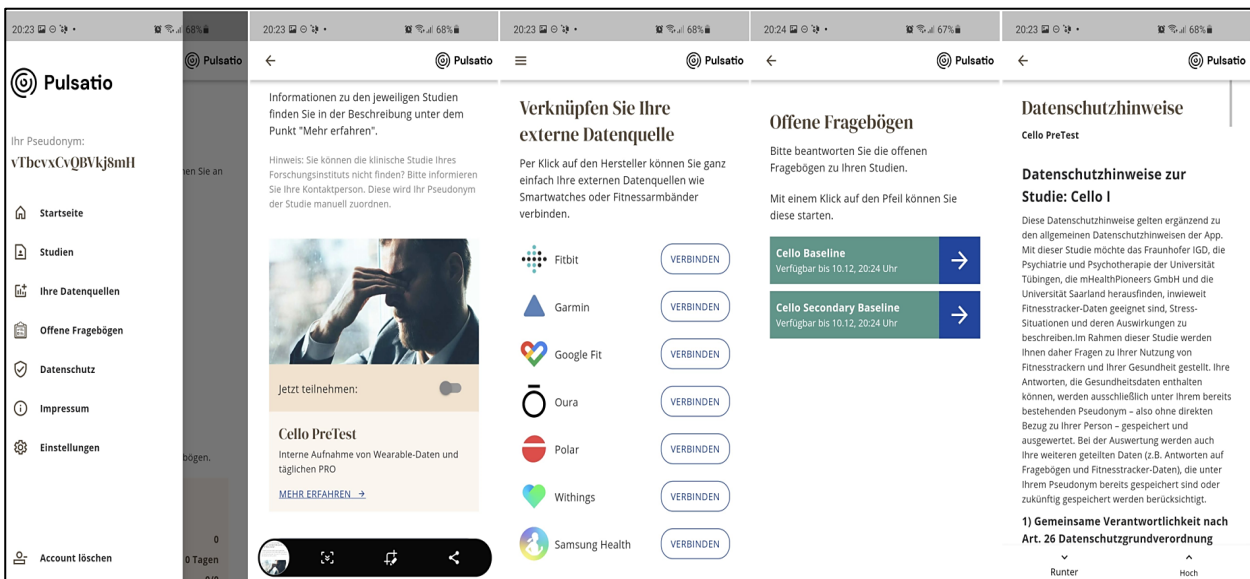


Abb. 1: Übersicht über die entwickelte App (Screenshots aus der Oberfläche der Pulsatio App); Thryve

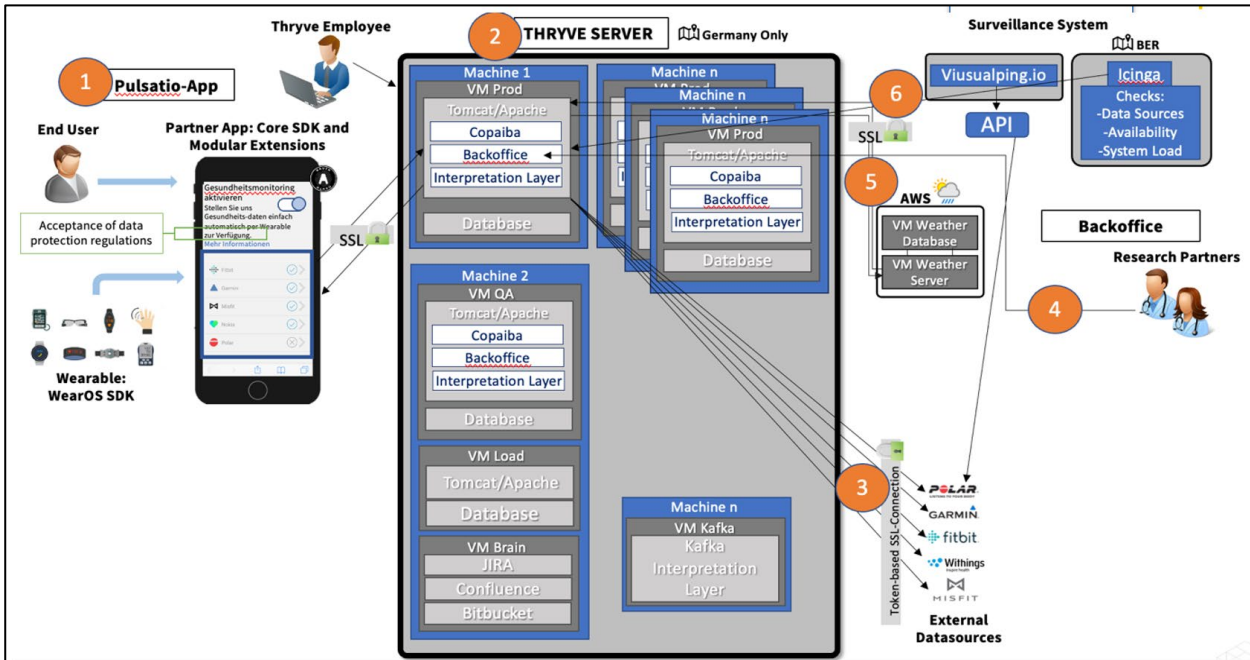


Abb. 2: Infrastruktur aus App, Server und verbundenen Geräten; Thryve

Weiterhin kann die zu diesem Zwecke entwickelte App über den Tag mehrere EMA-Assessments (EMA: engl. ecological momentary assessments; Momentanbewertungen in realer Alltagsumgebung) zu spezifischen Zeiten an die Endnutzer aussteuern (Abb. 3). Zeitpunkte, Häufigkeit, Fragetypen und Hinweistexte können dabei im Back-End konfiguriert werden, um auf neue Anforderungen und weitere Studienphasen flexibel reagieren zu können. Darüber gehen aus den Projektarbeiten große und kleine reichhaltige, qualitativ hochwertige, multimodale Datensätze hervor, die sowohl Labor- als auch Felddaten enthalten und die für die weitere Stressforschung und zur Analyse weiterer Fragestellungen zur Verfügung stehen. Es erfolgt bereits die Übertragung von Erkenntnissen in die Bearbeitung von laufenden oder geplanten Projekten (s.u.).

The screenshot shows the Pulsatio app interface for an EMA assessment. The screen is titled "Wie fühlst du dich gerade?" (How do you feel right now?) and "Was hat dich gestresst/ stressst Dich?" (What stressed you/stresses you?). The mood assessment is represented by a 10x10 grid with axes for "Hohe Erregung" (High arousal) vs. "Schläfrigkeit" (Drowsiness) and "Unangenehme Gefühle" (Unpleasant feelings) vs. "Angenehme Gefühle" (Pleasant feelings). Below the grid is a "WEITER" button. The stress assessment section lists several stressors with checkboxes: "Zwischenmenschliches (z.B. Streit, Missverständnisse)", "Druck (z.B. Leistung, Zeit, Verantwortung)", "Bewertung/Beurteilung durch andere", "Mangel an Kontrolle", and "Zu viel Input (z.B. Lärm, Menschen, neue Erfahrungen)". A "WEITER" button is also present at the bottom of the stress list.

Abb. 3: Bsp. EMA-Assessments in der Pulsatio-App; Thryve

Das Projekt hat wesentliche wissenschaftlich-technische Ergebnisse erzielt, die den ursprünglichen Zielen weitgehend entsprechen. Ein zentrales Ziel des Projekts war die Entwicklung eines Biofeedback-Systems zur Unterstützung der Stressbewältigung. Allerdings konnte dieses Ziel nicht vollständig erreicht werden. Es wurde ein Demonstrator entwickelt, der den Nutzern über Push-Benachrichtigungen auf dem Smartphone eine Rückmeldung zu ihren Stresszuständen bietet. Diese Funktion stellt eine reduzierte Version des ursprünglich geplanten Biofeedbacks dar, das in Echtzeit Feedback zur Stressbewältigung liefern sollte. Aufgrund technischer Herausforderungen, die außerhalb des Einflussbereichs des Projekts lagen, wie zum Beispiel Probleme bei der Integration von Geräten wie Withings, konnte die Echtzeit-Feedback-Funktion nicht wie geplant umgesetzt werden (s.u.). Zudem stellte sich heraus, dass die Umsetzung des Biofeedback-Systems als Medizinprodukt der Klasse 2B zu teuer für eine wirtschaftlich sinnvolle Vermarktung gewesen wäre. Eine Zulassung als Medizinprodukt dieser Klasse erfordert umfangreiche regulatorische Prozesse und damit verbundene hohe Kosten, die die Vermarktungsperspektiven erheblich einschränkt hätten. Diese Entscheidung hat Auswirkungen auf die ursprüngliche Zielsetzung, ein vollständig marktfähiges Biofeedback-System zu entwickeln. Das Biofeedback-Feature, das als Medizinprodukt zertifiziert werden könnte, stellt eine Hürde dar, da die regulatorischen Anforderungen für ein solches Produkt sehr hoch sind.

Ein weiteres zentrales Ziel war die kontinuierliche Erfassung von Vitaldaten zur Stressdetektion. Dies wurde erfolgreich umgesetzt. Die s.g. Pulsatio-App (s.o.) erfasst Vitaldaten wie die Herzratenvariabilität, Schlafmuster und Aktivitätslevel und ermöglicht es, daraus Rückschlüsse auf den Stresszustand der Nutzer:innen zu ziehen. Diese Funktion arbeitet einwandfrei und entspricht den ursprünglich angestrebten Zielen. Die App bietet eine benutzerfreundliche Oberfläche und ermöglicht es den Nutzer:innen, ihre Stresslevels kontinuierlich zu überwachen. Die Stressdetektion auf Basis dieser Vitaldaten stellt einen wichtigen Erfolg dar, da sie eine personalisierte und kontinuierliche Überwachung des Stresszustands ermöglicht. Ein zusätzlicher wichtiger Aspekt des Projekts war die Entwicklung einer robusten Datenschutzstrategie. Alle relevanten Datenschutzvorgaben, insbesondere die Datenschutz-Grundverordnung (DSGVO), wurden eingehalten. Die gesammelten Gesundheitsdaten wurden datenschutzkonform verarbeitet, was insbesondere bei den Feldstudien von entscheidender Bedeutung war. Die Teilnehmer:innen der Studien hatten die Gewissheit, dass ihre Daten geschützt sind, was zu einer hohen Teilnahmebereitschaft und einer zuverlässigen Datenerhebung führte. Datenschutz war ein zentraler Bestandteil der Entwicklung und wurde als wichtiges Vertrauen schaffendes Element in die App integriert.

Die technische Herausforderung lag in der Verlässlichkeit und Genauigkeit der erfassten Daten. Insbesondere bei der kontinuierlichen Messung der Herzratenvariabilität und anderen Vitalparametern traten Schwierigkeiten auf. Die Datenqualität war anfangs nicht immer optimal, was zu einer Verzögerung bei der Implementierung der Echtzeit-Feedback-Funktion führte. Die Daten aus den tragbaren Geräten waren in ihrer Qualität und Zugänglichkeit oft eingeschränkt, was die präzise Analyse und Nutzung in Echtzeit erschwerte. Die Stressdetektion selbst funktionierte jedoch gut, da sie auf diesen Daten basierte, jedoch konnte die Echtzeit-Rückmeldung nicht wie geplant erfolgen (s.u.).

Die PLG leitete die Entwicklung intelligenter und datengesteuerter Algorithmen des maschinellen Lernens (ML), um Nutzer:innen beim Verständnis, der Bewältigung und der Reduzierung von Stress durch eine plattformunabhängige, tragbare KI-Lösung zu unterstützen. Die Arbeit konzentrierte sich hier auf den Aufbau robuster Pipelines für die Stresserkennung, Stressvorhersage und Interventionsempfehlungen - jeweils zugeschnitten auf reale Daten von kommerziellen Smartwatches und validiert anhand kontrollierter Labordatensätze. Die Gruppe leistete einen umfangreichen Beitrag zur Analyse und Vorverarbeitung von diesen Datensätzen (z.B. WESAD; bestehende Datensätze aus Tübingen und Jena) und deckte dabei entscheidende Einschränkungen in den bestehenden Datenkettungsverfahren und eine Variabilität der Stressreaktionen auf. Diese Erkenntnisse wurden innerhalb des Konsortiums ausgetauscht, um die zukünftige Datenerfassung und die Versuchsplanung zu verbessern. Im Mittelpunkt des PLG-Beitrags stand die Entwicklung des MuStP (Multi-level Stress Predictor), eines personalisierten zweistufigen

ML-Modells, das Anomalie-Erkennung und EKG-basierte Stressklassifizierung kombiniert. PLG erkannte die signifikanten individuellen Unterschiede in der Stressreaktivität und die Verschiebung zwischen Labor- und Alltagssituationen, integrierte persönliche Anpassungstechniken (z. B. Feinabstimmung, Ähnlichkeitsabgleich) und gestaltete das System so, dass es mit niedrig- und hochauflösenden Signalen arbeiten kann, um die Belastung der Benutzer:innen zu reduzieren. Um die Interpretierbarkeit und das Engagement der Nutzer:innen zu verbessern, wurden aufmerksamkeitsbasierte Modellierungs- und Post-Hoc-Erklärungswerkzeuge untersucht und Dual-Mode-EMA-Prompting-Strategien entwickelt, um die Datenerfassung mit der Benutzerfreundlichkeit in Einklang zu bringen. PLG trug auch zur Entwicklung der Software und des Versuchsaufbaus bei, kodierte die gesamte Datenerfassungspipeline und arbeitete eng mit den Partnern zusammen, um zu definieren, wie die EMAs abgefragt, gespeichert und in der Modellierung verwendet werden sollten. Durch statistische Kalibrierung, Korrelation mit Cortisolmessungen und iterative Zusammenarbeit mit klinischen Expert:innen stellte PLG sicher, dass die Modellvorhersagen sowohl mit physiologischen als auch mit psychologischen Stressmarkern übereinstimmen. Letztlich legte die Arbeit der PLG die technische und methodische Grundlage für den Einsatz personalisierter, adaptiver und erklärbarer Stressvorhersagealgorithmen in der Praxis - ein Schlüssel zur Verwirklichung der langfristigen Vision des Projekts einer intelligenten, nutzerzentrierten Gesundheitsunterstützung.

Im Vorhaben-Verbund wurde die Einbindung der Sensorik für Pretests, klinische Aufnahmen und mögliche folgende Studienphasen besprochen. Dabei wurde insbesondere Wert daraufgelegt, ein System zu erproben, das in der Stressprävention mit derzeit verbreiteten Wearables arbeitet und keine zusätzliche Spezialsensorik benötigt. Dazu wurde die flexible Studien-App Pulsatio (s.u.) entwickelt, die die Einbindung gängiger Massenmarkt-Wearables und flexibel ausgestaltbarer Stress-Assessments mittels Fragebogen ermöglicht. Ziel dabei war es, dass diese App auf die Erkenntnisse der einzelnen Projektphasen flexibel reagieren kann, um für die nächste Studienphase geeignete EMA-Assessments anzubieten und einzuholen. Bei der Signalauswahl bestand die zusätzliche Herausforderung darin, dass mehrere Massenmarkthersteller, z.B. FitBit und Apple den Zugang zu PPG-Rohsignalen weitestgehend gesperrt haben und diese keinen Dritt-Apps mehr zugänglich machen. Dies hat nur sekundär wettbewerbstechnische Gründe, primär sehen sich die Hersteller außerstande, die Qualität der Messung außerhalb exakt auf den Anwendungsfall zugeschnittener und aufwendig kontrollierter Messbedingungen bereit zu stellen und in Einklang mit Ihrer Zertifizierung zu bringen. In zahlreichen externen Tests wurde festgestellt, dass die Verlässlichkeit optischer Messungen der HRV in nicht-kontrollierten Alltagssituationen mit aktuellen Geräten der Massenmarkthersteller nicht ausreichend zuverlässig ist. Im Gegenzug haben weit verbreitete Wearables der Massenmarkthersteller zunehmend EKG-Aufnahmemöglichkeiten, wobei ein hochqualitatives Signal gegeben ist und weniger Probleme durch fehlerhafte Messsituationen möglich sind. Die darauf aufbauend entwickelten Algorithmen können später ebenfalls auf optische Pulssignale erweitert werden (die Algorithmen nutzen bereits Herzfrequenz-Signale in Levels 1 & 2), sollte sich der Zugang zu diesen Signalen verbessern – zu Projektende gab es noch keinen Zugang zu kontinuierlichen und Rohdaten von PPG-Sensoren. Dazu wurde das Auslösen, Durchführen und die Datenübertragung von EKG auf Geräten von Withings und Apple für das Projekt entwickelt. Besonderes Augenmerk galt ebenfalls der datenschutzkonformen Einbindung dieser Datenquellen, sodass der Abruf der Informationen auf möglichst datensparsame Weise erfolgt. Dazu wurde eine Datenschutzfolgeabschätzung und eine Datenschutz-FAQ zum Produkt erstellt. Dies ist insbesondere mit dem Ziel einer möglichen Anschlussverwertung in Präventionsprogrammen der deutschen gesetzlichen Krankenversicherungen geschehen, die sehr hohe Datenschutzerfordernisse für eine Implementierung stellen. Die Studien-App mit der Cell-o-Studie wurde im Google Play Store und im App Store veröffentlicht und sammelt zuverlässig die gewünschten Daten. Die Fragebogenfunktionalität wurde nach der ersten Iteration in Absprache mit den Projektpartnern auf zweidimensionale visuelle Skalen erweitert. Die Serverarchitektur und die Datenübertragungswege wurden gemäß der initialen Planung umgesetzt. Die PLG arbeitete an Methoden für generative und diskriminative Modelle für multimodale Zeitreihendaten wie Signalmessungen von

Wearables. Es muss hierbei betont werden, dass der ursprüngliche Plan für die PLG darin bestand, Methoden für kontinuierliche Daten zu verwenden, da davon ausgegangen wurde, dass das Konsortium Zugriff auf die Rohdaten erhalten würde. Die PLG mussten die Pläne ändern und geeignete Methoden für die „halbkontinuierliche und kurze Natur“ der Smartwatches finden und verwenden. Ziel war es, diese Methoden so zu entwickeln, dass sie höchste Genauigkeit bei Aufgaben wie Stresserkennung, -vorhersage und -management bieten und gleichzeitig interpretierbar, robust und effizient sind. Diese Algorithmen wurden schließlich anhand von Datensätzen trainiert, die entweder im Rahmen von Laborstudien oder durch die von den Cell-o-Projektpartnern entwickelte mobile Anwendung (Pulsatio-App plus Smartwatch) in Alltagssituationen gesammelt wurden. Zu Beginn wurde davon ausgegangen, mit kontinuierlichen und Rohdaten zu arbeiten, um die ML-Algorithmen zu entwickeln und zu trainieren. Nachdem die Partner des Fraunhofer IGD und der PLG einige Zeit mit Aufzeichnungen von kommerziellen Smartwatches gearbeitet hatten, stellten sie fest, dass nur auf halbkontinuierlich und auf heruntergesampelte Versionen der Signale zugegriffen werden konnte. Die anderen Signaltypen wurden manuell ausgelöst und führten nur zu kurzen Datensegmenten. Darüber hinaus ändern digitale Apps und kommerzielle Smartwatches ständig die Regeln für den Zugriff auf Informationen und Signale für Wissenschaftler:innen. Derzeitige kommerzielle Smartwatches verfügten nur über begrenzte Datenerfassungsmöglichkeiten in Bezug auf die Granularität und Qualität der erfassten Signale. Genauer gesagt konnten die handelsüblichen Smartwatches nur eine begrenzte Anzahl von Signaltypen in einem Streaming-Verfahren mit einer relativ geringen Auflösung (etwa 10 Minuten) erfassen. Diese Einschränkung stellt eine große Hürde für die Ausführung präziser Algorithmen zur Erkennung von Anomalien dar, mit denen EMAs ausgelöst oder Stressepisoden erkannt werden können. Aus den Erfahrungen mit kommerziellen Smartwatches wurde geschlossen, dass sie nur eine begrenzte Anzahl von Signalen messen. Es besteht jedoch zukünftig auch die Möglichkeit, zusätzlich zur Smartwatch kommerzielle oder forschungsbasierte (Labor-)Sensoren zu verwenden, um wichtige Signale in der Stressforschung, wie z.B. die elektrodermale Aktivität (EDA) aufzuzeichnen. Es ist auch wichtig zu erwähnen, dass längere, umfangreichere Feldstudien notwendig sind, um genügend Daten für bestimmte leistungsfähige Algorithmen des maschinellen Lernens zu liefern.

Die PLG führte Forschungen zur Abschwächung des Domain-Shift-Problems durch, das auftritt, wenn Modelle des maschinellen Lernens mit Daten aus einer bestimmten Verteilung trainiert werden und dann zur Vorhersage mit Daten aus einer anderen Verteilung verwendet werden. Dieses Problem hängt auch mit der Personalisierung der Modelle zusammen, da die Verteilung der Daten für jede:n Nutzer:in einzigartig ist. Die PLG schlug zwei Hauptwege zur Entschärfung der Probleme vor, die sich als trainingsbasierte oder auf Verstärkungslernen basierende Ansätze zur Entschärfung von Domänenverschiebungen und/oder zum Erreichen von Personalisierung zusammenfassen lassen. Nach der Diskussion mit dem Konsortium konzentrierte sich PLG auf den trainingsbasierten Ansatz, d.h. die Feinabstimmung der Modelle. Nach der Entwicklung von MuStP, dem hierarchischen Modell für maschinelles Lernen (s.u., Abb. 4), das aus drei Ebenen besteht und von PLG im Rahmen des Projekts vorgeschlagen und entworfen wurde, entwickelte PLG Skripte für die Feinabstimmung, die für die Personalisierung, d.h. die personalisierte Stressvorhersage, entscheidend ist. PLG entwickelte zwei verschiedene Personalisierungsschemata: Training eines Pools von Modellen zur Erkennung von Anomalien und Auswahl des Modells, das den Nutzer:innen im Alltag am ähnlichsten ist, oder Feinabstimmung der Modelle anhand benutzerspezifischer Daten. Die PLG hat die Grundalgorithmen von MuStP finalisiert und sichergestellt, dass sie durch Training reproduzierbar sind. Um die Anwendung des vorgeschlagenen Modells im Alltag mit Smartwatches zu ermöglichen, entwickelte PLG eine Echtzeit-ML-Pipeline zur Stressvorhersage. Diese Implementierung umfasst die Ausführung der entwickelten Algorithmen auf Echtzeitdaten von kommerziellen Smartwatches, die Anpassung der Erfassungszeitpläne für die ökologische Momentaufnahme (EMA) an feste oder modellabhängige Versionen, das Herausfiltern von Messungen, die mit nicht bevorzugten Aktivitäten wie Bewegung oder Schlaf verbunden sind, die Durchführung von Datenimputationen, wenn Messungen fehlen, und die abendliche Rückmeldung/Zusammenfassung der Stressvorhersagen an die

Nutzer:innen. Nach Abschluss der Modellentwicklung führte PLG Simulationen mit historischen Daten durch, die von kommerziellen Smartwatches gesammelt wurden und berichtete dem Konsortium über das Verhalten der Modelle und lieferte Erkenntnisse darüber, wie man die Modelle abstimmen kann, insbesondere im Hinblick auf den Kompromiss zwischen hoher Wiedererkennung und Präzision bei der Festlegung von Entscheidungsschwellen. Da nicht genügend Daten für hochauflösende Herzfrequenzmessungen gesammelt wurden, entschied sich PLG für die vorgeschlagene Pipeline mit zwei Ebenen, d.h. niedrig aufgelöste Herzfrequenz und hochauflösendes EKG. Die PLG hat eine Forschungsarbeit über die Leistung des vorgeschlagenen Modells MuStP verfasst. MuStP wurde entwickelt, um den Aufwand für die Nutzer:innen zu minimieren, da es auf der ersten Ebene mit der niedrig aufgelösten Herzfrequenz arbeitet und nur bei Bedarf EKG-Messungen (durch Benachrichtigungen in der App) erfordert. Das vorgeschlagene Modell wurde zunächst anhand von Labordaten trainiert (s.o.), die mit dem TSST-Verfahren erhoben wurden. Anschließend wurde es mit Daten der kommerziellen Withings-Smartwatch und EMAs für den täglichen Gebrauch angepasst. Um Domänenverschiebungen zu berücksichtigen und eine personalisierte Stressvorhersage zu erreichen, wurde eine Post-hoc-Optimierung und Feinabstimmung angewendet, d.h. eine Modellübertragung. In der Ablationsstudie (s.u.) wurden die Auswirkungen des Modelltransfers auf die Leistung veranschaulicht. Empirische Ergebnisse auf Smartwatch-Daten zeigen, dass MuStP eine F1-Punktzahl von 52 % erreicht, eine signifikante Verbesserung gegenüber dem Modell ohne Personalisierung oder Übertragung, trotz der geringen Anzahl von Messungen, die als Stress gekennzeichnet sind. Die PLG reichte die Forschungsarbeit bei einem Workshop der International Conference on Machine Learning (s.u.) und bei einem weiteren Workshop des International Symposium on Wearable Computing ein.

Der gesamte von der PLG entwickelte oder korrigierte Code wurde dem Konsortium über GitHub mit begleitenden Readme-Dateien für die weitere Nutzung zur Verfügung gestellt.

In intensiven Treffen und Diskussionen zwischen den medizinischen und technischen Projektmitgliedern wurde schließlich das 3-stufige Stresserkennungssystem (Abb. 4) entwickelt und finalisiert, das sowohl die Stärken der digitalen App und kommerzieller Smartwatches, als auch modernste Ansätze des maschinellen Lernens für den Umgang mit der akuten Stresserkennung in einem Stressüberwachungssystem nutzt. Das dreistufige Stresserkennungssystem stellt bei Bedarf den Kontext zu den Verhaltensdaten her und bezieht schlussendlich auch subjektive Daten ein.

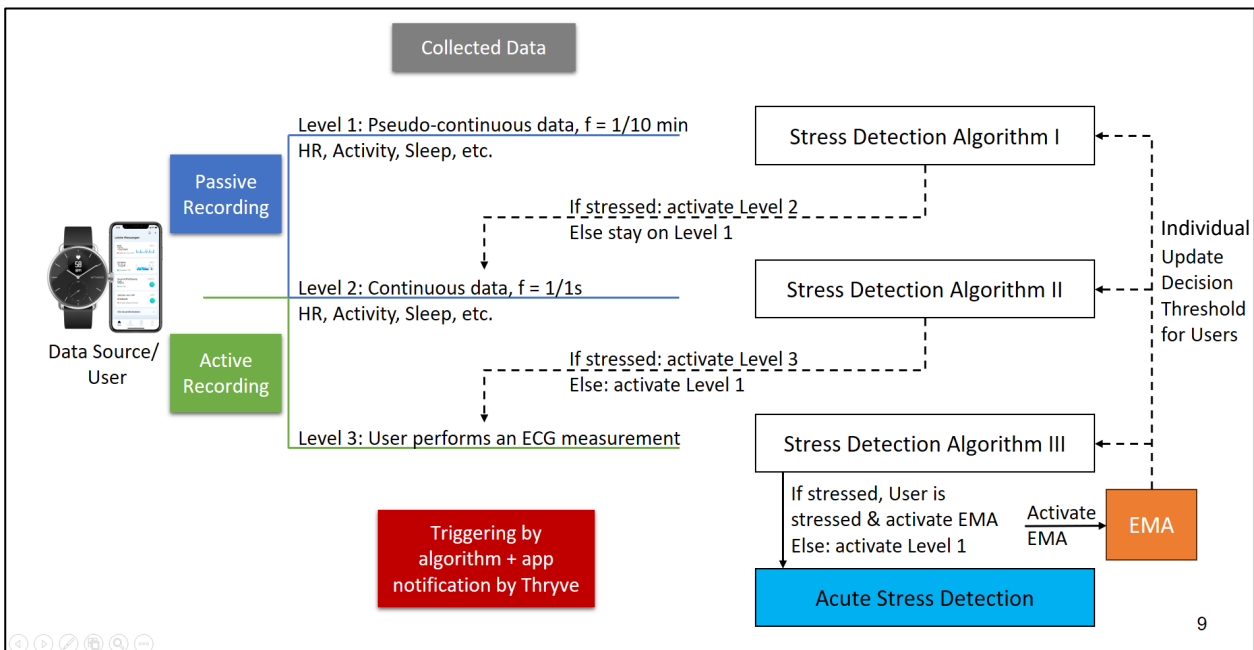


Abb. 4: 3-stufiges Stresserkennungssystem MuStP (PLG)

Das hierarchische maschinelle Lernmodell in drei Ebenen wurde durch die PLG so konzipiert, dass es die meisten der kommerziellen Smartwatch-Daten nutzt, ohne die Nutzer:innen zu belasten (erste, zweite Stufe), es sei denn, es besteht eine hohe Wahrscheinlichkeit von Stress (dritte Stufe). Das Modell nutzt niedrig aufgelöste Herzfrequenzmessungen in der ersten Ebene sowie Verhaltensdaten als Kontext, hoch aufgelöste Herzfrequenzmessungen in der zweiten Ebene und hoch aufgelöste EKG-Messungen in der dritten Ebene, um Stress zu erkennen. Die erste Ebene ist ein baumbasiertes Modell zur Anomalie-Erkennung (Entscheidungsbäume für prädiktive Modellierung), während die zweite und dritte Ebene Deep-Learning-Classifier sind. Da die Datenerfassung noch nicht abgeschlossen war, wurden vorhandene Bestandsdaten (Legacy-Dataset) analysiert und genutzt, um die Leistung des vorgeschlagenen Modells mit den Projektpartnern zu bewerten. Die PLG erstellte Backbone-Modelle für jede Ebene und übermittelte sie an die Kollege:innen des Fraunhofer IGD, damit diese weiter an der Trainingsleistung arbeiten konnten. Da im Rahmen der Feldstudien weiterhin Daten aus dem alltäglichen Umfeld gesammelt wurden, wurde alle zwei Wochen eine gründliche Überprüfung der Datensätze aus den Feldstudien durchgeführt, um fehlende Daten, Benutzerfehler und Unstimmigkeiten zu ermitteln. Um die Anwendung des vorgeschlagenen Modells im Alltag mit Smartwatches zu ermöglichen, begann die PLG mit der Arbeit an einer Echtzeit-ML-Pipeline zur Stressvorhersage. Der wichtigste Beitrag hierbei war, dass die Echtzeit-ML-Pipeline skalierbar und schnell sein sollte (es gibt keine Hardware, die in Echtzeit arbeiten kann), damit sie mit den Daten einer/eines einzigen Nutzer:in ausgeführt werden kann, was als Entwicklung des Endprodukts angesehen werden kann. Parallel erfolgte mit allen Projektpartnern die Implementierung von Daten-Imputation und Soft-Label-Extraktion aus Ecological Momentary Assessment (EMA).

Nachdem die gesammelten Daten ein beträchtliches Ausmaß erreicht hatten, führte die PLG Simulationen mit „alten“ Smartwatch-Daten durch (s.o.), um die anfängliche Leistung des 3-Level-Modells (MuStP) zu bewerten (Hinweis: es wurden nur PPG-Daten simuliert, EKG-Daten wurden direkt verwendet.) Diese Simulationen lieferten wertvolle Einblicke in das Verhalten des Modells und erleichterten die Abstimmung und Auswahl von Entscheidungsschwellen. In Anbetracht der Herausforderungen, die sich bei spärlichen Daten ergeben, weil die gesammelten Daten unzureichend sind, wurden Techniken zur Erzeugung synthetischer Datensätze implementiert, die eine robuste Modelleistung auch in weniger datenreichen Umgebungen gewährleisten.

Als wichtigen Teil der Datenanalyse untersuchte die PLG die statistischen Unterschiede zwischen Messungen im Labor und im Alltag, da unterschiedliche Geräte verwendet werden. Während zum Beispiel Herzfrequenzmessungen direkt mit verschiedenen Geräten aus jeder Umgebung verglichen werden können, gibt es einen signifikanten Unterschied in den Signaleigenschaften der EKG-Daten zwischen den beiden Umgebungen. Aufgrund dieser Beobachtung wurden die EKG-Messungen auf eine feste Skala skaliert, so dass sie in einer einheitlichen Modellpipeline verwendet werden können. Dies ermöglichte es, Modelle, die mit Labordaten trainiert wurden, im Alltag anzuwenden, was zum nächsten Schritt führt, Modelle mit Labordaten zu trainieren und sie in Zukunft auf reale Daten abzustimmen und anzuwenden.

Es wurden semikontinuierliche und passive batteriechonende Herzfrequenzmessungen, gelegentliche batterieintensive hochauflösende Herzfrequenzmessungen und modernste EKG-Messungen angewendet, um physiologische Daten für das Stresserkennungssystem im Alltag mithilfe von kommerziellen Smartwatches zu sammeln. Ziel des Systems war und ist es, die Nutzer:innen über ihren physiologischen Stresszustand und ihre subjektive Stresswahrnehmung zu informieren, um sie auf eventuelle Unterschiede zwischen ihrem physiologischen Stresszustand und ihrer subjektiven Stresswahrnehmung aufmerksam zu machen und sie über ihr Stressniveau zu informieren. Die subjektiven Daten wurden mit der entwickelten Pulsatio-App über EMA erhoben. Die App verfügt über zeitplan- und ereignisgesteuerte EMA-Pläne, die mit dem Cello-System zur Stressüberwachung verwendet werden können. Außerdem wurde mit der Vorhersage chronifizierenden Stresses anhand der im Cello-Projekt gesammelten multimodalen Datensätze begonnen und es wurden erste gute Ergebnisse erzielt. In zukünftigen Projekten können und sollten die Ergebnisse des akuten Stresserkennungssystems in das Verfahren zur Vorhersage

chronifizierenden Stresses integriert werden. Die relevanten Experimente bzw. ursprünglichen Studien im dritten und letzten Förderjahr zur Datenaufnahme und -analyse konnten abgeschlossen werden, jedoch werden nach Projektende weitere Tests und Datenerhebungen zur Testung der Modelle durchgeführt.

Als Innovation im Projekt standen neben der Integration von erweiterter künstlicher Intelligenz (Augmented Intelligence, AI) und der Kooperative Mensch-Maschine-Interaktionen die geräte- und plattformunabhängige Stresserkennung, das Benutzer-Feedback, die personalisierte Stresserkennung und -vorhersage und die wissenschaftliche Belegbarkeit im Vordergrund. Das kooperative Vorgehen von Mensch und Technik begründete sich in der gegenseitigen Beeinflussung der verschiedenen Ansätze und implementierten Funktionen durch Integration von Augmented Intelligence (AI). Als besondere Stärke ist hier die Aktivierung der App-basierten Einschätzungen durch passive und kontinuierliche Beobachtungen zu nennen, wodurch diese automatisch in ihrer Genauigkeit optimiert werden und ferner direkt zu einer indizierten Intervention führen sollten.

Regelmäßige Treffen, sowohl in Präsenz, als auch aufgrund damaliger Beschränkungen in digitaler Form mit allen Partnern wurden weitestgehend planmäßig organisiert und durchgeführt. Auch regelmäßige monatliches Treffen mit den Doktorand:innen wurden durchgeführt, um sicherzustellen, dass die korrekten Daten und Rückmeldungen der Mitarbeiter:innen zu Stress, Stress-Merkmalen, Interpretation der Ergebnisse und Veröffentlichungen zur Verfügung stehen. Es wurde eine Analyse der neurowissenschaftlichen und psychiatrischen Literatur für die Auswahl von Vitalsignalen und Merkmalen als Features für die zukünftige Entwicklung von Algorithmen aus dem Bereich des maschinellen Lernens durchgeführt. Es wurden im Verlauf des Projektes zahlreiche potentielle Geräte in den Fokus genommen, die für die hohen Anforderungen des Projektes geeignet erschienen. Nur wenige kommerzielle Smartwatches erfüllten die Datenschutzanforderungen und die Bedenken der Ethikkommissionen. Weiterhin wurden Anzahl und Art der Sensoren und Signale, auf die von der Smartwatch zugegriffen werden kann, diskutiert und definiert. In Bezug auf die App wurden die Schwerpunkte Datenlöschung, Datenschutzerklärung und die Implementierung der Cello-Studienprotokolle für die Labor- und Feldstudien in der App und der Datenzugriff für andere Projektpartner auf die von Smartwatches gesammelten Daten im Rahmen des Cello-Projekts besprochen. Es wurden u.a. Pilotstudien mit drei Proband:innen über drei Monate und mit drei Smartwatches unterschiedlicher Marken durchgeführt. Auf Basis der umfangreichen Testungen und der ausführlichen Evaluation der geeigneten Smartwatches wurde schließlich die Withings Scanwatch (französisches Unternehmen mit Hauptsitz und Servern in Frankreich) für die Labor- und Feldstudien für das Projekt ausgewählt und angeschafft. Es folgten verschiedene große sowie auch kleinere Laborstudien zu unterschiedlichen Schwerpunkten des Projektes, u.a. zur Charakterisierung von Stress, sowie mehrere Pilotfeldstudien und Feldstudien, z.B. zur Bereitstellung von Datensätzen für die Entwicklung und Testung von Algorithmen zur akuten Stresserkennung und für die Modellierung der mittelfristigen Stressvorhersage. Eine ursprünglich geplante Pilotstudie konnte durch die Analyse bereits vorhandener stressbezogener Daten aus alten und aktuellen Projekten ersetzt werden.

Es wurde im Projekt entschieden, die Leistung der generischen (nicht personalisierten) ML-Modelle im Rahmen des Cello-Projekts weiter zu verbessern und die Ergebnisse der verbesserten Modelle zu evaluieren und sie u.a. interessierten Krankenversicherungen in Deutschland und der EU vorzustellen. Die Unternehmen, denen es nicht erlaubt ist, ihre ML-Modelle auf der Grundlage der Daten ihrer Nutzer:innen zu personalisieren, bevorzugen die Verwendung guter generischer ML-Modelle. Personalisierte ML-Modelle waren eine Priorität für dieses Projekt. Als jedoch deutlich wurde, dass bei diversen Unternehmen und Forschungspartnern gute generische ML-Modelle gefragt sind, wurde entschieden, die Leistung der Modelle durch Modellpersonalisierung zu verbessern, jedoch als sekundäres Ziel. Es wurden (und werden auch weiterhin) mehrere ML-Techniken für die Modellpersonalisierung evaluiert. Die Herausforderung besteht darin, leistungsfähige Methoden zu finden, die nicht zu viele Benutzerdaten benötigen und dennoch die ML-Modelle mit einer akzeptablen Geschwindigkeit personalisieren und aktualisieren können.

Es haben Gespräche aller Projektpartner über die Zusammenarbeit mit der Techniker Krankenkasse (TK), der Allgemeinen Ortskrankenkasse (AOK) und der Barmer Krankenkasse stattgefunden, weiterführende Gespräche über potentielle Kooperationen sind angelaufen. Die deutschen Krankenkassen sind an einem einfachen und genauen Stresserkennungssystem interessiert. Darüber hinaus kann Cell-o zukünftig möglicherweise die Erkennung von chronischem Stress und die Vorhersage von chronifizierendem Stress anbieten (s.o.), um bestehende Ansätze zur Erkennung von akutem Stress zu ergänzen. Es wurden mehrere Tests der Algorithmen durchgeführt, um zu prüfen, ob die von den obengenannten Krankenkassen geforderten Genauigkeiten erreicht werden. Alle Tests wurden erfolgreich durchgeführt. Es werden zukünftige zusätzliche Tests von relevanten Merkmalen für maschinelles Lernen für Felddaten und weitere Evaluierungen durchgeführt, um die Anforderungen von Krankenkassen zu erfüllen und die im Projekt entwickelten Methoden in ihre Gesundheitsangebote und digitalen Apps zu integrieren. Es finden Gespräche und inhaltliche Diskussionen mit anderen nationalen und internationalen Forschungsgruppen statt, die daran interessiert sind, ihre Methoden anhand bestehender Datensätze aus den Cello-Projekten zu testen und gemeinsame Publikationen über die Ergebnisse zu verfassen; die Vorbereitungen dazu laufen. Im Verlauf des Projektes ergaben sich u.a. aufgrund von Corona-Beschränkungen, größeren (auch personalen) Umstrukturierungen an den verschiedenen beteiligten Institutionen und durch die großen Abhängigkeiten der Projektpartner untereinander bzgl. ihrer jeweiligen Arbeitspakete Veränderungen, Anpassungen des Arbeitsplans und Verzögerungen (vgl. II „Eingehende Darstellung“), die jedoch mithilfe einer engmaschigen Kommunikation mit allen Projektpartnern und gezielter Umstrukturierung innerhalb der Arbeitspakete aufgeholt oder anderweitig gelöst werden konnten – ausstehende Studien wurden zeitnah durchgeführt bzw. entsprechend angepasst oder ersetzt (s.u.). Größere Verzögerungen im Projekt wurden u.a. durch die sehr umfangreiche und langwierige Prüfung datenschutzrechtlicher und regulatorischer Auflagen (insbesondere an den großen Universitätskliniken) bedingt. Es erfolgte eine Datenschutzfolgeabschätzung und es wurden entsprechende Datenschutz-FAQ sowie Gerätezertifikate und Serverzertifizierungen sowie eine vollständige Beschreibung der IT-Infrastruktur für jeden Partner durch Thryve zur Pulsatio-App zur Verfügung gestellt.

Die Zusammenarbeit mit anderen neuen bzw. geplanten Projekten, u.a. mit dem DZPG (Deutsches Zentrum für Psychische Gesundheit), der WATCH-Studie und WeCare-DiGiP, hat im Jahr 2023 begonnen. Im Dezember 2023 wurde bzgl. der Bekanntmachung zur „Förderung von interdisziplinären Projekten zum Thema Entwicklung und Erprobung von neuen Ansätzen der Datenanalyse und des Datenteilens in der Long-/Post-COVID-19-Forschung“ ein neuer Antrag mit engem Bezug zu Cell-o eingereicht. Das Projekt „Föderierte Netzwerkmodellierung ökologischer komplexer dynamischer Muster in Post-COVID“ ist im Nov. 2024 angelaufen. Die Zusammenarbeit mit anderen BMBF-Projekten (A2I & Eghi) in den Fördermaßnahmen „Adaptive Technologien für die Gesellschaft - Intelligentes Zusammenwirken von Mensch und KI“ ist noch nicht abgeschlossen, da drei Projekte die Datenerfassung noch nicht beendet haben. Gespräche und Ideen zur Kooperationen und Gemeinsamkeiten haben stattgefunden. Es fanden bereits jährliche A2I/Cello-Workshops und einen regelmäßiger Kontakt zwischen den Projektkoordinatoren statt, um gemeinsame Interessen in Bezug auf Stress und Smartwatches zu diskutieren und um wissenschaftliche Outreach-Aktivitäten und gemeinsame Veröffentlichungen zu planen und voranzutreiben. Die Kontaktaufnahme zum Eghi-Projekt („Erweiterte Gesundheitsintelligenz für persönliche Verhaltensstrategien im Alltag“) war ebenfalls in Planung, um die Zusammenarbeit in den Bereichen Gesundheitsassistenzsysteme, KI und Gesundheitskompetenz zu etablieren. Wie eingangs beschrieben sind mehrere weitere Projekte angelaufen, welche die Erkenntnisse und Strukturen aus dem Cell-o-Projekt nutzen oder aufgreifen: Bsp. G-BA: WATCH; BMBF: FEDORA - „Föderierte Netzwerkmodellierung ökologischer komplexer dynamischer Muster in Post-COVID“, BMG REMIT – „Remit Remote Monitoring und -intervention für eine optimierte Versorgung des Post- und Long-COVID-Syndrom“; Industrie: NCT06278077: Neurexan - a Clinical Trial in Short-Term Insomnia Patients; weitere Anschlussprojekte, u.a. gemeinsam mit dem Deutschen Zentrum für Psychische Gesundheit (DZPG) am Standort

Halle-Jena-Magdeburg sind angebahnt/in Planung, z.B. Projekt JE7 („Frühe Erfassung und Erfolgssteigerung von Psychotherapie anhand pharmakologischer und digitaler Augmentation“), Realisierung durch Amendment zu Trial ID: 2024-512478-86-00, DFG: GRK-NeurInfekt (Skizze positiv begutachtet, Evaluation März 2025, MiVaS - Forschungsgruppe Mikrovaskulatur unter Stress). Durch den hohen Anwendungsbezug von Fraunhofer IGD werden praxistaugliche Ergebnisse auch in die Wirtschaftsforschung überführt.

Folgende Meilensteine wurden ursprünglich festgelegt:

1. Meilenstein (Ende des 4. Monats): Integration der Sensoren und Fertigstellung des vorläufigen App-Designs (Universitätsklinik Jena, Thryve);

2. Meilenstein (Ende des 12. Monats): Fertigstellung des operativen Systems zur internen Bewertung der App (Universitätsklinik Jena, Thryve, Fraunhofer IGD)

Diese beiden Meilensteine im ersten Jahr wurden in Zusammenarbeit zwischen UKJ und Thryve termingerecht erreicht. Für weitere Details s.u. (Arbeitspakete AP). Die Entwicklung der App zur Erhebung der EMA-Assessments und Einbindung der passiven Sensordaten konnte im geplanten Zeitbudget erfolgen (Thryve).

3. Meilenstein (Ende des 18. Monats): Interne Auswertung der Ergebnisse und anschließende Anpassung/Verbesserung an die Studie (Universitätsklinik Jena, Saarland Universität)

4. Meilenstein (Ende des 24. Monats): Überarbeitung des Systems für die Studie (Universitätsklinik Jena, Saarland Universität)

5. Meilenstein (Ende des 33. Monats): Abschluss der Forschungsstudie, Fertigstellung von drei Dissertationen, Veröffentlichung von wissenschaftlichen Artikeln in Bearbeitung (Universitätsklinik Jena).

Die Meilensteine 3 bis 5 sind, wie bereits in den vergangenen Zwischenberichten erläutert, im Verlauf des Projektes, verspätet bzw. verändert erfüllt worden, hierzu wird auf die folgende detaillierte Darstellung anhand der einzelnen Arbeitspakete in den Modulen „1. Konzeption und Koordination“, „2. Entwicklung und Validierung der Software“, „3. Stresscharakterisierung“ und „4. Kalibrierung, Feldtest, und Optimierung“ verwiesen. Abb. 5 zeigt im Überblick die Anpassungen, die im Projektverlauf insbesondere bzgl. der Studienplanung und -durchführung erforderlich waren, weitere Erläuterungen dazu finden sich in den jeweiligen Arbeitspaketen.

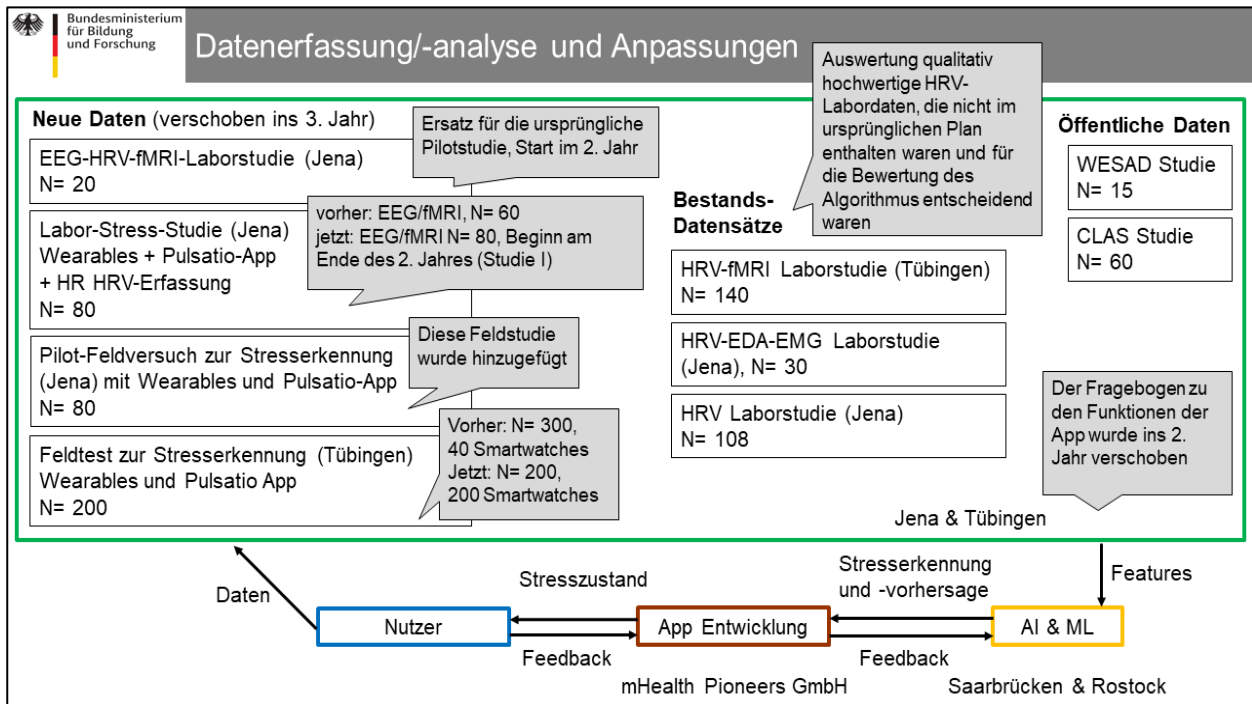


Abb. 5: Überblick über die Datenerfassung und entsprechende Anpassungen (UKJ)

Es folgt die Darstellung des Projektverlaufs im Vergleich zu den **(ursprünglichen) Arbeitspaketen:**

Modul 1: Konzeption und Koordination

AP 1.1 (Monat 1-36; verantwortlich: UKJ)

(ursprüngliches) Ziel: Projektkoordination, Einhaltung des Zeitplans, Koordination von Lösungsansätzen bei unerwarteten Problemen

UKJ: Im Rahmen der Projektkoordination stand das UKJ mit allen Forschungsgruppen und den einzelnen Projektpartnern in regelmäßigem Informationsaustausch, sorgte stetig für die Einhaltung des Zeitplans und der gesetzten Meilensteine (Ist-Soll-Analyse) und unterstützte durch regelmäßige Treffen bei der Suche nach Lösungen bei unerwarteten Problemen. Insbesondere mussten Anpassungen in den finanziellen und organisatorischen Projektplänen vorgenommen und bei Schwierigkeiten vor allem aufgrund der Corona-Beschränkungen, langen Krankheitszeiten und anderen pandemie-bedingten Verzögerungen entsprechende Lösungen gefunden werden, um den erfolgreichen Fortgang des Projektes zu gewährleisten. Diese sind im folgenden Abschnitt beschrieben. Der regelmäßige Informationsaustausch (Projektkoordination UKJ) mit allen Forschungsgruppen und den einzelnen Projektpartnern wurde auch im zweiten Förderjahr beibehalten, um den Zeitplan und die gesetzten Meilensteine in einer Ist-Soll-Analyse regelmäßig zu überprüfen und bei Störungen anzupassen und um Lösungen für unerwartete Problemen zu finden. Dazu wurden weiterhin regelmäßige, z.T. kleinere Treffen organisiert und durchgeführt. Die ursprünglichen Anpassungen wegen Schwierigkeiten aufgrund der Corona-Beschränkungen, der langen Krankheitszeiten und anderen pandemie-bedingten Verzögerungen blieben erhalten und wurden aufgrund weiterer großer Aufgabenpakete (Datenschutz, IT, Kooperationsvertrag, Datenverarbeitungsverträge mit Withings etc.) ergänzt.

Zu Beginn war eine gründliche Suche nach geeigneten Smartwatches und strenge Testungen der Geräte notwendig. Das ursprüngliche Vorgehen laut Antrag bestand darin, kommerzielle Smartwatches und Fitness-Tracker von mehreren Anbietern zu beschaffen, die mit modernsten Sensoren die wichtigen stressbezogenen Vitalparameter (EKG, PPG, EDA, Körpertemperatur,

Sauerstoffsättigung, Bewegung usw.) kontinuierlich messen und über die öffentliche API der Anbieter Zugang zu deren Daten zu erhalten. Die zu Forschungszwecken eingesetzten Empatica E4 Wearables messen neben der Bewegung (ACC) kontinuierlich physiologische Signale (BVP, EDA, TEMP). Sie sind (bzw. waren zum Zeitpunkt des Projektbeginns) jedoch teuer, ließen sich nicht ohne großen Entwicklungsaufwand mit der Pulsatio-App verbinden und hatten keine Perspektive zur Weiterentwicklung bis zur Marktreife als flächendeckend in der Versorgung nutzbares Medizinprodukt. Thryve empfahl die marktführenden Anbieter Apple, FitBit, Garmin, Withings und Polar. Für die Studien des Projektes wurde eine kommerziell verfügbare Smartwatch für den Alltag nötig, die günstig ist, eine geeignete Akkulaufzeit aufweist, die Datenschutzerfordernisse der Ethikkommission erfüllte und über moderne Sensoren zu Erfassung mehrerer Vitalparameter verfügte. Es war außerdem erforderlich, dass der jeweilige Anbieter den Zugriff auf die meisten bzw. bestenfalls alle dieser relevanten Signale über seine öffentliche API ermöglicht. Darüber hinaus setzte sich bei den Anbietern zunehmend die Erkenntnis durch, dass die Verwendung von EKG eine bessere Einschätzung der HRV ermöglicht als die Verwendung von PPG, da PPG sehr empfindlich auf Geräusche und Bewegungen reagiert. Polar musste daher ausgeschlossen werden, da dieser Anbieter nur Wearables für den Sport und nicht für den Alltag entwickelt hat. FitBit bot aufgrund der zahlreichen integrierten Sensoren und Vitalparameter attraktive Smartwatches. Leider teilte der Anbieter seit einiger Zeit die relevanten Daten nicht mehr mit Wissenschaftler:innen, da hier eine eigene App entwickeln werden soll. Apple hat moderne Sensoren und galt bereits zum damaligen Zeitpunkt in der Literatur als beliebter Anbieter auch für Forschungszwecke. Leider überstiegen die Kosten entsprechender Smartwatches die Kosten für die geplanten Studien. Garmin blockierte den Zugriff auf die wichtigsten Signale über die öffentliche API, bot kein EKG an und stellte Wissenschaftler:innen nur eine geringe und unflexible zeitliche Auflösung für die Zeitreihen zur Verfügung. Die Withings Scanwatch misst EKG, weist eine lange Akkulaufzeit auf, erfüllt die Datenschutzerfordernisse der Ethikkommission und hat (hatte) einen günstigen Preis im Rahmen der Projektmöglichkeiten und wurde daher für die Labor- und Feldstudien ausgewählt. Während der gesamten Zeit fanden regelmäßig Treffen mit mehreren Vertreter:innen von Smartwatch-Anbietern statt. Wir hatten Hilfe und technische Unterstützung von unseren Projektpartner:innen im IGD Rostock. Der Projektpartner Thryve stand ebenfalls mit Feedback und Kontakt zu den Anbietern zur Verfügung.

AP 1.2 (Monat 1-3; verantwortlich: UKJ)

(ursprüngliches) Ziel: „Use Case“-Analyse: Frühzeitige Validierung, damit die Merkmale des zu entwickelnden Produkts den Bedarf der Anwender im Kontext des Stressmanagements abdecken. „Product-Market Fit“ sowie Antrag auf Genehmigung durch die lokale Ethik-Kommission (Ethikvotum)

UKJ: Die im Antrag beschriebene „Use Case“ Analyse konnte aufgrund des zum betreffenden Zeitpunkt herrschenden hohen Infektionsaufkommens und der für das Projekt markanten Personalausfälle nicht durchgeführt werden. Stattdessen erfolgte eine ausführliche Diskussion des Fragebogendesigns unter dem Gesichtspunkt der Benutzerfreundlichkeit mit Fokusgruppen und externen Expert:innen mit langjähriger Erfahrung auf den Gebieten der EMA-basierten Datenerhebung und der Stresscharakterisierung. Die Ergebnisse wurden in die zu entwickelnde App integriert.

Des Weiteren stand eine Überprüfung des „Product-Market Fit“ zusammen mit Thryve sowie der Antrag auf Genehmigung durch die lokale Ethik-Kommission (Ethikvotum) auf der Agenda. Die Ethikanträge für die entsprechenden Studien wurden erarbeitet und geprüft.

Es erfolgte eine Anbahnung und Sicherstellung der Infrastruktur und des Zugangs zu späteren Nutzer:innen. Die Überprüfung von Frühversionen („proof-of-concept“) und die Analyse potentieller Hardware-Risiken konnte ebenfalls fristgerecht vorgenommen und abgeschlossen werden. Sowohl für die Laborexperimente als auch für die Nutzung der Smartwatch und der Pulsatio-App blieb das ausdrückliche Ziel des UKJs erhalten, die Datenschutz- und Sicherheitsanforderungen der Ethikkommission und der aktuellen

Rechtsnormen der EU-DSGVO zu erfüllen.

AP 1.3 (Monat 1-36; verantwortlich: UKJ)

(ursprüngliches) Ziel: ELSI Organisation

UKJ: Sowohl für die Laborexperimente als auch für die Smartwatch und Pulsatio-App war das ausdrückliche Ziel, die Datenschutz- und Sicherheitsanforderungen der Ethikkommission und der aktuellen Rechtsnormen der EU-DSGVO zu erfüllen. Thryve und das UKJ haben zusammen daran gearbeitet, Datenschutzerklärungen, Informationen für Proband:innen und Aufklärungsbögen zu erstellen. Die Kolleg:innen von Thryve haben eine Datenschutzfolgeabschätzung und eine Datenschutz-FAQ zur Pulsatio-App verfügbar gemacht und Withings hat ebenfalls einen Kooperationsvertrag bezüglich der Datenspeicherung und -verarbeitung mit dem UKJ unterzeichnet. Für jede Studie sind die Ein- und Ausschlusskriterien klar definiert worden. Die Freiwilligkeit der Teilnahme in den Studien wurde explizit ausgewiesen und die Proband:innen konnten ihre Teilnahme jederzeit ohne Angabe von Gründen beenden. Ihre Daten wurden in diesem Fall gelöscht.

AP 1.4 (Monat 1-3; verantwortlich: Thryve, Kollaboration: UKJ)

(ursprüngliches) Ziel: Erstellung einer Systemkonzeption zur Daten-(Signal-) Aufzeichnung der Wearables sowie einer grundlegenden Benutzer-App mit Back-End Service.

Thryve: Folgende Arbeiten bei Thryve wurden erfolgreich abgeschlossen: Erstellung einer Systemkonzeption zur Datenaufzeichnung der Wearables sowie die Entwicklung einer grundlegenden Benutzer-App (Pulsatio-App) mit Back-End-Service, deren Funktionalität getestet und deren Entwicklungsprozess (nach DIN EN ISO / IEC 27001: 2017 und ISO 9001 : 2015) zertifiziert wurde; Ziel: Vorläufige Tests im Labor zur Beurteilung des Designs des ersten Systems zur Verbesserung der Anwendungserfahrungen sowie zur Generierung erster Daten. Durch Verzögerungen bei den Testaufnahmen konnten die vorläufigen Labortests zur Beurteilung des Designs des ersten Systems zur Verbesserung der Anwendungserfahrungen sowie zur Generierung erster Daten nicht abgeschlossen werden.

Die Machbarkeit der geplanten Biofeedback-Komponenten in Echtzeit hat sich als weitestgehend unvereinbar mit der für die wirtschaftlichen Erfolgsaussichten geeigneten Geräteauswahl erwiesen (es gab keine Hardware für Anwendungen in Echtzeit, s.o.) und konnte technisch nicht umgesetzt werden. Stattdessen wurde eine Biofeedback-Komponenten mit einer Latenz < 3min umgesetzt. Für Biofeedback geeignete Gerätetypen benötigen qualitativ hochwertige (reliabel, valide) Messungen der HRV. Dies lässt sich zum einen über Elektrokardiogramm-Technologien wie in Brustgurten oder in sehr stark kontrolliert erfolgten Messsituationen über optische PPG-Sensoren erreichen.

Die für optische HRV-Messungen geeigneten Geräte wie Fitbit Sense oder Apple Watch hatten den Zugriff auf die Daten und die Steuerung der Messungen leider für externe Anwendungen gesperrt bzw. gaben keinen Datenzugriff auf Messungen, die nicht innerhalb eigener Applikationen mit Messprotokoll und Nutzerführung erstellt wurden. Das Ziel der Ausrichtung auf in der Bevölkerung bereits verbreitete Sensorik war ein klares Ergebnis der Geräteauswahl unter dem Gesichtspunkt der Erfolgsaussichten für eine wirtschaftliche Verwertung. Da diese primär an den Erstattungen des SGB V §20 ausgerichtet waren, war die Erstattung von Geräten nicht realisierbar, sodass auf bereits verbreitete Sensortypen abgestellt werden sollte.

Die Messungen über EKG waren mit den für den Feldtest geplanten Geräten Apple Watch (4,5,6) und Withings ScanWatch implementiert. Da die Auslösung der Messung und Synchronisierung der Daten durch die technischen Vorgaben der Hersteller leider nicht verlässlich in Echtzeit erfolgen konnten, war eine Echtzeit-Biofeedback-Funktion auch hier schwer machbar. Im Feldtest in zukünftigen Studien sollten daher im weiteren Verlauf die Latenzen ermittelt und dafür geeignete Feedbackverfahren getestet werden. Die Ermittlung der Latenz und Testung der Latenz-geeigneten Feedbackverfahren wurden von der PLG und dem

Fraunhofer IGD erbracht.

Durch die oben beschriebenen Schwierigkeiten bei der Datenfreigabe von optischen Pulssignalen ergab sich zusätzlicher Aufwand für die Einbindung der EKG-Signale von Withings und Apple.

UKJ: Für die Erstellung einer Systemkonzeption zur Daten-(Signal-) Aufzeichnung der Wearables sowie einer grundlegenden Benutzer-App mit Back-End Service stand das UKJ in engem Austausch mit Thryve. Alle datenschutzrechtlichen Dokumente wurden finalisiert.

AP 1.5 (Monat 4-6; verantwortlich: UKJ, Kollaboration: Thryve)

(ursprüngliches) Ziel: *Vorläufige Tests im Labor zur Beurteilung des Designs des ersten Systems zur Verbesserung der Anwendungserfahrungen sowie zur Generierung erster Daten*

UKJ: Die am UKJ geplanten vorläufigen Labortests zur Verbesserung der Anwendungserfahrungen sowie zur Generierung erster Daten zur Beurteilung des Designs des ersten Systems fanden wegen pandemie-bedingten Verzögerungen nicht wie geplant statt. Laut Originalplan sollten an 30 Proband:innen über die Dauer von jeweils einem Monat Testungen durchgeführt und ein erster vorläufiger Datensatz für weitere Testungen generiert werden. Wegen der Corona-Beschränkungen und weil das Testen und Bewerten von Smartwatches länger als erwartet dauerte, wurden in einer kleineren Pilotstudie mit drei Smartwatches unterschiedlicher Marken drei Datensätze über drei Monate analysiert (s.u.). Dabei wurden die Smartwatches FitBit Sense, Withings Scanwatch und Garmin Fenix 5x plus genutzt. Mittlerweile stellt FitBit die wichtigsten Zeitreihen nicht mehr für Wissenschaftler:innen zur Verfügung.

Die Hauptlaborstudie mit 80 Proband:innen (Laborstudie I) in Jena wurde vorbereitet.

Thryve/UKJ: Zwei Datensätze (Withings, Garmin) wurden vom UKJ und von Thryve getestet und evaluiert und es wurde festgestellt, dass die zeitliche Auflösung und die Datenqualität für die Signalverarbeitung und das maschinelle Lernen geeignet waren. Darüber hinaus ermöglichten die Art der Sensoren und die vielfältigen Modalitäten und Signale die Extraktion relevanter Merkmale für die Stresscharakterisierung. Der Datenzugriff und die Datenformate wurden mit Thryve und IGD Rostock geprüft und datenschutzkonform umgesetzt. Die Umsetzung des datenschutzkonformen Datenzugriffs war aufgrund der Pandemie-bedingten Home-Office-Arbeitssituation mit einem erhöhten Aufwand verbunden. Es musste sichergestellt werden, dass der Zugriff aus dem Home-Office kein Sicherheitsrisiko darstellt. Dies wurde durch die Implementierung und Nutzung eines IP-Whitelisting sichergestellt. Die Ergebnisse der ersten Datenerhebung und Evaluierung waren zufriedenstellend. Aus anderen Gründen, die wir oben erläutert haben, wurde schließlich entschieden, die Withings Scanwatch für die anstehenden Labor- und Feldstudien zu verwenden. Außerdem erfolgte in Rostock und Saarbrücken eine Analyse von Daten aus alten und aktuellen Projekten (aus Jena und Tübingen).

Modul 2: Entwicklung und Validierung der Software

AP 2.1 (Monat 5-9; verantwortlich: Fraunhofer IGD, PLG, Kollaboration: Thryve, UKJ)

(ursprüngliches) Ziel: *Bestimmung aussagekräftiger Vitalkennzeichen aus sensorischen Rohdaten (Voraussetzung: Systemdesign der App (AP 1.4))*

Fraunhofer IGD: Beim Fraunhofer IGD wurden Kommunikationsstrategien erfolgreich etabliert, u.a. in Fokusgruppen und auch Konsortialmeetings und in enger Abstimmung mit der PLG zur Entwicklung der Belastungserkennung und Anomalie-Erkennung. Ein Schwerpunkt beim Fraunhofer IGD war außerdem die Fertigstellung der Komponenten zur Datenfusion, Anomalie-Erkennung und Belastungserkennung. Gemeinsam mit den Verbundpartnern wurde eine Projektumsetzung wie in Abb. 12 (s.u.) dargestellt verfolgt. Durch Verzögerungen bei den Testaufnahmen wurde die Bestimmung aussagekräftiger Vitalkennzeichen aus

sensorischen Rohdaten durch das Fraunhofer IGD, die Universität des Saarlandes und Thryve verzögert abgeschlossen. Die Einrichtung der Datenpipeline durch Thryve ist erfolgt. Es wurden Optionen zu den Vitalkennzeichen von Thryve mit dem Fraunhofer IGD besprochen. Auch im Zeitplan der Fraunhofer IGD traten im Berichtszeitraum unerwartete Verzögerungen auf. Diese waren in großen Teilen auf die nach wie vor pandemie-bedingte schwierige Arbeitssituation mit fehlenden Vor-Ort-Terminen und ausschließlich virtuellem Austausch zurückzuführen. Dadurch war es vorerst schwierig, eine Pipeline zur automatisierten Datenfusion und Beanspruchungserkennung zu entwickeln und zu erproben. Diese Schritte wurden später nachgeholt.

Beim Fraunhofer IGD wurden Kommunikationsstrategien erfolgreich etabliert, u.a. in Fokusgruppen und auch Konsortialmeetings und in enger Abstimmung mit der PLG zur Entwicklung der Belastungserkennung und Anomalie-Erkennung. Ein Schwerpunkt beim Fraunhofer IGD war außerdem die Fertigstellung der Komponenten zur Datenfusion, Anomalie-Erkennung und Belastungserkennung. Gemeinsam mit den Verbundpartnern wurde folgende Projektumsetzung verfolgt (Abb. 6):

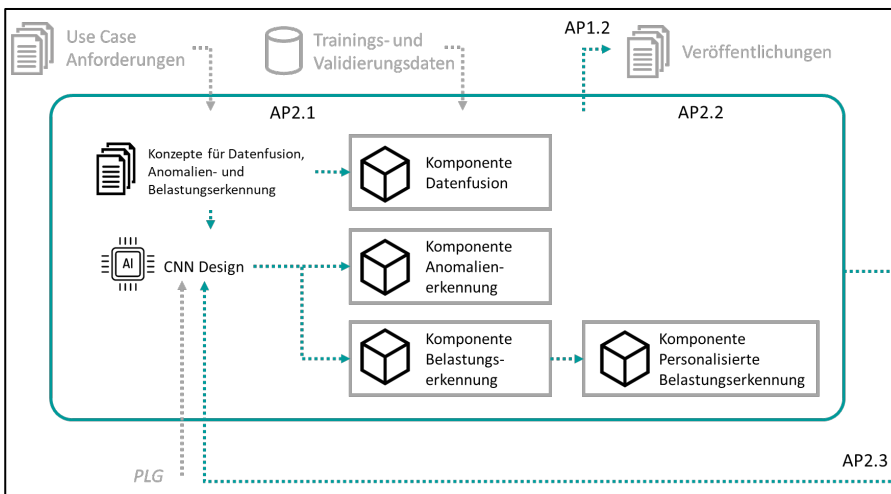


Abb. 6: Darstellung der geplanten Projektumsetzung durch das Fraunhofer IGD (Fraunhofer IGD)

Alle Projektarbeiten wurden mit dem Ziel durchgeführt, die geplanten Arbeitsaufgaben erfolgreich zu bearbeiten. Dazu zählen folgende wissenschaftliche und technische Arbeitsziele des Gesamtvorhabens:

- Erforschung und Entwicklung eines Verfahrens zur Fusion multivariater Vitalparameter aus unterschiedlichen Mobilgeräten und Plattformen insbesondere bei unvollständigen und qualitativ eingeschränkten Datenbeständen
- Erforschung und Entwicklung von adaptiven CNN-basierten Analyseverfahren zur personalisierten und automatisierten Anomalie-Erkennung, sowie Belastungserkennung in multivariaten Vitalparametern

Sowohl die Validierungsarbeiten der generellen Methodik mit dem WESAD-Datensatz (Schmidt, Reiss, Duerichen, Marberger, Van Laerhoven 2018, <https://doi.org/10.1145/3242969.3242985>) als auch die Validierung unterschiedlicher Smartwatch-Plattformen haben zur Zielerreichung unmittelbar beigetragen und damit zur Auswahl einer geeigneten Smartwatch (Withings) und der Umsetzung einer prototypischen ML-Pipeline geführt. Folgende Teilergebnisse wurden erreicht: Es wurde eine Pipeline zum Training der ML-Modelle implementiert und basierend auf der vom Projektpartner Thryve vorhandenen Datenstruktur prototypisch erstellt. Die finale Pipeline konnte zum damaligen Zeitpunkt aufgrund von Verzögerungen an der Datenschnittstelle noch nicht umgesetzt werden (s.u.). Die Belastungs- bzw. Anomalie-Erkennung wurde anhand der WESAD-Daten prototypisch umgesetzt. Erste Ergebnisse deuteten eine personalisierte Belastungserkennung mit 92% Genauigkeit an. Abb. 7 zeigt die Ergebnisse einer Leave-One-Subject-Out Kreuzvalidierung des ML-Modells. Das Modell zeigte gute Ergebnisse bei der Detektion von Stress. Zur Belastungserkennung wurden hierfür aufgezeichnete Daten einer Empatica E4 verwendet. Insbesondere die Verwendung des Photoplethysmography (PPG) und Beschleunigungssensor (ACC) Signals ermöglichte eine gute Erkennung von

Stressphasen. Die Limitationen dieses Ansatzes mussten jedoch weiterhin untersucht werden. Unter starker Bewegung leiden PPG Signale unter starkem Rauschen und anderen Artefakten. Algorithmen des ML würden dann hierbei das Rauschen lernen, d.h. Phasen, in denen Bewegung auftritt, sind dann automatisch als Stress klassifiziert. Um dem entgegenzuwirken, brauchte es umfänglichere Datenbanken, die auch Stressphasen ohne Bewegungs-, bspw. arithmetische Tests am Computer oder Ähnliches umfassen.

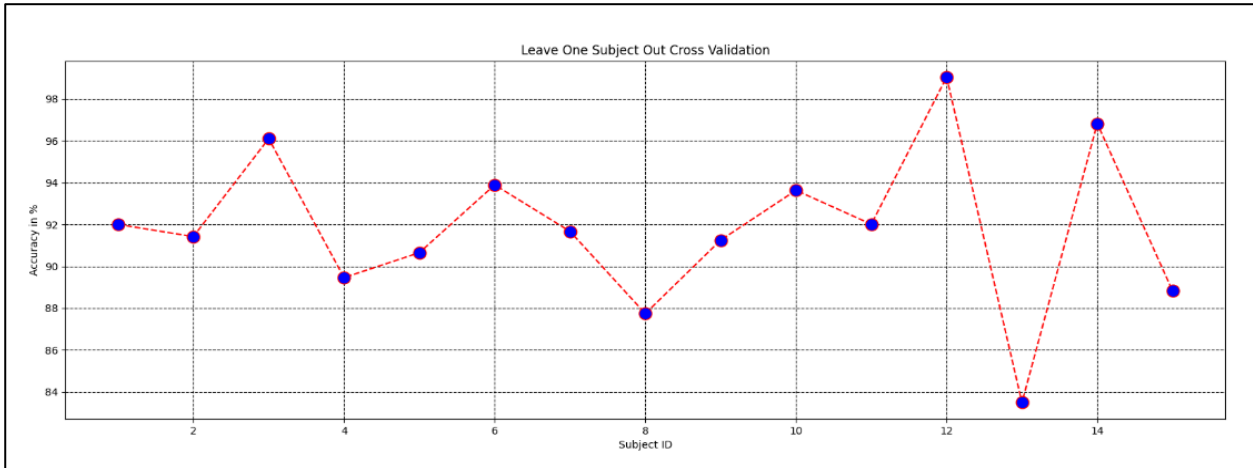


Abb. 7: Ergebnisse einer LOSO Kreuzvalidierung. Die Ergebnisse zeigen, dass Stress im Durchschnitt mit einer Genauigkeit von 92% erkannt werden kann (Fraunhofer IGD)

Die Herausforderungen einer Stresserkennung liegen in der Multimodalität und in der Subjektivität von Stress. Zur Verwendung von Modellen des ML war es wichtig, multimodale Daten in eine homogene Datenstruktur zusammenzuführen. Eine homogene Datenstruktur der Daten der Wearables, EKG-Daten und der EMAs wurde durch Thryve bereitgestellt. Dies stellte den nächsten Schritt bei der Entwicklung der Modelle dar. Ein Beispiel für ein Stresssignal kann den Abb. 8 und Abb. 9 entnommen werden. Hier zeigte sich, dass insbesondere die EDA ein guter akuter Stressmarker ist. Leider ist es derzeit unmöglich, mit verfügbaren kommerziellen Smartwatches die EDA zu messen. Die Verwendung physiologischer Signale mit anderen Modalitäten wie beispielsweise GPS-Standort, Aktivität und Schlaf sind der Schlüssel zur korrekten Klassifizierung von Stressphasen.

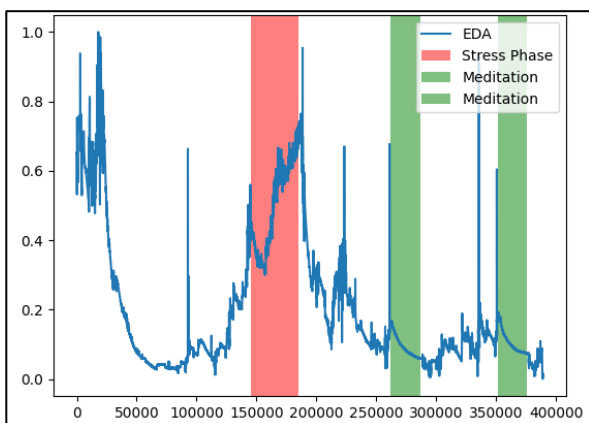


Abb. 8 - Auch zu Beginn des Experiments zeigte sich eine hohe EDA. Die horizontale Achse bildet hier die Zeit ab, die vertikale die Ausprägung der EDA, skaliert (Fraunhofer IGD)

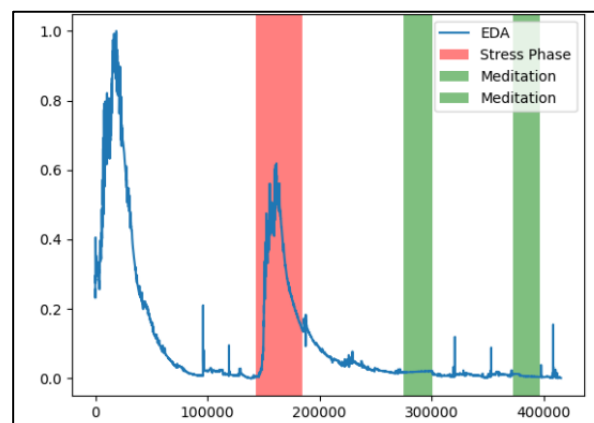


Abb. 9 - Stress äußert sich insbesondere in der Elektrodermalen Aktivität (EDA); (Fraunhofer IGD)

PLG: Die PLG trug maßgeblich zur Entwicklung und zum Prototyping von Algorithmen zur Stress- und Anomalie-Erkennung bei. Darüber hinaus hat die PLG während des gesamten Prozesses wichtige Herausforderungen im Zusammenhang mit der multimodalen Datenintegration und dem Rauschen in den Sensordaten von Wearables identifiziert und angegangen. Bewegungsbedingte Artefakte,

insbesondere in PPG-Signalen, wurden als eine der wichtigsten Einschränkungen identifiziert, was auf die Notwendigkeit erweiterter Sensordatenerfassungsmethoden und verbesserter Artefaktmanagement-Techniken in zukünftigen Datenbankiterationen hinweist (Abbildungen 10 bis 15). In Zusammenarbeit mit dem Fraunhofer IGD wurde eine umfassende vorläufige Analyse des WESAD-Datensatzes durchgeführt. Die aus dieser Analyse gewonnenen Erkenntnisse wurden in Präsentationen und anschließenden Diskussionen mit den Projektpartnern ausgetauscht, wobei zwei kritische Beobachtungen im Vordergrund standen: Erstens zeigen die Personen im WESAD-Datensatz sehr unterschiedliche Stressreaktionen, was die Notwendigkeit personalisierter Modellierungsansätze unterstreicht (Abbildung 10).

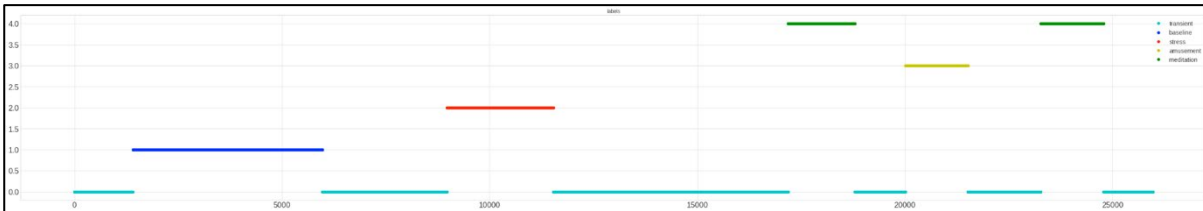


Abb. 10: WESAD-Datenvisualisierung der verschiedenen Episoden (PLG)

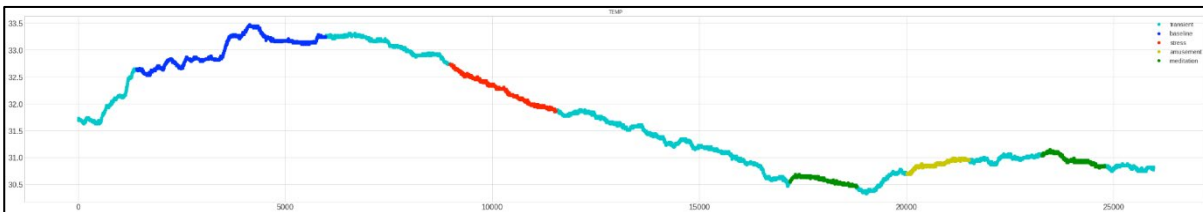


Abb. 11: WESAD-Datenvisualisierung des Temperaturkanals (PLG)

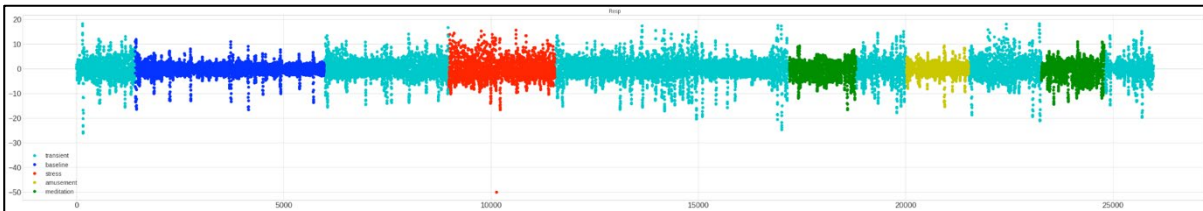


Abb. 12: WESAD-Datenvisualisierung des Respirationskanals (PLG)

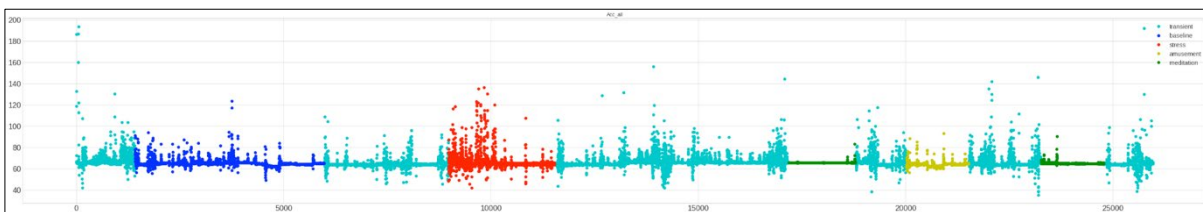


Abb. 13: WESAD-Datenvisualisierung des Beschleunigungskanals (PLG)

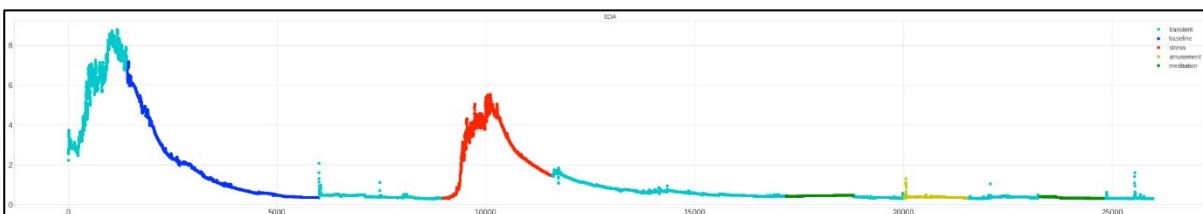


Abb. 14: Datenvisualisierung des EDA-Kanals (PLG)

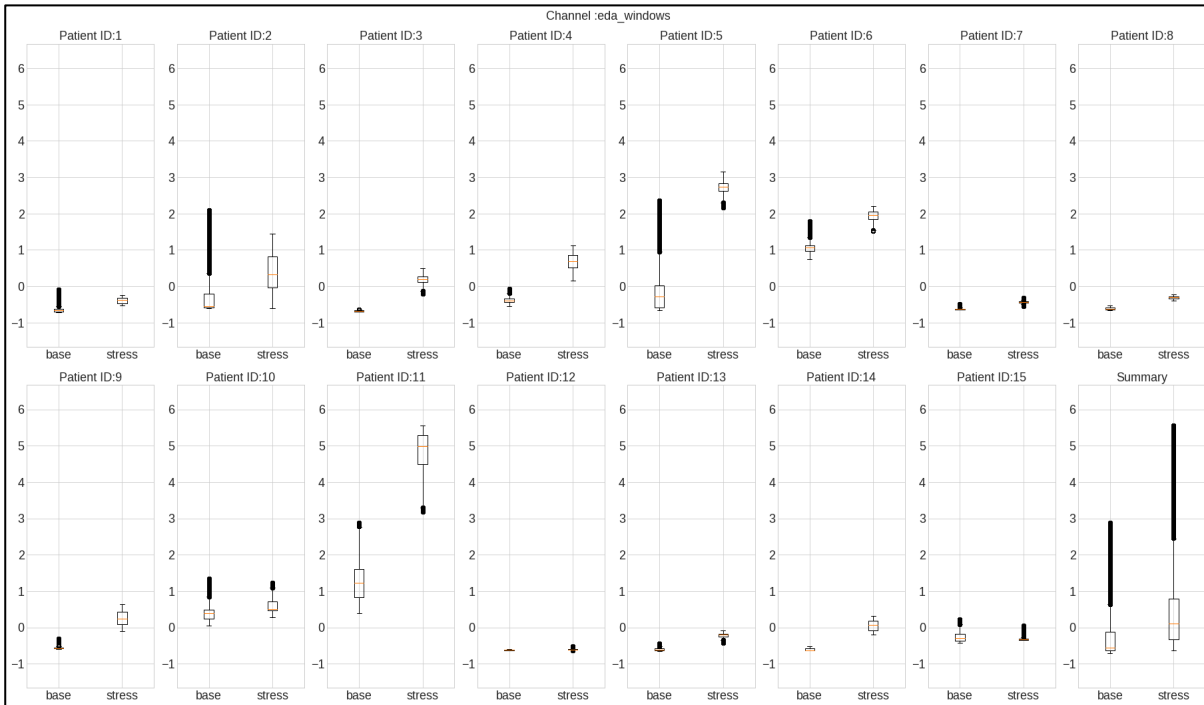


Abb. 15: Verteilung der EDA-Merkmale für jeden Patienten in der Grundlinie und im Stresszustand (PLG)

Zweitens erschien die Markierung der Grundlinie im Datensatz unzuverlässig, da die Probanden zu Beginn der experimentellen Sitzungen erste Stressreaktionen zeigten (Abbildung 11 und 15). Diese Ergebnisse flossen in Empfehlungen an das Konsortium ein, die darauf abzielten, solche Artefakte während der nachfolgenden Datenerhebungsphase für das Cell-o-Projekt abzuschwächen.

Unter Nutzung des WESAD-Datensatzes sowie zusätzlicher Datensätze, die von Thryve zur Verfügung gestellt wurden, entwickelte und implementierte die PLG eine prototypische Pipeline für maschinelles Lernen, die speziell auf die Analyse physiologischer Signale, insbesondere Photoplethysmographie- (PPG) und Beschleunigungsmesserdaten (ACC), zugeschnitten ist. Die Gruppe erreichte eine personalisierte Stresserkennungsgenauigkeit von ca. 92 % in Leave-One-Subject-Out-Validierungsszenarien, was die erheblichen Vorteile der personalisierten Modellierung unterstreicht. Im Folgenden (Tab. 1-5) werden die Leistungskennzahlen für den WESAD-Datensatz unter Verwendung von Beschleunigungs- und PPG-Kanälen mit Deep-Architekturen vorgestellt, die auf der Verwendung von Multilayer Perceptrons (MLPs), Convolutional NNs (CNNs) und Transformers basieren.

Tab. 1: Leistung bei der Rekonstruktion für Dense AE (PLG)

MSE Loss		LR	
		0.00010	0.00001
Layers	(1024, 2048)	0.092735 \mp 0.067625	0.161432 \mp 0.104841
	(2048, 1024)	0.069407 \mp 0.048114	0.104300 \mp 0.066275

Tab. 2: Leistung bei der Klassifizierung mit Merkmalen von Dense AE (PLG)

Accuracy		LR	
		0.0001	0.0003
Layers	(128, 1)	0.763133 ± 0.071842	0.749124 ± 0.066504
	(128, 32, 1)	0.753594 ± 0.064182	0.725327 ± 0.076476
	(256, 32, 4, 1)	0.747416 ± 0.070298	0.738607 ± 0.065656

Tab. 3: Leistung bei der Rekonstruktion für 1-D CNN AE (PLG)

MSE Loss		LR		
		0.00010	0.00030	0.00001
Layers	(16, 32)	0.011757 ± 0.007364	0.004900 ± 0.003076	0.138050 ± 0.091092
	(32, 64)	0.009382 ± 0.005995	0.002994 ± 0.001618	0.127733 ± 0.088156
	(64, 128)	0.008443 ± 0.005294	0.002922 ± 0.001843	0.129111 ± 0.089010

Tab. 4: Leistung bei der Klassifizierung mit Merkmalen von 1-D CNN AE (PLG)

Accuracy		LR		
		0.0001	0.0003	0.00001
Layers	(128, 1)	0.758439 ± 0.057623	0.753857 ± 0.052356	0.768883 ± 0.051918
	(128, 32, 1)	0.748576 ± 0.051025	0.741515 ± 0.045679	0.766871 ± 0.052274
	(256, 32, 4, 1)	0.748919 ± 0.054678	0.750374 ± 0.053468	0.767708 ± 0.052971
	(512, 256, 32, 4, 1)	0.753857 ± 0.056126	0.756039 ± 0.055794	0.725475 ± 0.136627

Tab. 5: Leistung bei der Klassifizierung mit überwachtem Transformator (PLG)

Accuracy		LR		
		0.001	0.0003	0.0001
Embedding Dimension	32	0.858899 ± 0.091376	0.868967 ± 0.055755	0.854194 ± 0.068290
	64	0.867304 ± 0.055612	0.871690 ± 0.046510	0.859510 ± 0.081629
	128	0.849472 ± 0.063206	0.871910 ± 0.034956	0.858449 ± 0.062995

Um das Verständnis der für die Stressvorhersage wichtigen Signale zu vertiefen, führte die PLG eine umfassende Literaturrecherche durch und untersuchte die wichtigsten Forschungsarbeiten, um die wesentlichen physiologischen Signale, die bei der Stressvorhersage verwendet werden, die Leistung verschiedener Algorithmen des maschinellen Lernens und die Unterschiede zwischen Labor- und Alltagssituationen zu ermitteln. Insbesondere wurden die jüngsten Fortschritte bei der Modellierung von Zeitreihen untersucht, wobei die Verarbeitung asynchroner und multivariater Zeitreihendaten sowohl im diskreten als auch im kontinuierlichen Zeitbereich behandelt wurde. Der Schwerpunkt lag dabei auf Methoden, die mit fehlenden Daten umgehen können und durch aufmerksamkeitsbasierte Rahmenwerke (z.B. Transformer-Modelle) interpretierbare Erkenntnisse liefern. Zu den wichtigsten Beiträgen zählten Arbeiten von Shchur et al. (2021), Zuo et al. (2020), Zhang et al. (2020), Salehi et al. (2019), Boyd et al. (2020), Zerveas et al. (2021) und Grigsby et al. (2021). Darüber hinaus konzentrierte sich die PLG auf die jüngste Forschung im Bereich des unüberwachten Repräsentationslernens für multivariate Zeitreihendaten und hob wichtige Methoden von Yazdanbakhsh und Dick (2019), Franceschi et al. (2019), Yue et al. (2021) und Yèche et al. (2021) hervor. Die Gruppe untersuchte auch ausführlich die Literatur zu Methoden der Kausalinferenz und des Verstärkungslernens, die für personalisierte Stressinterventionen anwendbar sind, und diskutierte Rahmenwerke für die Inferenz individualisierter Behandlungseffekte (Alaa & van der Schaar, 2017; Bica et al., 2020; Zhang et al., 2020) und Anwendungen des Verstärkungslernens im Gesundheitswesen (Gottesman et al., 2019; Yu et al., 2021; Upadhyay et al., 2018; Li et al., 2018). Schließlich untersuchte die PLG den neuesten Stand der Forschung zur Stresserkennung sowohl in kontrollierten (Labor) als auch in unkontrollierten (in-the-wild) Umgebungen. Bemerkenswerte Arbeiten von Arsalan et al. (2022), Can et al. (2019), Siirtola (2019) und Gedam & Paul (2021) wurden geprüft, um die gängigen Datenvorverarbeitungstechniken, die mit tragbaren Sensordaten verbundenen Herausforderungen und die etablierten Grundlagen für eine effektive Stresserkennung besser zu verstehen. Im Anhang dieses Dokuments finden Sie eine vollständige Literaturübersicht, die von der PLG durchgeführt wurde.

AP 2.2 (Monat 6-12; verantwortlich: Fraunhofer IGD, Kollaboration: UKJ, Thryve)

(ursprüngliches) Ziel: Entwicklung personalisierter Vitalmarker für die Interpretation der Sensorsignale Voraussetzung: Zwischenergebnisse der „Use Case“-Analyse (AP 1.2)

Durch Verzögerungen bei den Testaufnahmen konnte beim Fraunhofer IGD die Entwicklung personalisierter Vitalmarker für die Interpretation der Sensorsignale und bei Thryve die Entwicklung personalisierter Vitalmarker für die Interpretation der Sensorsignale verzögert abgeschlossen werden.

PLG: In diesem Arbeitspaket analysierte die PLG einen freigegebenen Datensatz der Projektpartner aus Jena, der Cortisol-

Messungen enthält. In der Analyse arbeitete die PLG an der Korrelation zwischen den EMAs und den Cortisolwerten der Nutzer:innen sowie anderen Tests zur Stressbewertung wie dem State Trait Anxiety Inventory (STAI). Die PLG informierte Jena über die Statistiken und die Qualität der Messungen, die von verschiedenen Wearables im Labordatensatz erfasst wurden. In Bezug auf lebenswichtige Signale aus Rohdaten stützte sich die PLG im endgültigen vorgeschlagenen Modell auf ein Deep-Learning-Modell, um die Merkmale zu extrahieren, oder verwendete ein baumbasiertes ML-Modell, um Anomalien anhand von Herzfrequenzmessungen zu erkennen.

Die PLG leistete zusätzliche Arbeit und führte eine umfassende Analyse eines von den Projektpartnern am UKJ in Jena bereitgestellten Datensatzes durch, wobei sie sich speziell auf Cortisolmessungen als wichtige physiologische Indikatoren für Stress konzentrierte. Die Analyse umfasste die Untersuchung von Korrelationen zwischen Ecological Momentary Assessments (EMAs), Cortisolwerten und anderen psychologischen Stressindikatoren, insbesondere dem State-Trait Anxiety Inventory (STAI). Zu den Beiträgen der PLG gehörte ein detailliertes statistisches Feedback an die Kolleg:innen des UKJ hinsichtlich der Datenqualität und Zuverlässigkeit der Sensormessungen, die von verschiedenen tragbaren Geräten unter Laborbedingungen gewonnen wurden. Zur effektiven Ableitung personalisierter Vitalmarker aus sensorischen Rohdaten implementierte PLG fortschrittliche Methoden des maschinellen Lernens, wie z.B. Faltungsneuronale Netze (CNNs), CNNs für Punktwolken, um Lücken zu schließen, und Transformatoren (siehe Tabellen 1 bis 5). Deep-Learning-Modelle wurden für die anspruchsvolle Merkmalsextraktion eingesetzt, um die Interpretierbarkeit und Genauigkeit der von tragbaren Sensoren erfassten Vitaldaten zu verbessern. Darüber hinaus wurden baumbasierte maschinelle Lernmodelle eingesetzt, um Anomalien innerhalb der Herzfrequenzmessungen zu identifizieren, was wesentlich zur Robustheit und Präzision der im Rahmen des Projekts entwickelten Stresserkennungsverfahren beitrug. Diese Ansätze unterstützten gemeinsam das Ziel, personalisierte, zuverlässige Stressindikatoren aus Sensordaten zu erstellen.

Thryve unterstützte das Fraunhofer IGD bei der Entwicklung der personalisierten Vitalmarker.

UKJ: Das UKJ stand hier im Rahmen seiner möglichen und notwendigen Zuarbeiten und seiner Expertisen in enger Kooperation mit den anderen, hier hauptverantwortlichen Projektpartnern. Hinsichtlich der Kommunikation zwischen Thryve und den Psychiater:innen und Psycholog:innen in Jena und Tübingen wurden zu diesem Zweck Fokusgruppentreffen organisiert. Diese Treffen zum Thema Pulsatio-App und Datenerhebung halfen, viele offene Fragen zur Nutzung der App in Labor- und Feldstudien, zur Auswahl der Smartwatches, zum experimentellen Design und zu den Studienprotokollen, zur Implementierung der EMAs in der App und Bedenken über den Datenschutz und die Ethikanträge zu klären. Der Standort Jena unterstützte diesbezüglich bei der Kommunikation zwischen Thryve und den Machine-Learning-Experten in Rostock und Saarbrücken. Die Themen reichten vom Datenzugriff auf den Thryve-Server, auf welche Zeitreihen über die öffentliche API des Smartwatch-Anbieters zugegriffen werden konnte, die zeitliche Auflösung der Zeitreihen von Smartwatches bis hin zur Auswahl und dem Testen der Smartwatches. Darüber hinaus half das UKJ bei der Bereitstellung von Daten aus alten und aktuellen Projekten für die Machine-Learning-Experten in Rostock und Saarbrücken, um sie bei der Implementierung ihrer Algorithmen zu unterstützen, ihre Leistung zu testen und ihre Ergebnisse zu interpretieren. Darüber hinaus halfen die Anforderungen der Expert:innen für maschinelles Lernen dem UKJ, die Labor- und Feldstudien so zu gestalten, dass optimale Datensätze für die Stresserkennung und -vorhersage bereitgestellt werden. Außerdem wurden die offenen Fragen der Machine-Learning-Experten zu Stress, seinen Arten, seiner Diagnose und den relevanten Vitalsignalen und Merkmalen in der Stressforschung beantwortet.)

AP 2.3 (Monat 8-12; verantwortlich: Thryve, Kollaboration: UKJ)

(ursprüngliches) Ziel: *Integration der personalisierten Vitalmarker-Interpretationen in die App, Optimierung der Bedienung der App, skalierbares Back-End*

Thryve: Die Integration der personalisierten Vitalmarker-Interpretationen in die App wurde durch Thryve mit der Implementierung des Algorithmus zur automatischen Stresserkennung teilweise verspätet abgeschlossen (Implementierung des Algorithmus zur automatischen Stresserkennung (AP2.3 /AP3.4), da aufgrund der verzögerten Tests die Auswahl der personalisierten Vitalmarker-Interpretationen verspätet vorlag. Thryve hat das bereitgestellte Script überarbeitet, eine Umgebung in der Cloud eingerichtet, die das Ausführen des Scriptes anstern, durchführen und validieren kann. In Folge dessen wurde das Script zur Vitalmarker-Interpretation integriert und in Iterationsschleifen für verschiedene Szenarien der Datenverfügbarkeit robuster gestaltet. Durch die oben beschriebenen Schwierigkeiten bei der Integration der personalisierten Vitalmarker-Interpretationen in die App hat sich der Abschluss des AP 2.3 verzögert. Dadurch wurde eine Verlängerung des Projektes im Jahr 2024 notwendig. Durch die Verlängerung sollte möglich gemacht werden, den kompletten Datensatz zu analysieren, die personalisierten Vitalmarker in die App zu integrieren und dann die finale App in einer kleinen Feldstudie mit 60 Proband:innen zu validieren. Mit der Verlängerung wurde AP 2.3 schlussendlich abgeschlossen. Es erfolgte eine Optimierung der App nach ersten Funktionstests (QR-Code-Onboarding eingerichtet, 2D-Visual Scale als Frage-Items eingerichtet, EKG-Modul für EMA-Fragebögen integriert, vollständige Datenanbindung der Wearables, Datenexport-Server Bereitstellung damit alle Studienpartner Zugriff zu den Daten erhalten können).

UKJ: (vgl. AP 2.3)

AP 2.4 (Monat 12-36; verantwortlich: PLG, Kollaboration: UKJ)

(ursprüngliches) Ziel: Schätzen und Vorhersagen des Stressniveaus unter Verwendung eines unbeaufsichtigten und kontrollierten Algorithmus des maschinellen Lernens

PLG: PLG entwickelte und kodierte die gesamte Pipeline für maschinelles Lernen und übertrug sie an das Konsortium, das sich auf überwachte und halbüberwachte Algorithmen für die kontinuierliche Stressvorhersage spezialisiert hat (siehe Abb. 16). Der Algorithmus zur Stresserkennung wurde jedoch aus dieser Pipeline entfernt, da es an geeigneten gesammelten Daten mangelte (siehe Abbildung 17).

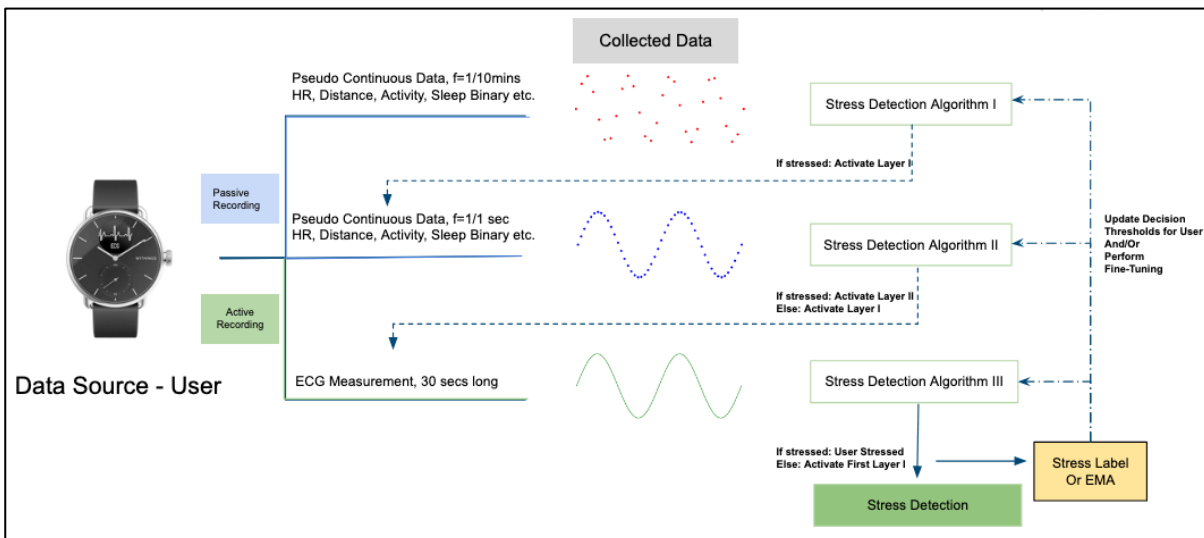


Abb. 16: Ursprüngliche, von der PLG vorgeschlagene Pipeline (PLG)

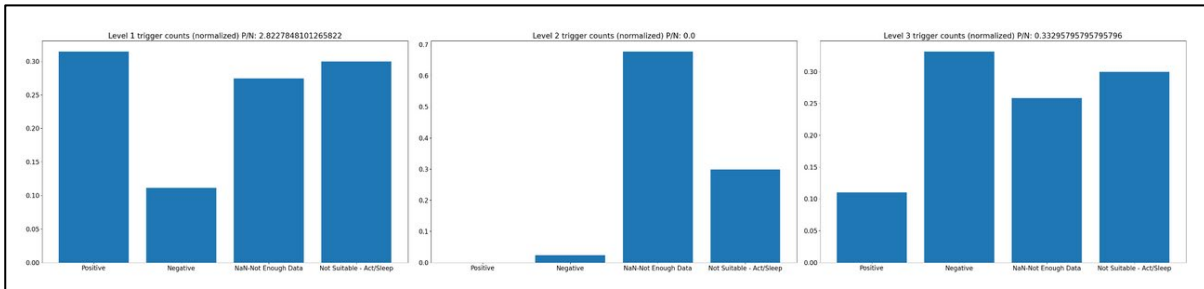


Abb. 17: Auslösungsverhältnisse pro Stufe der ursprünglichen Pipeline (PLG)

Der Eckpfeiler dieser Bemühungen war das Multi-Level Stress Predictor (MuStP)-Modell (Skizze siehe Abb. 18), das sorgfältig entwickelt wurde, um subjektive EMAs (Soft Labels) in einen hierarchischen Rahmen zu integrieren. Diese mehrstufige Struktur verarbeitete effizient physiologische Daten von Smartwatches mit unterschiedlichen zeitlichen Auflösungen.

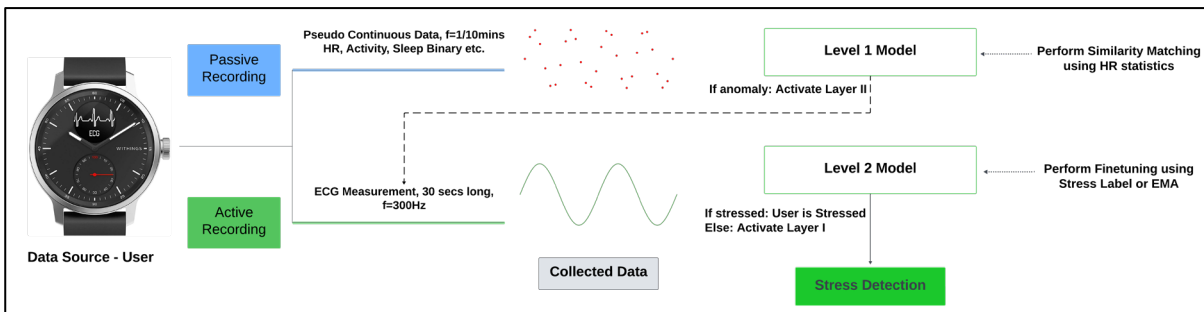


Abb. 18: Skizze für Multi-level Stress Predictor (MuStP) (PLG)

Um die Nutzerbeteiligung weiter zu erhöhen und das Stressbewusstsein zu fördern, implementierte die PLG einen nutzerzentrierten Feedback-Mechanismus. Dieser liefert tägliche Zusammenfassungen, die algorithmische Stressvorhersagen klar mit den subjektiven Stresseinschätzungen der Nutzer:innen vergleichen. Diese Funktionalität steigerte das Nutzerbewusstsein deutlich und verbesserte die praktische Anwendbarkeit des entwickelten Stresserkennungssystems. Zusätzlich arbeitete die PLG mit dem Fraunhofer IGD zusammen, um grundlegende Algorithmen für unüberwachte und überwachte Stresserkennung zu entwickeln, die zukünftige vergleichende Evaluationen erleichtern sollen. Gemeinsam definierten die Teams mögliche psychologische Interventionen, die in die entwickelten Machine-Learning-Modelle integriert wurden. Angesichts der Herausforderungen der kontinuierlichen Datenerfassung konzipierte die PLG die Pipeline sorgfältig, um sowohl im Labor als auch in der Praxis (Smartwatch) eine hohe Vorhersageleistung zu gewährleisten. Die Pipeline wurde erfolgreich fertiggestellt, getestet und implementiert und zeigte eine robuste Leistung in überwachten und halbüberwachten Szenarien.

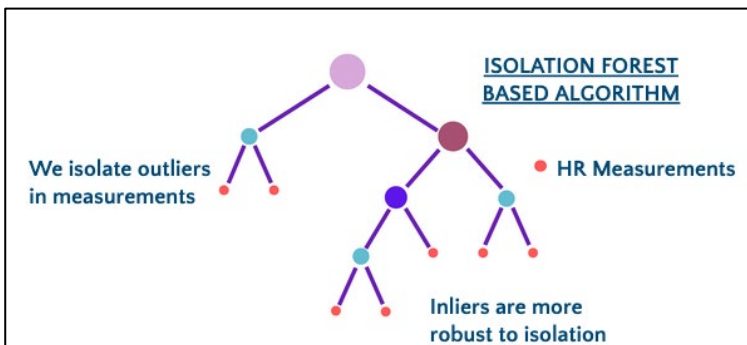


Abb. 19: Auf Isolation Forest basierender Algorithmus zur Anomalieerkennung, verwendet in Level 1 des MuStP-Modells („Vom Labor in den Alltag: Personalisierte Stressvorhersage über Smartwatches“; International Conference of Machine Learning'24; Workshop „ML for Life and Material Science: From Theory to Industry Applications“: <https://openreview.net/forum?id=XBtTaeQOWs>) (PLG)

Die PLG hat mit dem Fraunhofer IGD zusammengearbeitet, um Algorithmen für unüberwachte und überwachte Stresserkennungsmodelle zu verwenden. Diese Ergebnisse können in Zukunft als Baseline verwendet werden. Gemeinsam wurde daran gearbeitet, eine Reihe möglicher psychologischer Interventionen zu definieren, die in den entwickelten ML-Algorithmen empfohlen werden. Die PLG entwickelte ein Modell in überwachten und halbüberwachten Einstellungen zur Vorhersage von Stress anhand von Labor- und Smartwatch-Daten in einem mehrstufigen ML-Modell. Das vorgeschlagene Modell wurde entwickelt, um das Problem der kontinuierlichen Datenerfassung im Rahmen des Projekts zu überwinden und gleichzeitig Stress mit hoher Leistung vorherzusagen.

Die PLG hat das Modell in überwachten und halbüberwachten Einstellungen fertiggestellt, um Stress anhand von Labor- und Smartwatch-Daten in einem mehrstufigen ML-Modell vorherzusagen. In der Praxis verwendet PLG subjektive Nutzerbewertungen als Soft Labels. In die gesamte Pipeline hat PLG tägliche Zusammenfassungen für die Nutzer:innen aufgenommen, in denen ihre subjektive Einschätzung von Stress und die Vorhersagen des Algorithmus verglichen wurden, um ein Bewusstsein für die Nutzer:innen zu schaffen.

UKJ: (vgl. AP 2.3)

AP 2.5 (Monat 12-36; verantwortlich: PLG)

(ursprüngliches) Ziel: *Verwendung erklärbarer Algorithmus des maschinellen Lernens*

PLG: Die PLG konzentrierte sich hier auf die Aufmerksamkeit-basierten Modelle, da sie einen erklärbaren Rahmen für die Ausgaben des neuronalen Netzwerks boten. Diese könnten sowohl für die Erkennung von Stressursachen als auch für die Stressvorhersage geeignet sein. PLG untersuchte zunächst aufmerksamkeitsbasierte Deep-Learning-Frameworks, um die Interpretierbarkeit von Stressvorhersagemodellen zu verbessern. Aufgrund der diskreten Natur und Spärlichkeit der verfügbaren Datensätze, darunter auch zahlreiche fehlende Daten, waren diese jedoch nicht dicht genug, um aufmerksamkeitsbasierte Modelle effektiv zu trainieren. Daher verlagerte PLG seinen Fokus auf post-hoc-Erklärungsmethoden, wie beispielsweise die Interpretation persönlicher Entscheidungsschwellen der Nutzer:innen, um deren Tendenz zu Stressreaktionen zu erkennen. Diese Methoden wurden eingesetzt, um Muster aus EMAs neben Stressvorhersagen zu analysieren und so wertvolle Erkenntnisse über individuelle Stressauslöser und -reaktionen zu gewinnen. Die Erkenntnisse aus diesen Post-hoc-Analysen verbesserten die Modelltransparenz deutlich, erleichterten eine klarere Kommunikation mit dem Konsortium und stärkten das Nutzervertrauen sowie die Einhaltung der empfohlenen Stressinterventionen.

AP 2.6 (Monat 13-26; verantwortlich: UKJ)

(ursprüngliches) Ziel: *ELSI Organisation*

UKJ: Das UKJ wurde kurzfristig gebeten, weitere Details über die App und die Wearables in den Ethikantrag und die zugehörigen Unterlagen aufzunehmen und zusätzliche Zertifikate für die Wearables, die App und die IT-Infrastruktur einzureichen. Dies wurde bearbeitet und abgeschlossen und es waren schlussendlich nur noch der unterzeichnete erweiterte Kooperationsvertrag und die Forschungsvereinbarung mit Withings einzureichen, um mit der Hauptfeldstudie beginnen zu dürfen. Im weiteren Verlauf wurden Kooperationsvertrag, die Vereinbarung zur gemeinsamen Verantwortung über die Datenverarbeitung und die Forschungsvereinbarung mit Withings eingereicht und akzeptiert.

AP 2.7 (Monat 13-26; verantwortlich: UKJ)

(ursprüngliches) Ziel: *Überprüfung der App unter Realbedingungen*

UKJ: Ziel dieses Arbeitspaketes war die Überprüfung/Testung der App unter Realbedingungen und damit die Validierung der App und ihrer Prognosefähigkeit. Vor diesem Hintergrund sollten über einen Zeitraum von sechs Monaten mit der App 300 (150 männliche, 150 weibliche) Proband:innen mit gleichmäßiger Altersverteilung zwischen 20 und 70 Jahren gemessen werden. Hierzu sollten 40 Smartwatches zur Messung bereitgestellt werden. Die Smartwatches standen schlussendlich zur Verfügung, jedoch verschob sich die Durchführung dieses Arbeitspaketes aufgrund der Corona-Beschränkungen und der darauffolgenden bereits beschriebenen Startschwierigkeiten sowie wegen der unvorhergesehen weiteren Verzögerungen nach Mai 2023. Diese Hauptfeldstudie (Feldstudie I) wurde in Tübingen durchgeführt (Sept. 2023-Dezember 2023) und mit 147 Personen (davon 100 weiblich) abgeschlossen. Mithilfe der Withings Scanwatches und der Pulsatio-App wurden über einen Zeitraum von drei Monaten physiologische, verhaltensbezogene und subjektive Daten erhoben. Diese Ergebnisse der Studie wurden verwendet, um die Stresserkennung zu validieren, Daten für die mittel- und langfristige Stressvorhersage bereitzustellen und Einblicke zu geben, welche psychologischen Interventionen in der App implementiert werden mussten.

Modul 3: Stresscharakterisierung

AP 3.1 (Monat 1-6; verantwortlich: UKJ)

(ursprüngliches) Ziel: Bestimmen der allgemeinen zeitlichen Muster der Stressreaktivität, Definieren der Hypothese

UKJ: Das Ziel war für die ersten Monate die Bestimmung der allgemeinen zeitlichen Muster der Stressreaktivität, die Hypothesenbildung und abschließend die Testung und Nutzung der Hypothese in der entsprechenden App. Die detaillierte Literaturrecherche zur Stresscharakterisierung und Stressreaktivität sowie die Analyse des öffentlichen WESAD Datensatzes in Rostock und Saarbrücken erfolgten planmäßig. Da die Pilotstudie durch Messungen zur Geräteevaluation und durch Daten aus alten und aktuellen Projekten ersetzt wurde, wurden den Projektpartnern der IGD Rostock und PLG Saarbrücken diese Datensätze verfügbar gemacht und dort analysiert.

PLG/UKJ: Die PLG und das UKJ arbeiteten hier eng zusammen, um detaillierte Analysen bestehender Datensätze durchzuführen, darunter der öffentlich zugängliche WESAD-Datensatz sowie weitere Datensätze aus Tübingen und Jena. Diese Analysen identifizierten und charakterisierten erfolgreich unterschiedliche zeitliche Muster der Stressreaktivität und lieferten so wichtige Informationen für nachfolgende Versuchsdesigns sowie eine verbesserte Genauigkeit der Stresserkennungsmodelle. Die PLG führte eine umfassende Literaturrecherche zur Stresscharakterisierung und -reaktivität und dazu eingehende explorative Analysen des WESAD-Datensatzes durch. Diese in Rostock und Saarbrücken durchgeführten Analysen verliefen wie geplant. Erste Ergebnisse zeigten deutliche Unterschiede in den individuellen Stressreaktionen und unterstrichen die Bedeutung einer personalisierten Modellierung. Darüber hinaus identifizierte die PLG Inkonsistenzen bei der Kennzeichnung sowie Baseline-Probleme im WESAD-Datensatz. Diese wurden den Projektpartnern mitgeteilt, um Verbesserungen zukünftiger Datenerhebungsmethoden im Projekt zu ermöglichen. Zu den spezifischen numerischen Ergebnissen der von PLG durchgeführten Analysen gehören überwachte Klassifizierungsexperimente mit Transformer-Modellen, bei denen Genauigkeiten von bis zu 87 % erreicht wurden (mit Einbettungsdimensionen von 64 und Lernraten von 0,0003). Darüber hinaus wurden bei der unbeaufsichtigten Merkmalsextraktion mit Faltungs-Autoencoder-Architekturen Rekonstruktionsfehler (MSE) von nur etwa 0,0029 erzielt (bei einer Konfiguration mit 32 bis 64 Schichten und einer Lernrate von 0,0003), was eine effektive Merkmalsextraktion aus spärlichen physiologischen Daten belegt. Aufgrund von Änderungen in der Projektplanung wurde die ursprünglich geplante Pilotstudie durch Auswertungen von Messgeräten und Analysen vorhandener Datensätze ersetzt. Diese Datensätze wurden anschließend dem IGD Rostock und der PLG Saarbrücken zur weiteren Analyse zur Verfügung gestellt. Die aus diesen Untersuchungen gewonnenen Erkenntnisse trugen zusammen mit laufenden Datenqualitätsbewertungen und Merkmalsextraktionsstrategien wesentlich zur Verfeinerung und Verbesserung der im Projekt verwendeten Versuchsprotokolle und Stressvorhersagemodelle bei.

Nachdem die Datensätze im Rahmen des Projekts gesammelt worden waren, hat die PLG die Strategie der Feinabstimmung pro Nutzer:in übernommen (Abb. 20), die auch im MuStP-Modell zu sehen ist. Dieser Ansatz hat die Verallgemeinerung verbessert, indem er die Unterschiede in den Stressreaktionen zwischen den Nutzer:innen erfasst hat.

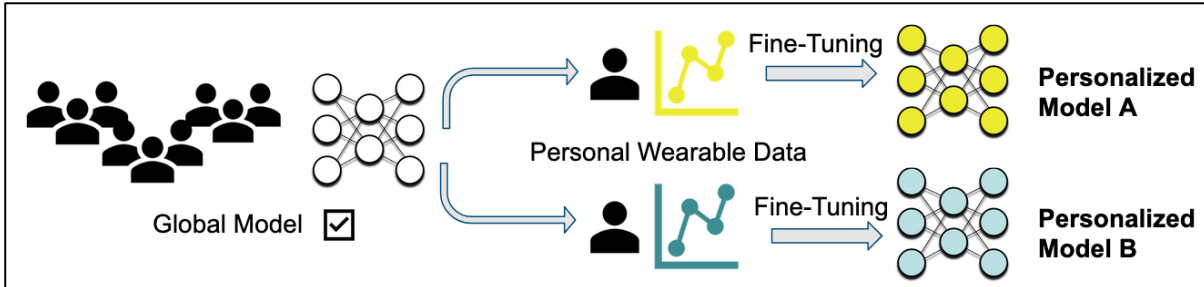


Abb. 20: Feinabstimmung der Modelle für die Personalisierung (PLG)

AP 3.2 (Monat 1-8; verantwortlich: UKJ)

(ursprüngliches) Ziel: Erstellung des Studienprotokolls zur Charakterisierung von Stress sowie Antrag auf Genehmigung durch die lokale Ethik-Kommission (Ethikvotum)

UKG: Die Studienprotokolle zur Charakterisierung von Stress wurden erstellt. Der Antrag auf Genehmigung der ersten Laborstudie durch die lokale Ethik-Kommission (Ethikvotum) wurde eingereicht und genehmigt und ein Amendement über die Smartwatches wurde vorbereitet. Der Antrag auf Genehmigung der zweiten Laborstudie durch die lokale Ethik-Kommission (Ethikvotum) wurde vorbereitet und eingereicht. Die Vorbereitungen der für die Durchführung der Experimente erforderlichen Software und Hardware und die Anpassung des Experiments wurden planmäßig vorgenommen. Die Aufzeichnung der Pilotdaten und die Durchführung einer vorläufigen Analyse der aufgezeichneten Pilotdaten sind oben beschrieben (siehe AP 3.1).

AP 3.3 (Monat 8-14; verantwortlich: UKJ, PLG)

(ursprüngliches) Ziel: ELSI Organisation

PLG: PLG arbeitete aktiv mit anderen Konsortialpartnern zusammen, um die Qualität und Effektivität von Ecological Momentary Assessments (EMAs) zu verbessern. Da EMAs wichtige Soft-Label-Daten sind, die von Nutzer:innen in realen Alltagskontexten erhoben werden, trug die PLG somit maßgeblich zur Optimierung des EMA-Designs und zur Analyse der EMA-Reaktionen mit Cortisol-Spiegeln bei (siehe Abb. 21 und 22). Dies umfasste Empfehlungen zu effektiven Fragestrategien, Antwortspeichertechniken und Ansätzen zur Minimierung von Datenverlusten aufgrund fehlender Teilnehmerreaktionen.

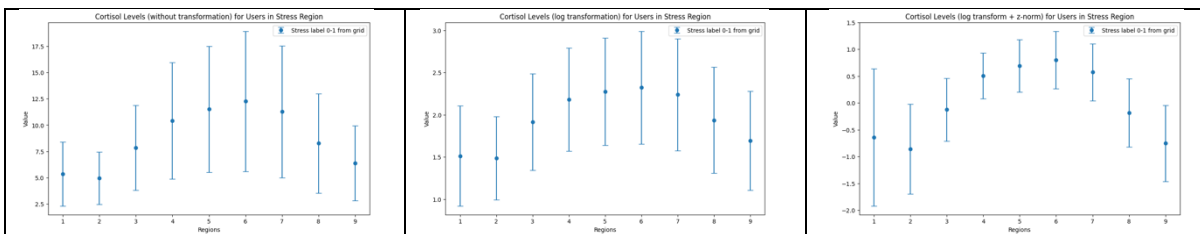


Abb. 21: Cortisolspiegel aus Zeitfenstern der Laborstudie (PLG)

Darüber hinaus spielte die PLG eine wichtige Rolle bei der Definition eines umfassenden Satzes möglicher psychologischer Interventionen, die in die entwickelten Machine-Learning-Modelle integriert werden sollten. Die PLG analysierte und lieferte detaillierte Einblicke in Datensätze aus Laborstudien von Partnern in Tübingen und Jena sowie in vorläufige Pilotdaten von Thryve. Diese Analyseergebnisse wurden proaktiv mit den Konsortialpartnern kommuniziert und ermöglichten fundierte Entscheidungen über die zukünftige Datenerhebung und Interventionsplanung.

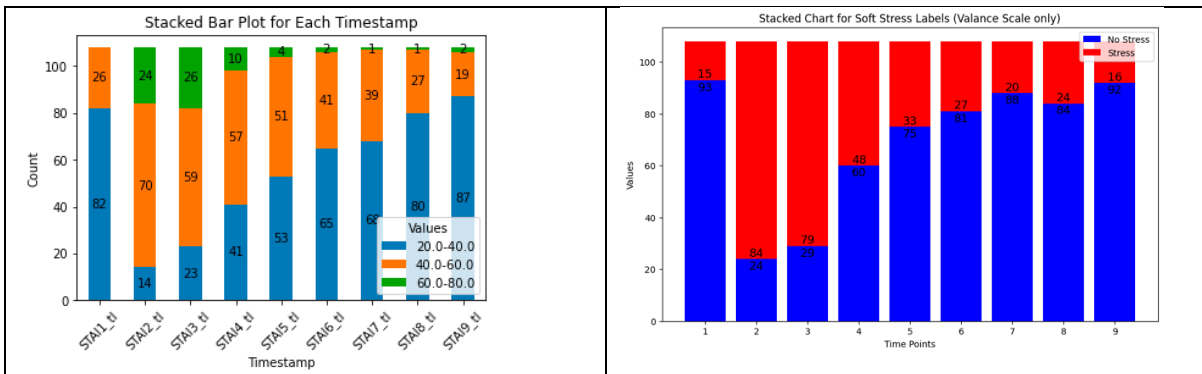


Abb. 22: Vergleich von STAI-Ergebnissen und Soft-Labels, die mit der Valenzskala erhoben wurden (PLG)

UKJ: Die Organisation der ethischen, rechtlichen und sozialen Aspekte (ELSI) am UKJ schritt weiter voran. Es lagen formulierte Einverständniserklärungen für die Proband:innen für beide Laborstudien zur Erfassung von Stressfaktoren und Charakterisierung individueller Stressunterschiede sowie entsprechende Aufklärungsbögen, die Erläuterungen zu Ein- und Ausschlusskriterien, zur Anonymisierung und Pseudonymisierung und zur Freiwilligkeit der Proband:innen und zur Gewährleistung von Datenschutz und -sicherheit vor. In Bezug auf Datenschutzerklärungen, Informationen für die Proband:innen und Ethikanträge kooperierte das UKJ mit Thryve.

AP 3.4 (Monat 3-9; verantwortlich: UKJ, Kollaboration: Thryve, PLG)

(ursprüngliches) Ziel: Vorbereitung der Hardware und Software Setups für die Experimente

UKJ: Am UKJ wurden die Hardware- und Teile des Software-Setups für die Experimente vorbereitet sowie eine entsprechende detaillierte Literaturrecherche durchgeführt. Anstelle von Daten aus der Pilotstudie wurden Daten aus Messungen der Geräteevaluierung sowie Daten aus älteren und aktuellen Projekten (s.u.) analysiert. Daran schloss sich die Anpassung der Experimente, Tests mit den Laborgeräten und den Wearables und die Schulung der Doktorand:innen und Hiwis zur Durchführung der Aufzeichnung und der Datenanalyse an.

PLG: Die PLG nahm aktiv an gemeinsamen Fokusgruppensitzungen mit dem UKJ und Thryve teil und leistete einen wichtigen Beitrag zu den Diskussionen über den Versuchsaufbau und die statistischen Eigenschaften, die für die Datenerfassung sowohl im Labor als auch im Alltag erforderlich sind. Um die Hardware gründlich zu verstehen und die Durchführbarkeit von Algorithmen für maschinelles Lernen zu testen, erhielt die PLG ein eigenes Gerät, das eine detaillierte Untersuchung und das Testen von Algorithmen ermöglichte. Entscheidend ist, dass die PLG die gesamte Datenerfassungspipeline unabhängig kodierte und alle Regeln und Verfahren, die für die Ausführung von maschinellen Lernmodellen, die effektive Nutzung von EMAs und die Bereitstellung genauer Stressvorhersagen erforderlich sind, sorgfältig integrierte. Darüber hinaus übernahm die PLG Aufgaben wie das Einlesen der Daten, die Pflege der Codebasis des Projekts, die Extraktion wichtiger Statistiken aus konsortialweiten Studien und die unverzügliche Übermittlung der Ergebnisse an alle Partner. Fehlende Nutzer und Fehler bei der Datenerfassung wurden identifiziert und dem Konsortium umgehend mitgeteilt, um die Datenintegrität zu gewährleisten. In Zusammenarbeit mit den Rostocker Partnern wurden umfassende Berichte erstellt, in denen auf fehlende Daten und die Verteilung von Soft-Labels der EMAs hingewiesen wurde.

Angesichts der Einschränkungen, die bei der kontinuierlichen Datenerfassung mit kommerziellen Smartwatches auftraten, passte die PLG ihren Fokus strategisch an. Während die ursprünglich geplanten Literaturübersichten über die Modellierung kontinuierlicher Zeitreihen vor allem für im Labor erhobene Daten (EKG, EDA, HRV, PPG, ACC-Sensoren) relevant waren, verlagerte die PLG den Schwerpunkt auf Modelle, die heterogene Daten mit unterschiedlichen zeitlichen Auflösungen und systematischen Ausfällen verarbeiten können.

Thryve: Bei Thryve wurden Teile des Software-Setups für die Experimente vorbereitet. Diese Teile umfassten: die Einrichtung der Studie und der Studiengruppe in der App, die Generierung der QR-Codes für das QR-Code-Onboarding der Studienteilnehmer:innen, das Implementieren der speziell für das Projekt entwickelten EMA-Fragebögen in der App, Implementation des Algorithmus zur automatischen Stresserkennung in die App, Einrichten des Datenexports-Servers und Erteilung der Zugriffsrechte zum Datenexport-Server (für alle Studienpartner).

AP 3.5 (Monat 9-12; verantwortlich: UKJ)

(ursprüngliches) Ziel: *Pilotstudie zur Erfassung von Stressfaktoren und Charakterisierung individueller Stressunterschiede*

UKJ: Um die ursprünglich geplante Pilotstudie zu ersetzen, wurde entschieden, vorhandene stressbezogene Daten aus alten und aktuellen Projekten (s.o.) zu nutzen. Aufgrund der Corona-Pandemie und der intensiven Suche nach geeigneten Smartwatches für das Projekt konnten, wie bereits beschrieben, die im Antrag formulierte Pilotstudie nicht durchgeführt werden. Stattdessen wurden über einen Zeitraum von drei Monaten Daten von drei Projektmitgliedern mit drei verschiedenen Smartwatches (Garmin, Withings und FitBit) aufgezeichnet (s.o.). Zusätzlich wurden drei verschiedene Datensätze aus alten und aktuellen Projekten zur Verfügung gestellt, um Stresserkennungsansätze zu validieren und Stress zu charakterisieren. Die Projektpartner IGD Rostock und PLG Saarbrücken haben 2021 mit der Arbeit am öffentlichen Datensatz WESAD (s.o.) begonnen. Er enthält Aufnahmen von 15 Proband:innen. Die vorliegenden Daten wurden mit dem RespiBAN Professional Brustgurt und einem am Handgelenk tragbaren Empatica E4 Gerät gemessen, um sowohl physiologische (BVP, EKG, EDA, EMG, RESP und TEMP) als auch Bewegungsdaten (ACC) zu erfassen.

Außerdem wurde den Projektpartnern aus Rostock und Saarbrücken ein Datensatz aus Tübingen (Legacy Dataset I) und ein Datensatz aus Jena (Legacy Dataset II) mit 30 Proband:innen zur Verfügung gestellt, um ihre Ansätze zur Stresserkennung zu validieren. Der erstgenannte Datensatz enthält Daten von 140 Proband:innen. Im Experiment (Legacy Dataset I) wurden Hormondaten (Cortisol), subjektive Daten, fMRI und PPG und im zweitgenannten Datensatz (Legacy Dataset II) wurden HRV-, EDA- und EMG-Daten über das Multisensorsystem NeXus10-Gerät erhoben. Das Paradigma erzeugt Stresslevel unterschiedlicher Intensität in aufsteigender Reihenfolge. Dieser Datensatz sollte helfen, eine verbesserte Stresserkennung zu validieren.

Alle bestehenden wie auch im Projekt noch zu erfassenden Daten dienen der vollständigen Testung der in der Entwicklung befindlichen App und dem Vergleich der Ergebnisse mit hochwertigen Daten der Laborsystemen.

Das UKJ beabsichtigte außerdem, eine interne Datenerhebung mit 20 Mitgliedern des Cello-Projekts durchzuführen. Schließlich wurde eine Laborstudie (Laborstudie II in Jena) entworfen, die den Ansprüchen der geplanten Pilotstudie entsprach. Diese zweite, ergänzende Laborstudie begann im Sommer 2022 und wurde Anfang des Jahres 2023 weitergeführt. Aufgrund der beschriebenen Störungen in der EEG-Anlage im Nov. 22 (vgl. Zwischenbericht) konnten die Scans für die Studie erst im Januar 2023 fortgesetzt und die gesamte Stichprobe (n=22) wurde bis Ende April 2023 gescannt. Es wurden die Parameter Cortisol, EKG und EEG sowie MRT, MRS, fMRT und subjektive Daten erfasst und ausgewertet. Diese Studie erfüllte die gleichen Ziele wie die ursprünglich geplante Pilotstudie zur Stresscharakterisierung. Dieser Datensatz war von hoher Relevanz für die Untersuchung wichtiger Merkmale in Bezug auf maschinelles Lernen in der Stressforschung.

Darauf folgte die Vorverarbeitung der Daten (April-August 2023). Es wurden auch in dem Zeitraum ab September 2023 modalitätsübergreifende Datenanalysen durchgeführt. In diesem Zeitraum wurde planmäßig eine Doktorandenstudie ausgewertet, in der der Zusammenhang zwischen Stress in der Kindheit, aktuellem Stress und der Dynamik der Herzfrequenzvariabilität untersucht wurde (andauernd, siehe Punkt 5.). Außerdem wurde eine Analyse und ein Manuskript erstellt, in denen die Dynamik des Ruhezustands unter Verwendung der Metastabilitätsmetrik untersucht wurde. In einer weiteren Doktorarbeit wurde der Zusammenhang zwischen Kindheitsstress, Glutamat und subjektivem Stress analysiert (siehe Punkt 5., Veröffentlichung zu

dynamischer Konnektivität und im Labor gemessenen physiologischen Gehirn- und Herzdaten, um subjektive Selbsteinschätzungen und Cortisolwerte vorherzusagen, ist in Arbeit). Darüber hinaus arbeiteten weitere Doktorand:innen an der Analyse physiologischer, verhaltensbezogener und subjektiver Daten. Außerdem wurde eine Analyse und ein Manuskript erstellt, in denen die Dynamik des Ruhezustands unter Verwendung der Metastabilitätsmetrik untersucht wurde (zu Veröffentlichungen siehe Punkt 5.).

PLG: Als Ersatz für die ursprünglich geplante Pilotstudie wurden bestehende stressbezogene Datensätze aus vergangenen und laufenden Projekten verwendet. Aufgrund von Einschränkungen durch die COVID-19-Pandemie und Verzögerungen bei der Suche nach geeigneten Smartwatches wurde die ursprüngliche Pilotstudie nicht durchgeführt. Stattdessen wurden die Daten von drei Projektmitgliedern über drei Monate hinweg mit Garmin-, Withings- und FitBit-Smartwatches aufgezeichnet. Zusätzlich wurden drei alte Datensätze aus vergangenen und laufenden Projekten verwendet. Die Probabilistic Learning Group (PLG) analysierte alle zur Verfügung gestellten Datensätze umgehend und gründlich. Dazu gehörte der öffentlich zugängliche WESAD-Datensatz, der physiologische (BVP, EKG, EDA, EMG, RESP, TEMP) und Beschleunigungsdaten (ACC) von 15 Teilnehmern enthält, die mit dem RespiBAN Professional-Brustgurt und dem Empatica E4-Armband erfasst wurden. Darüber hinaus führte PLG eine umfassende Validierung und Entwicklung von Algorithmen für maschinelles Lernen mit dem Tübinger Legacy-Datensatz I durch, der Hormon- (Cortisol), subjektive, fMRI- und PPG-Daten von 140 Teilnehmern umfasst, sowie mit Legacy-Datensatz II aus Jena, der HRV-, EDA- und EMG-Daten von 30 Teilnehmern enthält. Alle diese alten Datensätze wurden von der PLG systematisch zur Validierung der Stresserkennungsmethoden und zur Verfeinerung der Stresscharakterisierungsmodelle verwendet, wie die verschiedenen Stressreaktionen in Abbildung 13 (links) zeigen. Diese Analysen bildeten eine solide Grundlage für die anschließende Entwicklung und Validierung der App im Rahmen des Cell-o-Projekts. Darüber hinaus unterstützten diese Ergebnisse detaillierte Vergleiche mit hochwertigen Labordaten. Die gründliche und prompte Validierung durch PLG gewährleistete die Zuverlässigkeit und Bereitschaft der maschinellen Lernmodelle für die reale Anwendung in den laufenden Studien des Cell-o-Projekts.

AP 3.6 (Monat 11-14; verantwortlich: UKJ, PLG)

(ursprüngliches) Ziel: *Analyse der Daten der Pilot-Stress-Studie, Aktualisierung des generellen Studienprotokolls, Aktualisierung/Anpassung des Ethikantrags* **Lösungsansatz:** *Detaillierte Literaturrecherche, Anpassung des Experiments, Vorbereitung der für die Durchführung der Experimente erforderlichen Software und Hardware, Schulung der Doktoranden und Hiwis*

UKJ: Die Daten der verschiedenen Smartwatches sowie die Analyse der Daten aus alten, öffentlichen und aktuellen Projekten (s.o., Ersatz für Pilot-Laborstudie aus 3.5) wurden abgeschlossen. Nach Rücksprache mit den Projektpartnern wurde eine Aktualisierung des Studienprotokolls und des Ethikantrags unternommen. Nach der Literaturrecherche über Stressdetektion und Smartwatches konnten Einblicke in die Merkmale und Signale gewonnen werden, die für Stresscharakterisierung relevant waren. All dies floss in die Vorbereitung und Durchführung der nachfolgenden Laborstudien in Jena ein. Die Datenanalyse aus der kleineren Laborstudie (N=22, siehe AP 3.5) erfolgte zeitnah nach Abschluss durch den Doktoranden.

PLG: PLG spielte eine zentrale Rolle bei der Analyse aller vom Konsortium gesammelten Daten. In Fällen, in denen studienspezifische Daten noch nicht zur Verfügung standen - insbesondere bei verzögerten Pilotstudien - führte die PLG unabhängig robuste Simulationen unter Verwendung vorhandener Datensätze durch. Diese Simulationen waren entscheidend, um die Forschungsdynamik aufrechtzuerhalten und dem Konsortium verwertbare Erkenntnisse zu liefern.

Die von der PLG durchgeführten Simulationen untersuchten umfassend die Leistung der entwickelten Pipeline für maschinelles Lernen unter verschiedenen Eingabebedingungen. Insbesondere bewertete die PLG das Verhalten des Algorithmus bei verschiedenen Signaltypen, einschließlich niedrig aufgelöster Herzfrequenz, hoch aufgelöster Herzfrequenz und EKG-Messungen. Darüber hinaus wurden detaillierte Analysen der Feuerungsraten in verschiedenen Phasen der Stressreaktion und der gesamten End-to-End-Modellpipeline durchgeführt, die für die Analyse des Verhaltens des entwickelten Algorithmus für maschinelles Lernen unter

verschiedenen Bedingungen unerlässlich sind. Diese Ergebnisse zeigten nicht nur die Widerstandsfähigkeit der PLG-Algorithmen bei unterschiedlicher Datenqualität, sondern gaben auch Anhaltspunkte für künftige Feld- und Labor Datensammlungen. Nach einer detaillierten Literaturrecherche (siehe Anhang) zu Stresserkennungs- und Smartwatch-Funktionen arbeiteten PLG und UKJ gemeinsam an der Aktualisierung des allgemeinen Studienprotokolls und der Verfeinerung des Ethikantrags, um ihn an die neuesten Erkenntnisse und Projektanforderungen anzupassen. Die aus diesen Überprüfungen und Simulationen gewonnenen Erkenntnisse waren für die Vorbereitung und Durchführung der nachfolgenden Laborstudien in Jena von entscheidender Bedeutung, um robuste, ethisch konforme und wissenschaftlich fundierte experimentelle Rahmenbedingungen zu gewährleisten. Die umfangreichen Simulationen und Analysen der PLG lieferten dem Konsortium wichtige Informationen, die eine fundierte Entscheidungsfindung und Protokolloptimierung ermöglichten.

AP 3.7 (Monat 14-24; verantwortlich: UKJ)

(ursprüngliches) Ziel: ELSI Organisation

UKJ: Hier bestand die Aufgabe insbesondere vor allem darin, die Einverständniserklärung der Teilnehmer:innen für die Hauptstudie (s.u., AP 3.8) zu überprüfen und ggf. anzupassen. Die Hauptstudie diente der Erfassung von Stressfaktoren und der Charakterisierung individueller Stressunterschiede. Das Arbeitspaket wurde weitestgehend abgeschlossen

AP 3.8 (Monat 14-20; verantwortlich: UKJ)

(ursprüngliches) Ziel: Hauptstudie zur Erfassung von Stressfaktoren und Charakterisierung individueller Stressunterschiede

UKJ: Hier bestand das Ziel darin, die Hauptstudie mit 60 Proband:innen (30 Männer, 30 Frauen) zur Erfassung von Stressfaktoren und zur Charakterisierung individueller Stressunterschiede durchzuführen und basierend darauf Verbindungen zwischen neutraler und psychophysischer Metrik von Stressreaktivität vorherzusagen. Hierfür sollten Experimente durchgeführt werden, in denen Stress ausgelöst und die Erholungszeit der Proband:innen ohne Stressintervention analysiert werden. Dabei sollten neuronale (EEG/fMRI), physiologische (EKG, Atemfrequenz, Hautwiderstand), genetische und hormonelle Daten erfasst werden. Diese Hauptlaborstudie (Laborstudie I) zu Stress mit N=80 (mit den Withings-Smartwatches, der App, HRV-Signalen vom Biopac-System) hat mit Verzögerung im Juni 2023 gestartet. Aufgrund der bereits beschriebenen Probleme (Lieferverzögerungen beim Hersteller der Smartwatches, Datenschutz-Anforderungen am UKJ, Überarbeitungsanforderungen seitens der IT-Abteilung im UKJ und entsprechende Datenschutz-Neuvereinbarungen bzgl. der gemeinsamen Verantwortlichkeit mit Thryve) musste der Startpunkt der Studie mehrfach verschoben werden. Es wurden schlussendlich Parameter wie bestimmte stressspezifische Hormone (Cortisol), EKG und weitere physiologische Signale (mit Biopac-System) sowie subjektive Daten erhoben. Parallel dazu trugen die Proband:innen Withings Scanwatches und Empatica E4 Wrist Wearables. Die Daten, die hier generiert wurden, orientierten sich an den im Antrag beschriebenen zu erfassenden Parametern und dienten der Entwicklung und späteren Testung der Stresserkennung und der Stressvorhersage. Dieser Datensatz war für den Vergleich und das Ranking von stressrelevanten Merkmalen wertvoll. Insbesondere die parallele Erfassung physiologischer und verhaltensbezogener Daten (über die Wearables) war von hohem Wert für die Weiterentwicklung der Stresserkennung.

Die Stichprobe aus Laborstudie I (N=80) hat zusätzlich an einer neuen, ursprünglich nicht geplanten, vier Tage dauernden ersten (Pilot)-Feldstudie im häuslichen Umfeld teilgenommen. Sechsmal täglich führten die Proband:innen (Beginn im Mai 2023) eine Cortisolprobe, eine hochauflösende Herzfrequenz-Messung (5 Minuten) und anschließende EKG-Messung (30 Sekunden) und EMAs durch. Diese Studie (Pilotfeldstudie I) diente als Pilotprojekt für die größere Feldstudie und zur Bereitstellung von Datensätzen für die Erkennung von akutem Stress und die Modellierung der mittelfristigen Stressvorhersage.

AP 3.9 (Monat 16-24; verantwortlich: UKJ, PLG)

(ursprüngliches) Ziel: Analyse der Daten der Hauptstudie, Testung der Hypothese

Im Anschluss an AP 3.8 sollten in AP 3.9 die Daten der Hauptstudie analysiert und die Hypothese getestet werden. Hierfür sollte eine detaillierte Datenanalyse durchgeführt und die bestehenden Befunde und Daten ausgewertet und darüber hinaus eine intensive Literaturrecherche in Vorbereitung zur Testung der Hypothese vorgenommen sowie Doktorand:innen und Hiwis geschult werden.

UKJ: Dieses Arbeitspaket erfolgte erst nach Abschluss der Studie Ende des Jahres 2023 (siehe oben). Im Anschluss an AP 3.8 wurde seit Ende des Jahres 2023 die detaillierte Datenanalyse durchgeführt.

PLG: Die PLG teilte die Ergebnisse der Stressvorhersage durch MuStP unter Verwendung der verfügbaren Daten mit den Projektpartnern. Die Ergebnisse umfassen die Auswirkungen verschiedener Personalisierungs- und Modellübertragungstechniken (vgl. Abb. 10 sowie Veröffentlichungen Punkt 5.). PLG legte großen Wert darauf, sowohl kontextuelle (Labor vs. Alltag; vgl. Abb. 24) als auch individuelle Unterschiede bei Stressreaktionen zu verstehen (Abb. 23). Es ist wichtig hervorzuheben, dass Labordaten verlässliche Stress-Labels zum Trainieren maschineller Lernmodelle liefern, während Alltagsdaten Soft-Labels liefern.



Abb. 23: Skizze von Nutzer:innen mit unterschiedlichen Grundeinstellungen und Stressreaktionen (PLG)

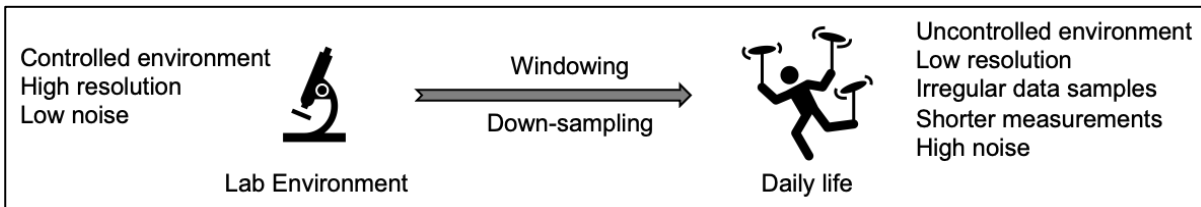


Abb. 24: Vergleich von Laborumgebung und Alltagsumgebung in Bezug auf die Datenverfügbarkeit (PLG)

Das Team stellte fest, dass sich der Stress bei den einzelnen Teilnehmer:innen und in verschiedenen Umgebungen unterschiedlich manifestierte - ein Hinweis darauf, dass ein allgemeines, bevölkerungswieites Modell für maschinelles Lernen möglicherweise nicht zuverlässig funktioniert, wenn es in realen Umgebungen eingesetzt wird. Als Reaktion auf diese Herausforderung entwickelte und implementierte die PLG eine Strategie, die sich auf die Personalisierung des Modells und die Anpassung an den jeweiligen Bereich konzentrierte, indem sie die MuStP-Pipeline (Multi-level Stress Predictor, s.o.) entwickelte. MuStP wurde als zweistufiges Modell aufgebaut:

- Ebene 1 verwendete niedrig aufgelöste Herzfrequenzdaten zur Erkennung von Anomalien durch Isolationswälder.
- Stufe 2 war ein faltungsbasierter LSTM-Klassifikator, der hochauflösende EKG-Daten für die endgültige Stressvorhersage verarbeitete.

Das Modell wurde anhand von Labordaten (LabData), einem Datensatz aus einer kontrollierten Laborstudie mit 99

Teilnehmer:innen, die mit dem Trier Social Stress Test (TSST) Stress ausgesetzt waren, trainiert und dann anhand von Alltagsdaten

(EVERYDAYDATA) von 131 Nutzer:innen mit kommerziellen Smartwatches (WITHINGS Scanwatch) auf die reale Welt übertragen.

Um den Domänenwechsel zwischen Labor- und Alltagsumgebung zu bewältigen, nutzte die PLG:

- Similarity Matching (SM) zur Auswahl benutzerspezifischer Anomalie-Detektoren für Stufe 1.
- Modell-Feintuning (FT) der letzten Klassifikatorschicht in Stufe 2 zur Anpassung der Vorhersagen an individuelle physiologische Muster.
- Post-hoc-Schwellenwert-Optimierung (PO) zur Maximierung der individuellen F1-Ergebnisse während der Inferenz.

Zusammenfassung der Ergebnisse (aus der MUSTP-Studie):

- Die personalisierte Modellübertragung und -optimierung steigerte die F1-Punktzahl auf 52 % bei Smartwatch-Daten aus der realen Welt - eine beeindruckende Verbesserung angesichts der unausgewogenen Verteilung der Stressbezeichnungen (nur ~31 % wurden als Stress bezeichnet).
- Die Feinabstimmung der Stufe 2 steigerte die Genauigkeit bei Labordaten von 77 % auf 82 %.
- Die hierarchische Architektur ermöglichte es dem Modell, den Aufwand für den Benutzer:innen zu minimieren, indem es EKG-Messungen nur bei Bedarf anforderte und sich hauptsächlich auf niedrig aufgelöste HR-Signale für eine effiziente Echtzeitnutzung stützte.

Durch den Aufbau einer flexiblen Pipeline, die sich an verschiedene Personen und Kontexte anpassen lässt, hat die PLG eine der größten Herausforderungen bei der Stressmodellierung - Generalisierung versus Personalisierung - erfolgreich bewältigt und eine Modellarchitektur entwickelt, die sowohl wissenschaftlich robust als auch praktisch einsetzbar ist.

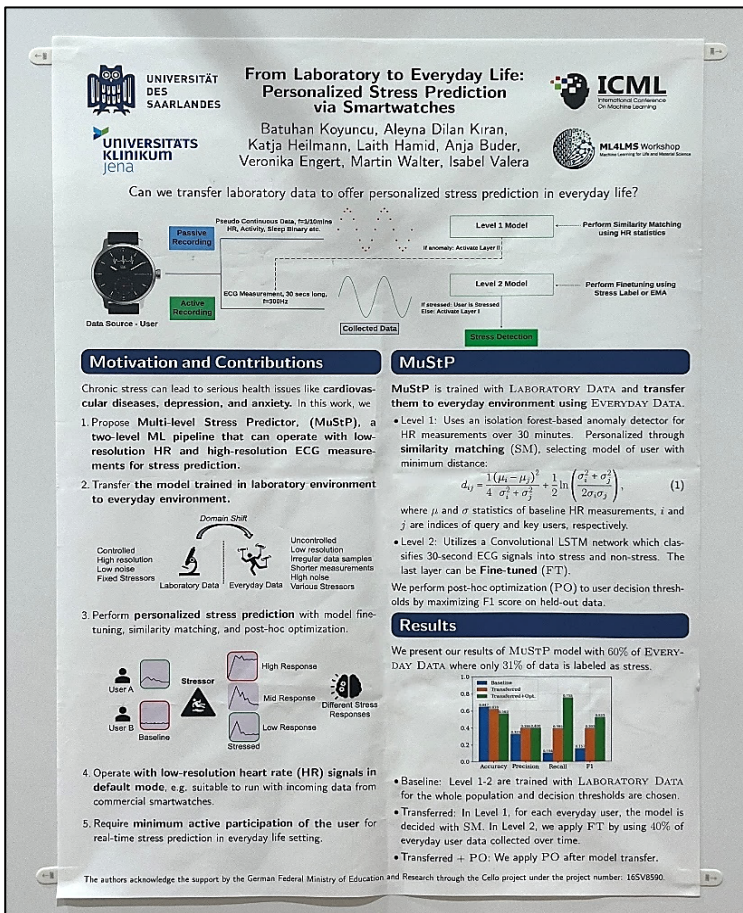


Abb. 25: ML4LMS Poster Juni 2024: „From Laboratory to Everyday Life: Personalized Stress Prediction via Smartwatches“ Accepted at the 1st Machine Learning for Life and Material Sciences, Workshop at ICML 2024. Copyright 2024 by the author(s) (PLG)

Diese Erkenntnisse wurden innerhalb des Konsortiums verbreitet und in einer Präsentation mit dem Titel "From Laboratory to Everyday Life: Personalized Stress Prediction via Smartwatches" auf dem ICML 2024 Machine Learning for Life and Material Sciences Workshop in Wien vorgestellt (Abb. 25).

AP 3.10 (Monat 16-26; verantwortlich: UKJ)

(ursprüngliches) Ziel: *ELSI Organisation*

UKJ: Aufgrund der Rückmeldungen aus dem Konsortium bzgl. der technischen Nichtumsetzbarkeit wurde gemeinsam entschieden, nicht mehr auf Biofeedback hinzuarbeiten. Stattdessen sollte ein umfangreicher Feldtest mit der finalen Version der App und der selbstlernenden KI durchgeführt werden (N=60). Da hierfür der damalige Zeithorizont ca. sechs bis acht Monate nach Projektende lag, wurde die Beantragung eines Verlängerungsantrages mit Kostenneutralität hingewiesen, der schließlich an den Projektträger gestellt und genehmigt wurde. Statt Biofeedback wurde ein umfangreicher Feldtest mit der finalen Version der App und der selbstlernenden KI durchgeführt (Ziel: N=60). Im Jahr der Projektverlängerung wurde mit Live-Tests begonnen, um die App und die Algorithmen zu testen.

AP 3.11 (Monat 24-26; verantwortlich: UKJ)

(ursprüngliches) Ziel: *Pilotstudie zur Untersuchung der Effektivität des HRV-Biofeedbacks*

UKJ: Diese Studie wurde, wie oben beschrieben, nicht in der geplanten Form durchgeführt, da bereits zu einem früheren Zeitpunkt entschieden werden musste, Biofeedback auszuklammern. Die Machbarkeit der notwendigen Biofeedback-Komponenten in Echtzeit war nicht mit der erforderlichen Wirtschaftlichkeit bzgl. der Auswahl geeigneter Geräte vereinbar. Für Biofeedback geeignete Gerätetypen benötigen qualitativ hochwertige (reliabel, valide) Messungen der HRV. Wie eingangs bereits beschrieben haben die für optische HRV-Messungen geeigneten Geräte den Zugriff auf die Daten und die Steuerung der Messungen für externe Anwendungen gesperrt, ebenso den Datenzugriff auf Messungen, die nicht innerhalb eigener Applikationen mit Messprotokoll und Nutzerführung erstellt wurden. Für Cell-o standen somit keine geeigneten Gerätetypen für das geplante Biofeedback in Echtzeit zur Verfügung. Das Ziel der Ausrichtung auf in der Bevölkerung bereits verbreitete Sensorik war ein klares Ergebnis der Geräteauswahl unter dem Gesichtspunkt der Erfolgsaussichten für die wirtschaftliche Verwertung. Da diese primär an den Erstattungen des SGB V §20 ausgerichtet waren, war die Erstattung von Geräten nicht realisierbar, sodass auf bereits verbreitete Sensoriktypen abgestellt wurde. Die Messungen über EKG waren mit den für den Feldtest geplanten Geräten Apple Watch und Withings ScanWatch implementiert. Da die Auslösung der Messung und Synchronisierung der Daten durch die technischen Vorgaben der Hersteller leider nicht verlässlich in Echtzeit erfolgen konnten, war eine Echtzeit-Biofeedback-Funktion auch hier schwer machbar. Im zukünftigen Feldtest sollten daher die Latenzen ermittelt und dafür geeignete Feedbackverfahren getestet werden.

AP 3.12 (Monat 26-28; verantwortlich: UKJ, PLG)

(ursprüngliches) Ziel: *Analyse der Daten der Pilotstudie der Interventionsmaßnahmen und Aktualisierung des Hauptstudienprotokolls sowie des Ethikantrags*

PLG: Mit den aktualisierten Zielen in dieser Phase des Projekts wurden die Interventionen neu konzeptualisiert, indem die Nutzer:innen über positive Stressprognosen informiert wurden, die implizit als Stressmarker fungierten und signalisierten, dass die/der Nutzer:in möglicherweise in einen Stresszustand gerät und achtsam sein oder vorbeugende Maßnahmen ergreifen sollte. Diese Vorhersagen sollten das Bewusstsein der Nutzer:innen fördern und ihn möglicherweise dazu ermutigen, sich stressreduzierend zu verhalten (z. B. Atemübungen, Achtsamkeit). Die PLG spielte eine zentrale Rolle bei der Sicherstellung, dass diese Interventionen nicht nur genau und wissenschaftlich fundiert, sondern auch akzeptabel und nicht aufdringlich für die Endnutzer:innen sind. Da die

MuStP-Pipeline so konzipiert war, dass sie bei jeder Erkennung von Stress eine ökologische Momentaufnahme (EMA) auslöst, wurde deutlich, dass häufige Modellaktivierungen - insbesondere falsch-positive Ergebnisse - zu einer Belastung für die Nutzer:innen werden könnten, da sie zu oft aufgefordert werden (Abb. 26-28).

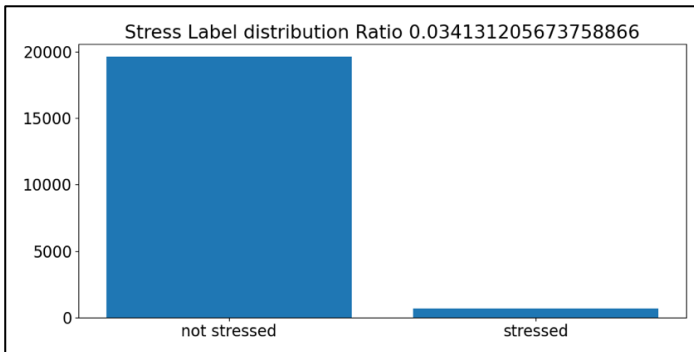


Abb. 26: Verteilung der EMAs (Soft Labels) für den Bestands-Datensatz I (PLG)

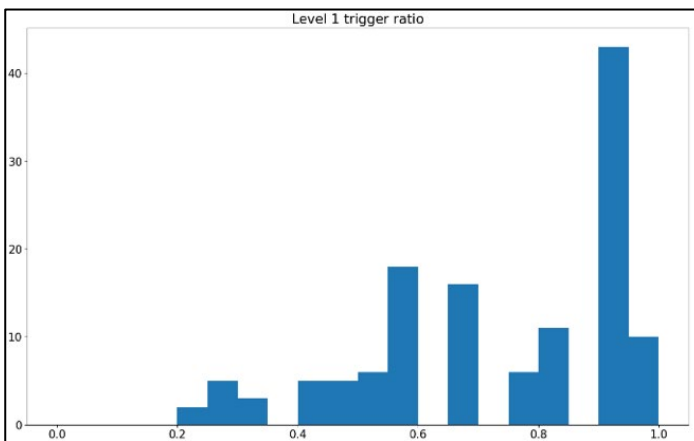


Abb. 27: Histogramm der Auslösungsverhältnisse für Stufe 1 des MuStP (PLG)

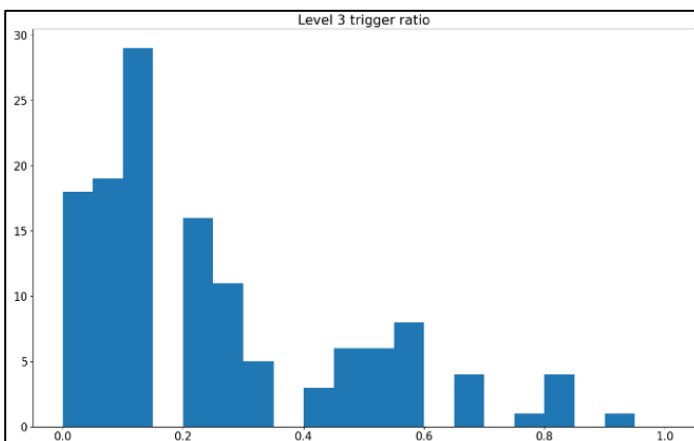


Abb. 28: Histogramm der Auslöserquoten für Stufe 2 des MuStP (PLG)

Um dieses Problem zu lösen, entwickelte und integrierte PLG eine Betriebsstrategie mit zwei Betriebsmodi in das Stresserkennungssystem:

- im Normalmodus löste das System EMAs bei der Erkennung von Stress und zu festen Tageszeiten (morgens, mittags, abends) aus, um eine konsistente Basiserfassung der subjektiven Stresszustände zu gewährleisten.

- im reduzierten Modus, der automatisch ausgelöst wurde, wenn die Erregungsrate des Modells einen von der/vom Benutzer:in einstellbaren Schwellenwert überschritt, wurden die Aufforderungen zu den EMAs vorübergehend unterdrückt, um den Benutzer weniger zu stören.

Dieser Mechanismus war entscheidend für:

- Verhinderung von „EMA-Ermüdung“ und des Benutzer-Ausscheidens,
- Erhaltung der Datenqualität durch Förderung überlegter und weniger reaktiver EMA-Antworten,
- Erhöhung der praktischen Anwendbarkeit und der Akzeptanz des Systems durch die Nutzer:innen.

Der Entwurf und die Implementierung dieser systemweiten Logik wurden mit den Projektpartnern geteilt und flossen direkt in die Aktualisierungen des Studienprotokolls und der ethischen Dokumentation ein. Dadurch wurde sichergestellt, dass die nutzerorientierte Interaktionsebene des Algorithmus ethisch vertretbar, personalisiert und möglichst wenig aufdringlich war, während gleichzeitig die Erhebung zuverlässiger Daten für die wissenschaftliche Auswertung unterstützt wurde. Diese Arbeit stärkte den Interventionsrahmen des Projekts, indem sie eine technisch robuste und dennoch menschenzentrierte Brücke zwischen maschinellen Vorhersagen und realem Verhalten schuf.

UKJ: (siehe 3.11)

AP 3.13 (Monat 26-30; verantwortlich: UKJ)

(ursprüngliches) Ziel: *ELSI Organisation*

(siehe 3.11)

AP 3.14 (Monat 26-30; verantwortlich: UKJ)

(ursprüngliches) Ziel: *Hauptstudie zur Testung der Effektivität von HRV-Biofeedback, Meditation, Mindfulness*

(siehe 3.11)

AP 3.15 (Monat 28-32; verantwortlich: UKJ)

(ursprüngliches) Ziel: *Analyse der Daten der Hauptstudie, Hypothesentestung*

(siehe 3.11)

Modul 4: Kalibrierung, Feldtest, und Optimierung

AP 4.1 (Monat 28-33; verantwortlich: UKJ, Kollaboration: PLG)

(ursprüngliches) Ziel: *Kalibrierung der Beurteilungen und Vorhersagen des App-Systems*

PLG/UKJ: Die PLG leitete die Kalibrierungsarbeiten für das Stressvorhersagesystem, indem sie eine umfassende statistische Analyse durchführte, bei der physiologische Messungen in Laborumgebungen (z.B. TSST-induzierter Stress) mit denen verglichen wurden, die im Alltag über Smartwatches erfasst wurden. Diese Analyse bestätigte bemerkenswerte Verteilungsunterschiede zwischen kontrollierten und realen Umgebungen, insbesondere bei der Herzfrequenz und den EKG-abgeleiteten Merkmalen. Diese Ergebnisse wurden umgehend mit den Partnern des Konsortiums geteilt und bildeten die Grundlage für die Aktualisierung der Eingangsverarbeitungspipeline des Modells. In der Phase der Systemkalibrierung war PLG federführend bei der Anpassung der Vorhersagen des maschinellen Lernmodells an die Gegebenheiten sowohl im kontrollierten Labor als auch in der realen Welt. Ziel war es, sicherzustellen, dass die Stressvorhersagen zuverlässig, interpretierbar und aussagekräftig bleiben, unabhängig von Kontext und Benutzer:in. Zunächst führte die PLG eine vergleichende statistische Analyse physiologischer Signale (z.B. Herzfrequenz und EKG) durch, die während der Stresstests im Labor (LABDATA) und unter realen Bedingungen (EVERYDAYDATA) erfasst wurden. Diese

Analysen bestätigten erhebliche Verteilungsverschiebungen, die Umwelt- und Verhaltensunterschiede widerspiegeln (z.B. ruhende Labormessungen vs. mobile Alltagsnutzung). Um dem zu begegnen, implementierte die PLG Skalierungs- und Normalisierungsverfahren direkt in die Pipeline, um sicherzustellen, dass die Eingaben aus den verschiedenen Bereichen konsistent und für die Modelle interpretierbar sind. Eine weitere wichtige Kalibrierungsdimension war die Verfeinerung von Soft-Labels, die aus EMAs abgeleitet wurden. Die PLG untersuchte mehrere EMA-Bezeichnungsschemata, die auf Variationen innerhalb des Affect Grid basieren, unter Verwendung von über 19.000 EMA-Einträgen aus dem EVERYDAYDATA-Datensatz. Diese Etikettierungsstrategien wurden dann anhand von im Labor gemessenen Cortisolwerten validiert, um festzustellen, welche Definitionen die physiologischen Stressreaktionen am besten widerspiegeln. Die Ergebnisse - gestützt durch eine quantitative Korrelationsanalyse - wurden den Konsortialpartnern vorgestellt und ausführlich mit Psychiater:innen und Stressforscher:innen (UKJ) diskutiert, um die klinische Interpretierbarkeit der ausgewählten Bezeichnungen sicherzustellen.



Abb. 29: Erfassen von Soft-Labels von EMAs mit Affect Grid (PLG)

Zusätzlich zur Kalibrierung auf der Eingabeebene implementierte die PLG die persönliche Anpassung durch Modellfeinabstimmung als Kernkomponente der Systemkalibrierung. In der MuStP-Pipeline (Multi-level Stress Predictor) wurden die vortrainierten Modelle anhand der Daten der Benutzer:innen weiter optimiert - eine Strategie, die die Generalisierung verbessert und falsch-positive Ergebnisse minimiert. Für Stufe 1 geht die PLG davon aus, dass die Basis-Herzfrequenzmessungen einer/eines Benutzer:in einer univariaten Normalverteilung mit den Parametern μ und σ folgen; daher berechneten die PLG die Ähnlichkeit mithilfe des Bhattacharyya-Abstands:

$$d_{ij} = \frac{1}{4} \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2} + \frac{1}{2} \ln \left(\frac{\sigma_i^2 + \sigma_j^2}{2\sigma_i\sigma_j} \right)$$

wobei i und j die Indizes der Test- bzw. Trainingsnutzer bezeichnen. Nachdem wir die Abstände $d_{ij}=[d_{i1}, \dots, d_{ij}]$ berechnet haben, finden wir den Index j^* , der d_{ij} minimiert.

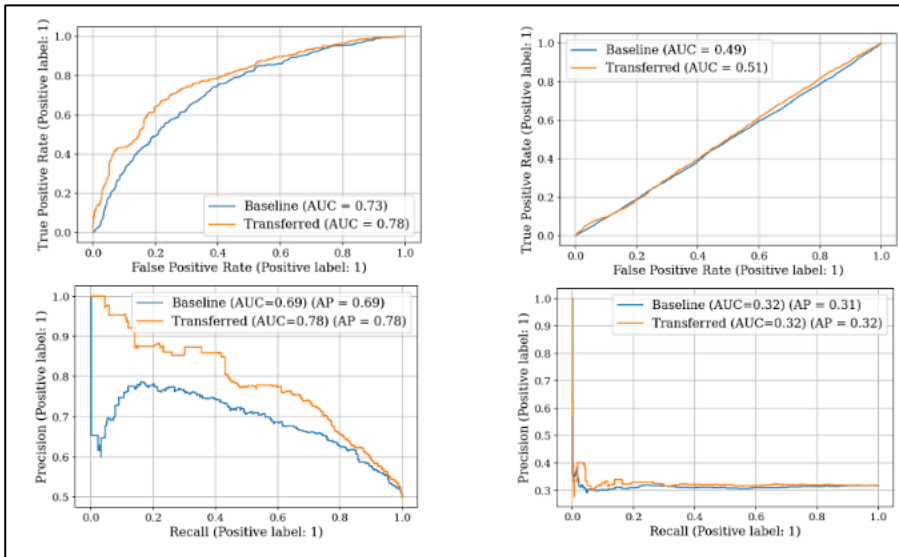


Abb. 30: Auswirkung des Ähnlichkeitsabgleichs in Stufe 1 des MuStP in einer Laborumgebung (links) und einer Alltagssituation (rechts); (PLG)

Für Stufe 2 verwendete die PLG eine Feinabstimmungsstrategie mit begrenzten Daten der Testnutzer:innen. Bei dieser Feinabstimmung wurden speziell die letzten Klassifizierungsschichten der tiefen neuronalen Netze angepasst, um die Ausgabe mit der einzigartigen physiologischen Grundlinie und der Stressreaktion jeder/jedes Nutzer:in in Einklang zu bringen. Wie in der MuStP-Evaluierung gezeigt wurde, erhöhte diese Personalisierungsstrategie die Modellgenauigkeit von 77 % auf 82 % bei Labordaten und trug dazu bei, einen F1-Score von 52 % bei Alltagsdaten zu erreichen, trotz der spärlichen und unausgewogenen Stresskennzeichnungen.

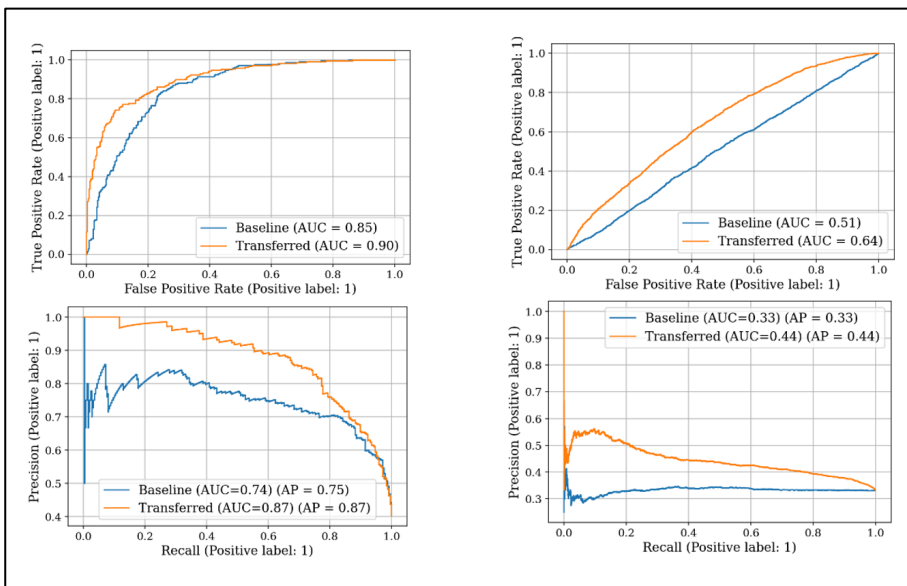


Abb. 31: Auswirkung der Feinabstimmung in Stufe 2 des MuStP in einer Laborumgebung (links) und einer Alltagssituation (rechts); (PLG)

Diese Kalibrierungsstrategien - statistische Normalisierung, Soft-Label-Optimierung und personalisierte Feinabstimmung des Modells - stellten sicher, dass das CELLO-App-System für einen zuverlässigen und ethisch vertretbaren Betrieb unter verschiedenen realen Bedingungen vorbereitet war. Alle Methoden und Ergebnisse wurden dokumentiert und mit den Mitgliedern des

Konsortiums geteilt und direkt in die aktualisierte Vorhersage-Pipeline des Systems integriert.

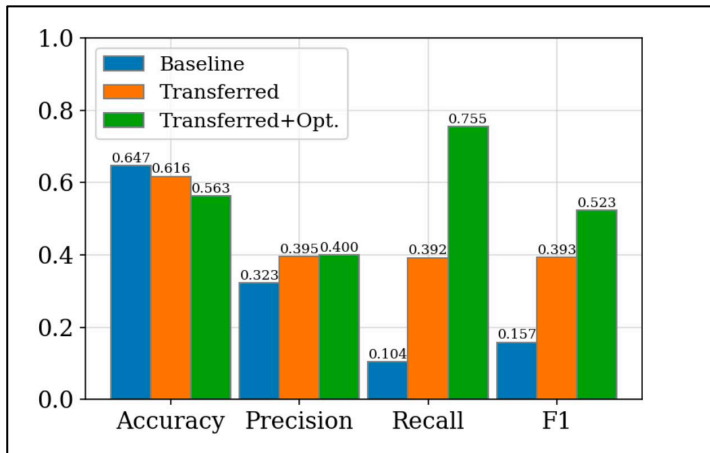


Abb. 31: Vergleich von Modellen im alltäglichen Umfeld (PLG)

Zur Bewältigung dieser Bereichsverschiebungen führte die PLG Skalierungs- und Normalisierungstechniken ein, um ein konsistentes Verhalten der Merkmale in verschiedenen Kontexten zu gewährleisten, bevor sie in maschinelle Lernmodelle eingespeist werden. Diese Kalibrierungsschritte verbesserten die Stabilität und Interpretierbarkeit der Vorhersagen in der Praxis.

Ein Hauptaugenmerk in dieser Phase lag auch auf der Verfeinerung von Soft-Label-Definitionen, die aus EMAs abgeleitet wurden. Die PLG untersuchte mehrere Kennzeichnungskriterien, die auf subjektiven Reaktionen beruhen (z.B. verschiedene Affektschwellen und Erregungs-/Valenzkombinationen), mit anderen Worten, wie die gesammelten EMAs zu einem einzigen binären Soft-Label zusammengefasst werden können. Die endgültige Entscheidung wurde mit den Konsortialpartnern des UKJ getroffen. Um festzustellen, welche Etikettierungsstrategie den tatsächlichen Stresszustand am genauesten widerspiegelt, führte die PLG Korrelationsanalysen zwischen EMA-basierten Soft-Labels und den in Laborstudien erhobenen Cortisolwerten durch (siehe Abb. 21, 22). Diese Ergebnisse wurden allen Konsortialpartnern mitgeteilt. Anschließend führte die PLG gemeinsame Evaluierungstreffen mit Psychiater:innen und klinischen Expert:innen durch, um die Implikationen zu interpretieren. Basierend auf diesem Expertenfeedback wählte das Team das am besten geeignete, von der EMA abgeleitete Kennzeichnungsschema aus, das in die MuStP-Pipeline zur Vorhersage realer Stressbedingungen integriert werden sollte. Dadurch stellte die PLG sicher, dass das finale App-System sowohl auf physiologische Variabilität kalibriert als auch klinisch in seiner Stressinterpretation ausgerichtet war und einen präzisen und ethisch verantwortungsvollen Einsatz im Alltag ermöglichte.

AP 4.2 (Monat 30-36; verantwortlich: UKJ)

(ursprüngliches) Ziel: ELSI Organisation

UKJ: AP 4.2 wurde entsprechend der angepassten Studiendurchführung erfüllt.

AP 4.3 (Monat 30-36; verantwortlich: UKJ, Kollaboration: Thryve, Fraunhofer IGD)

(ursprüngliches) Ziel: Umfangreicher Feldtest, Vergleich der Ergebnisse von Laborexperimenten und der realen Szenarien, Optimierung der Benutzeroberfläche, Optimierung des maschinellen Lernalgorithmus

UKJ: Die Messungen dauern z.T. noch an. Die Feldstudie in Tübingen ist abgeschlossen. Die Feldstudie in Jena wird im August '25 abgeschlossen.

Thryve: Aufgrund der Veränderungen im Ablauf des umfangreichen Feldtests wurden von Thryve neue QR-Codes für das QR-Code-Onboarding der Studienteilnehmer:innen erstellt und Änderungen der EMA-Fragebögen implementiert. Zusätzlich mussten aufgrund der verlängerten Studiendauer die Zugriffsrechte auf den Datenexport-Server angepasst und für neue Studienpartner erteilt werden.

AP 4.4 (Monat 13-36; verantwortlich: UKJ)

(ursprüngliches) Ziel: Dissemination der Ergebnisse

UKJ: Die Dissemination der Ergebnisse über Konferenzen, Manuskripte und sonstige Veröffentlichungen läuft (vgl. auch "Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF"). Um die Entscheidungen des Algorithmus zu kalibrieren, schlug die PLG zwei Hauptansätze vor. Sie nutzen den Ähnlichkeitsabgleich (SM) oder wählten den am besten geeigneten Anomalie-Detektor für jede:n Testnutzer:in während der Inferenz aus, anstatt auf einen einzigen Anomalie-Detektor zu setzen, der auf dem gesamten Labordatensatz trainiert wurde. Bei den Deep-Learning-Algorithmen, die in der vorgeschlagenen Pipeline verwendet wurden, wurde eine Feinabstimmung (FT) angewandt, um die letzten linearen Schichten des Modells zu aktualisieren, um die Personalisierung zu verbessern und Domänenverschiebungen zu reduzieren, da das Modell zunächst auf Labordaten trainiert wird, bevor es auf reale Daten angewendet wird. Schließlich wurde eine Post-hoc-Optimierung (PO) der Entscheidungsschwellen unter Verwendung von 40 % der im realen Leben gesammelten Nutzerdaten vorgenommen. Durch die Implementierung von SM, FT und PO verbesserten PLG die Leistung ihres Algorithmus erheblich und erreichten einen F1-Wert von 0,52, obwohl nur 31 % des Datensatzes aus Stressetiketten bestanden. Die PLG reichte das entsprechende Forschungspapier beim ML for Life and Material Science: From Theory to Industry Applications Workshop auf der International Conference on Machine Learning 2024 ein (s.u.). Das UKJ arbeitet derzeit mit Kooperationspartnern in Rostock und Saarbrücken zusammen, um ein Paper über das neuartige 3-stufige Stresserkennungssystem und die Bewertung der verbesserten Modelle und Genauigkeiten anhand von Labor- und Felddaten aus dem Cello-Projekt zu veröffentlichen. Es wird auch an einem Artikel über die Stabilität der HRV-basierten Merkmale beim Übergang von Labor- zu Felddaten gearbeitet. Zusätzlich arbeitet das UKJ an Veröffentlichungen zu dynamischer Konnektivität und nutzen im Labor gemessene physiologische Gehirn- und Herzdaten, um subjektive Selbsteinschätzungen und Cortisolwerte vorherzusagen (Meilicke,* Golbabaee,* Krylova, Winter, Izyurov, Colic; in Vorbereitung).

Folgende Medizinische Doktor- und Masterarbeiten werden/wurden im Rahmen des Projektes bearbeitet:

- Medizinische Doktorarbeit über hormonelle Daten (Cortisol) – läuft noch
- Medizinische Doktorarbeit über EMA-Ratings – läuft noch
- Medizinische Doktorarbeit mit EKG und Fragebögen - läuft noch, MD-Doktorandin N. Winter; Abschluss vorgesehen Ende 2026
- Medizinische Doktorarbeit unter Verwendung von Hormondaten (Cortisol) und Fragebögen - läuft noch, MD-Doktorandin L. McClain; Abschluss vorgesehen Ende 2025
- Medizinische Doktorarbeit mit MRS und Fragebögen - läuft noch, MD-Doktorand H. Meilicke; Abschluss vorgesehen Ende 2025
- Medizinische Doktorarbeit über Verhaltensdaten mithilfe von Fragebögen – läuft noch; MD-Doktorandin F. Bestehorn; Abschluss vorgesehen Mitte 2026
- Medizinische Doktorarbeit über Verhaltensdaten mithilfe von Fragebögen - läuft noch; MD-Doktorand B. Buske; Abschluss vorgesehen Mitte 2026
- in Tübingen läuft noch eine Masterarbeit über Verhaltensdaten mithilfe von Fragebögen und eine Medizinische Doktorarbeit über Verhaltens- und physiologische Daten mittels Fragebögen, EMA und EKG EMA

1. wichtigste Positionen des zahlenmäßigen Nachweises

Für die Erreichung der oben genannten Ziele wurden in allen Teilprojekten folgende Ausgaben getätigt (vgl. Verwendungsnachweise):

1. Personalmittel,
2. Reisemittel
3. und Sachmittel.

2. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Stress ist Deutschland- und EU-weit das zweithäufigste arbeitsbedingte Gesundheitsproblem. Wissenschaftliche Belege für die konkrete Wirksamkeit von Präventionsprogrammen sind unumgänglich, die Charakterisierung neuronaler, physiologischer und verhaltensbezogener Signaturen von Stress, um Präventionsprogramme zu konzipieren und zu verbessern, stellt die Basis für derartige Programme zur Stressreduktion und auch Stressmanagement dar. Die gemeinsame Zielstellung des Projektes bestand daher darin, einen Demonstrator für eine softwarebasierte Gesundheitsanwendung über Wearables und Smartphones zu entwickeln, denn Smartphone-basierte Apps bergen ein großes Potenzial, Menschen in ihrer individuellen Stressbewältigung zu unterstützen und die Ursachenfindung von Stress und auch von Resilienz zu ermöglichen. Dieses projektspezifische Ziel bediente den übergreifenden Zweck, die Mensch-Technik-Interaktion zu fördern und KI-Methoden zu nutzen, um Menschen bei Problemlösungen optimal zu assistieren und ihnen neue Möglichkeiten und Werkzeuge zur Stärkung eigener Fähigkeiten und Kompetenzen an die Hand zu geben. Smartphones insbesondere im Zusammenhang mit Smartwatches stehen jederzeit zur Verfügung und können in allen Situationen getragen und genutzt werden. Auf diese Weise kann die Charakterisierung von Stress vorangetrieben und Auswirkungen von Interventionen, wie z.B. Biofeedback-Training in verschiedenen Altersgruppen und mit unterschiedlichen Lebensstilen erfasst und analysiert werden.

Stresserkennungssysteme wiesen jedoch zum Zeitpunkt der Antragstellung für eine derartige Nutzung noch zu niedrige Präzision auf und für die Mehrheit der verfügbaren Apps gab es noch keine systematische Wirksamkeitsanalyse. Systeme oder Modelle, die sich auf Gesundheitsprognosen von Stress beziehen, gab es bis dahin ebenfalls nicht. Es sollten Methoden und Technologien des maschinellen Lernens bzw. der künstlichen Intelligenz entwickelt und erprobt werden, mit denen die gesammelten Vitalparameter der Nutzer:innen z.B. auf Anomalien bzw. eine akute Belastungssituation analysiert werden können. Die ursprünglich geplanten notwendigen und letztlich durchgeführten Arbeitsschritte waren notwendig, um das Projekt und die übergreifenden Ziele des BMBF voranzubringen. Die Forschungs- und Entwicklungsarbeiten des Projektes setzten die Weiterentwicklung und Optimierung von Augmented Intelligence entsprechen der Förderrichtlinien zielorientiert um.

3. voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses - auch konkrete Planungen für die nähere Zukunft - im Sinne des fortgeschriebenen Verwertungsplans

(siehe auch III - „Erfolgskontrollbericht“). Das Projekt sollte das Verständnis der Mechanismen, die der Stressaktivität zugrunde liegen, erheblich verbessern. Die im Rahmen dieses Projekts aufgezeichneten multimodalen Daten (App-Daten, neuronale Daten, Verhaltensdaten, physiologische Daten), die teilweise über lange Testperioden und bei einer großen Anzahl von Proband:innen gesammelt wurden, sollten es ermöglichen, auf persönlicher Ebene die Gründe für Stressreaktionen und die Beziehung zwischen Neuroimaging-Daten und den Daten, die wir vom tragbaren App-System erhalten haben (Ziel des Projekts Cello), zu erklären. Eine starke und klare Verbindung zwischen diesen Modalitäten zu finden, sollte dazu dienen, zu erklären, wie sich die Proband:innen fühlen und welche Ursachen diese Gefühle haben können. Die Projektpartner werden auf Basis der Projektergebnisse und -erkenntnisse zukünftig weitere Forschung betreiben. Darüber hinaus ergeben die Daten, die von den App-Anwender:innen gesammelt wurden, einen großen Datensatz mit hohem Forschungspotenzial, der in Zukunft zur Beantwortung gültiger

wissenschaftlicher Fragen verwendet werden kann. Im kommerziellen Bereich wird das Cello-Stressmanagementsystem plattformunabhängig sein und den vorhandenen Stressmanagementsystemen dank seiner Echtzeit-Stressmessungen, personalisierten Interventionen und Prognosen überlegen sein. Dies kann den Stressabbau optimieren, die Kosten stressbedingter Krankheiten senken und gesünderes Leben ermöglichen.

Projektergebnis ist ein Demonstrator eines App-basierten Stressmanagementsystems zur Erkennung und Vermeidung individueller Stressfaktoren und zur Vorhersage der eigenen Gesundheitsentwicklung und des Wohlbefindens. Das System könnte es den Nutzern nicht nur ermöglichen, ihr Verhalten zu ändern, um das Wohlbefinden zu verbessern, sondern es könnte auch dazu dienen, Frühwarnsymptome von stressassoziierten psychischen Erkrankungen zu erkennen. Um eine marktfähige Innovation zu entwickeln, wird das Forschungsteam mit Versicherungsgesellschaften und Gesundheitseinrichtungen kooperieren, um das System zu fördern und die Nutzer:innen zu unterstützen und die Kosten durch stressbedingte Erkrankungen auf ein Minimum zu reduzieren. Die Genauigkeit, Zuverlässigkeit und Wirksamkeit des App-Systems und dessen alltäglicher Einsatz wird derzeit noch getestet. Die Markteinführung wird in der ersten Stufe für den App Store (Apple) und Play Store (Android) geplant und sieht eine Verfügbarkeit der App in vorerst deutscher Sprache vor.

4. während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Aus Sicht verfügbarer Lösungen für Stress-Tracking in Smartwatches nutzen/nutzen Unternehmen wie Garmin, Samsung, Mi Band, FitBit, Withings oder die für Apple Watch verfügbaren Apps hauptsächlich EKG-, PPG-Herzfrequenz- und EDA-Kanäle, körperliche Aktivität und Daten zur Schlafqualität, um den Stressindex der Nutzer:innen vorherzusagen. Der Hauptaspekt, der bei diesen Algorithmen fehlte, ist die Verwendung subjektiver Daten (EMAs), um das Stressniveau der Benutzer:innen persönlich einzuschätzen. Da sie jedoch von privaten Unternehmen entwickelt werden, sind diese Algorithmen nicht öffentlich und werden für gerätespezifische Zwecke entwickelt. Es ist wichtig hervorzuheben, dass das Cell-o -Projekt darauf abzielte, einen Stressmonitorings-/Erkennungsalgorithmus zu entwickeln, der in einer geräteunabhängigen Umgebung verwendet werden konnte. Entsprechende Entwicklungen sowie weitere relevante FE-Ergebnisse oder Fortschritte von dritter Seite sind diesbezüglich nicht bekannt.

5. erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF

- Interview im Sept. 2022 zum Welttag der psychischen Gesundheit: „Betroffene in Therapie miteinbeziehen und ihre Autonomie stärken“ ([Interview zum Welttag der psychischen Gesundheit: „Betroffene in Therapie miteinbeziehen und ihre Autonomie stärken“ — Miteinander durch Innovation](#))
- Im Rahmen der NØRD Messe 2024 erfolgte am 30.05.2024 ein Fachvortrag für das Messepublikum durch den Wissenschaftlichen Mitarbeiter Erik Endlicher (Fraunhofer IGD) mit dem Thema „Berührungslose Vitaldatenanalyse für medizinische und berufliche Einsatzgebiete“.
- ML4LMS Poster im Juni 2024: „From Laboratory to Everyday Life: Personalized Stress Prediction via Smartwatches“ Accepted at the 1st Machine Learning for Life and Material Sciences, Workshop at ICML 2024. Copyright 2024 by the author(s). <https://openreview.net/pdf?id=XBtTaeQOWs>
- Publikationen noch in Vorbereitung: neue BA-Arbeit EMA-Bewertungen und Fragebogen (K. Bösendorfer) in 2025
- Meilicke, H. M. X.,* Golbabaie, S.,* McClain, L., Winter, N., Chand, T., Herrmann, L., ... & Colic, L. (in preparation). Brain activity and connectivity during a psychosocial stress task: effects of neurometabolites, cortisol, childhood adversity and recent chronic stress.

- Doerner, M.,* Bernal, J.,* Mattern, H., Golbabaei, S., Winter, N., Herrmann, L., ...& Engert, V., Schriber, S., Colic, L. (in preparation). Perivascular spaces as a link among vascular function and psychosocial stress: a two-sample study.
- geplantes Paper (gemeinsam mit N. Opel, V. Engert zu TSST + Feldstudie I) in 2025
- Symposium zur DGPPN Tagung (Berlin 2024)
- zu noch laufenden Master- oder Doktorarbeiten siehe oben (AP 4.4)

Weitere Quellen/Zitationen:

1. Arsalan, A., Anwar, S. M., & Majid, M. (2022). Mental Stress Detection using Data from Wearable and Non-wearable Sensors: A Review. *arXiv preprint arXiv:2202.03033*.
2. Behinaein, B., Bhatti, A., Rodenburg, D., Hungler, P., & Etemad, A. (2021). A Transformer Architecture for Stress Detection from ECG. In *2021 International Symposium on Wearable Computers* (pp. 132-134).
3. Bica, I., Alaa, A., & Van Der Schaar, M. (2020, November). Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International Conference on Machine Learning* (pp. 884-895). PMLR.
4. Boyd, A., Bamler, R., Mandt, S., & Smyth, P. (2020). User-dependent neural sequence models for continuous-time event data. *Advances in Neural Information Processing Systems*, 33, 21488-21499.
5. Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics*, 92, 103139.
6. Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). Do we really need deep learning models for time series forecasting?. *arXiv preprint arXiv:2101.02118*.
7. Fan, X., Zhang, S., Chen, B., & Zhou, M. (2020). Bayesian attention modules. *Advances in Neural Information Processing Systems*, 33, 16362-16376.
8. Franceschi, J. Y., Dieuleveut, A., & Jaggi, M. (2019). Unsupervised scalable representation learning for multivariate time series. *Advances in neural information processing systems*, 32.
9. Gedam, S., & Paul, S. (2021). A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. *IEEE Access*, vol. 9, pp. 84045-84066
10. Grigsby, J., Wang, Z., & Qi, Y. (2021). Long-Range Transformers for Dynamic Spatiotemporal Forecasting. *arXiv preprint arXiv:2109.12218*.
11. Li, L., Zhang, Y., & Chen, L. (2021). Personalized transformer for explainable recommendation. *arXiv preprint arXiv:2105.11601*.
12. Li, S., Xiao, S., Zhu, S., Du, N., Xie, Y., & Song, L. (2018). Learning temporal point processes via reinforcement learning. *Advances in neural information processing systems*, 31.
13. Moreno-Pino, F., Olmos, P. M., & Artés-Rodríguez, A. (2021). Deep Autoregressive Models with Spectral Attention. *arXiv preprint arXiv:2107.05984*.
14. Salehi, F., Trouleau, W., Grossglauser, M., & Thiran, P. (2019). Learning Hawkes processes from a handful of events. *Advances in Neural Information Processing Systems*, 32.
15. Shchur, O., Türkmen, A. C., Januschowski, T., & Günnemann, S. (2021). Neural temporal point processes: A review. *arXiv preprint arXiv:2104.03528*.
16. Siirtola, P. (2019, September). 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* (pp. 1198-1201).

17. Uadhyay, U., De, A., & Gomez Rodriguez, M. (2018). Deep reinforcement learning of marked temporal point processes. *Advances in Neural Information Processing Systems*, 31.
18. Yazdanbakhsh, O., & Dick, S. (2019). Multivariate time series classification using dilated convolutional neural network. *arXiv preprint arXiv:1905.01697*.
19. Yèche, H., Dresdner, G., Locatello, F., Hüser, M., & Rätsch, G. (2021, July). Neighborhood contrastive learning applied to online patient monitoring. In *International Conference on Machine Learning* (pp. 11964-11974). PMLR.
20. Yu, C., Liu, J., Nemati, S., & Yin, G. (2021). Reinforcement learning in healthcare: A survey. *ACM Computing Surveys (CSUR)*, 55(1), 1-36.
21. Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. (2021). TS2Vec: Towards Universal Representation of Time Series. *arXiv preprint arXiv:2106.10466*.
22. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021, August). A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 2114-2124).
23. Zhang, G., Ashrafi, R. A., Juuti, A., Pietiläinen, K., & Marttinen, P. (2020). Errors-in-variables modeling of personalized treatment-response trajectories. *IEEE Journal of Biomedical and Health Informatics*, 25(1), 201-208.
24. Zhang, Q., Lipani, A., Kirnap, O., & Yilmaz, E. (2020, November). Self-attentive Hawkes process. In *International conference on machine learning* (pp. 11183-11193). PMLR.
25. Zuo, S., Jiang, H., Li, Z., Zhao, T., & Zha, H. (2020, November). Transformer hawkes process. In *International conference on machine learning* (pp. 11692-11702). PMLR.