

Die nächste Generation hochleistungsfähiger Edge AI für intelligente Gebäude (OctoAI)

Berichte aus Energie- und Umweltforschung 57/2025

Wien, 2025

Impressum

Medieninhaber, Verleger und Herausgeber:

Bundesministerium für Innovation, Mobilität und Infrastruktur,
Radetzkystraße 2, 1030 Wien

Verantwortung und Koordination: Abteilung III/3 - Energie und Umwelttechnologien

Leitung: DI (FH) Volker Schaffler, MA, AKKM

Autorinnen und Autoren:

Franz Wotawa (TU Graz, Institut für Softwaretechnologie)

Theresa Kohl, Christoph Siegl, Thomas Hirsch, Thomas Schwengler, Gerald Schweiger (DiLT Analytics GmbH)

Matej Gustin, Christina Hopfe (TU Graz, Institut für Bauphysik, Gebäudetechnik und Hochbau)

Wien, Graz 2025. Stand: Jänner 2025

Ein Projektbericht gefördert im Rahmen von



Rückmeldungen:

Ihre Überlegungen zu vorliegender Publikation übermitteln Sie bitte an iii3@bmimi.gv.at.

Rechtlicher Hinweis

Dieser Ergebnisbericht wurde von die/der Projektnehmer:in erstellt. Für die Richtigkeit, Vollständigkeit, Aktualität sowie die barrierefreie Gestaltung der Inhalte übernimmt das Bundesministerium für Innovation, Mobilität und Infrastruktur (BMIMI) keine Haftung.

Mit der Übermittlung der Projektbeschreibung bestätigt die/der Projektnehmer:in ausdrücklich, über sämtliche für die Nutzung erforderlichen Rechte – insbesondere Urheberrechte, Leistungsschutzrechte sowie etwaige Persönlichkeitsrechte abgebildeter Personen – am bereitgestellten Bildmaterial zu verfügen.

Die/der Projektnehmer:in räumt dem BMIMI ein unentgeltliches, nicht ausschließliches, zeitlich und örtlich unbeschränktes sowie unwiderrufliches Nutzungsrecht ein, das übermittelte Bildmaterial in allen derzeit bekannten sowie künftig bekannt werdenden Nutzungsarten für Zwecke der Berichterstattung, Dokumentation und Öffentlichkeitsarbeit im Zusammenhang mit der geförderten Maßnahme zu verwenden, insbesondere zur Veröffentlichung in Printmedien, digitalen Medien, Präsentationen und sozialen Netzwerken.

Für den Fall, dass Dritte Ansprüche wegen einer Verletzung von Rechten am übermittelten Bildmaterial gegen das BMIMI geltend machen, verpflichtet sich die/der Projektnehmer:in, das BMIMI vollständig schad- und klaglos zu halten. Dies umfasst insbesondere auch die Kosten einer angemessenen rechtlichen Vertretung sowie etwaige gerichtliche und außergerichtliche Aufwendungen.

Vorbemerkung

Der vorliegende Bericht dokumentiert die Ergebnisse eines Projekts aus dem Forschungs- und Technologieprogramm „Stadt der Zukunft“ des Bundesministeriums für Innovation, Mobilität und Infrastruktur (BMIMI). Dieses Programm baut auf dem langjährigen Programm „Haus der Zukunft“ auf und hat die Intention, Konzepte, Technologien und Lösungen für zukünftige Städte und Stadtquartiere zu entwickeln und bei der Umsetzung zu unterstützen. Damit soll eine Entwicklung in Richtung energieeffiziente und klimaverträgliche Stadt unterstützt werden, die auch dazu beiträgt, die Lebensqualität und die wirtschaftliche Standortattraktivität zu erhöhen. Eine integrierte Planung wie auch die Berücksichtigung aller betroffener Bereiche wie Energieerzeugung und -verteilung, gebaute Infrastruktur, Mobilität und Kommunikation sind dabei Voraussetzung.

Um die Wirkung des Programms zu erhöhen, sind die Sichtbarkeit und leichte Verfügbarkeit der innovativen Ergebnisse ein wichtiges Anliegen. Daher werden nach dem Open Access Prinzip möglichst alle Projektergebnisse des Programms in der Schriftenreihe des BMIMI publiziert und elektronisch über die Plattform www.NachhaltigWirtschaften.at zugänglich gemacht. In diesem Sinne wünschen wir allen Interessierten und Anwender:innen eine interessante Lektüre.

Inhaltsverzeichnis

1	Kurzfassung	7
2	Abstract	10
3	Ausgangslage	12
	3.1. Ausgangslage (Status Quo, Motivation)	12
	3.2. Vorprojekte, Stand der Technik und Zielsetzungen des Projekts	12
	3.2.1. Vorprojekte.....	12
	3.2.2. Zusammenfassung des nationalen und internationalen Wissensstands	13
	3.2.3. Zielsetzungen und Innovationsgehalt von OctoAI.....	16
4	Projekthalt	17
	4.1. Anforderungsanalyse	17
	4.2. Entwicklung von Modellen und Technologien.....	19
	4.2.1. Belegungsdetektion	19
	4.2.2. Komfortvorhersage.....	20
	4.3. Prototypische Implementierung, verwendete Datensätze und Erhebung	22
	4.4. Validierung und Deployment	25
5	Ergebnisse	27
	5.1. Use Case 1 - Belegungserkennung	27
	5.1.1. Experimente zur Verringerung der Modellgrößen	31
	5.1.2. Transfer Learning.....	32
	5.2. Use Case 2 - Komfortbewertung- und Vorhersage	33
	5.3. Nutzerfreundliche Dashboards und Usability Tests.....	35
6	Schlussfolgerungen	37
	6.1. Belegungserkennung in Büros	37
	6.2. Komfortzonen	38
	6.3. Zielgruppen und Relevanz der Projektergebnisse	38
	6.4. Bisherige Verwertungs- und Verbreitungsaktivitäten	39
	6.5. Marktpotenzial.....	41
7	Ausblick und Empfehlungen	43
8	Verzeichnisse	45

1 Kurzfassung

Motivation und Forschungsfrage

Der Gebäudebestand in der Europäischen Union ist für etwa 40 % des Endenergieverbrauchs und 36 % der CO₂-Emissionen verantwortlich. Die zunehmende Urbanisierung, steigende Energiekosten und der wachsende Bedarf an energieeffizienten Lösungen erfordern innovative Ansätze zur Reduktion des Energieverbrauchs und zur Integration erneuerbarer Energien. Gleichzeitig bergen bestehende Cloud-basierte Systeme Herausforderungen hinsichtlich Datenschutz, Latenzzeiten und Verfügbarkeit.

Das Forschungsprojekt OctoAI untersucht die Anwendung von Edge-KI-Technologien für intelligente Gebäude, um datengestützte Entscheidungen effizient und sicher zu ermöglichen. Ziel ist es, alternative Rechenmodelle zu entwickeln, die sowohl die Energieeffizienz als auch den Komfort in Gebäuden verbessern.

Ausgangssituation/Status Quo

Viele aktuelle Smart-Building-Anwendungen basieren auf Cloud Computing, was zentrale Herausforderungen mit sich bringt: hohe Latenzzeiten, Abhängigkeit von einer stabilen Internetverbindung sowie Datenschutz- und Sicherheitsrisiken. Edge Computing bietet eine Alternative, bei der KI-Modelle direkt auf Endgeräten verarbeitet werden. Dadurch werden Reaktionszeiten verkürzt, die Datenhoheit bleibt bei den Nutzer:innen, und die Verlässlichkeit steigt. Trotz dieser Vorteile ist der Einsatz von Edge KI in Gebäuden noch nicht weit verbreitet, und es fehlen praxisnahe Anwendungen mit validierten Anwendungsfällen.

Projekthalte und Zielsetzungen

Das Projekt OctoAI konzentriert sich auf die Entwicklung und Erprobung von Edge-KI-Technologien in intelligenten Gebäuden. Dabei wurden zwei zentrale Anwendungsfälle identifiziert, die durch innovative maschinelle Lernverfahren optimiert werden sollen.

Der erste Anwendungsfall ist die Belegungsdetektion. Hierbei wurde untersucht, wie Sensordaten genutzt werden können, um die Anwesenheit von Personen in Innenräumen möglichst genau zu bestimmen. Die erfassten Daten, darunter CO₂-Konzentration, Geräuschpegel, Lichtintensität und Türöffnungsstatus, wurden mit verschiedenen KI-Methoden verarbeitet, um eine präzise Klassifikation der Raumauslastung zu ermöglichen. Das Hauptziel war die Entwicklung ressourcenschonender Modelle, die auf Edge-Geräten effizient laufen können, ohne auf eine Cloud-Anbindung angewiesen zu sein. Neben der Verbesserung der Gebäudeautomation durch intelligente Heiz- und Lüftungssysteme lag der Fokus auch auf der Minimierung des Energieverbrauchs.

Der zweite Anwendungsfall befasst sich mit der Vorhersage des thermischen Komforts. Hierbei wurde ein Modell entwickelt, das die Raumtemperatur für die kommenden 24 Stunden mit hoher Genauigkeit prognostizieren kann. Durch die Einbindung von Wetterprognosen und Echtzeit-Sensordaten konnte eine adaptive Regelung der Klimatisierung erreicht werden. Das Modell basiert auf der EN 16798-1-Norm für Komfortzonen und bietet eine datengetriebene Entscheidungsgrundlage für eine effiziente Steuerung der Gebäudeheizung und -kühlung.

Die Zielsetzung des Projekts bestand darin, robuste und skalierbare Edge-KI-Modelle zu entwickeln, die sowohl in neuen als auch in bestehenden Gebäuden implementiert werden können. Die Optimierung der Rechenleistung, die Sicherstellung einer hohen Datenqualität und die praktische Validierung der entwickelten Modelle in realen Testumgebungen standen dabei im Mittelpunkt. Zusätzlich sollte das Projekt zur allgemeinen Akzeptanz und Verbreitung von Edge-KI-Technologien beitragen, indem die Vorteile dieser Technologie für den Gebäudebetrieb greifbar gemacht wurden.

Methodische Vorgehensweise

Das Projekt OctoAI folgte einem iterativen und datengetriebenen Ansatz, der die Entwicklung, Implementierung und Validierung von KI-gestützten Modellen in verschiedenen realen Testumgebungen umfasste. Zunächst wurden durch eine umfassende Anforderungsanalyse die wichtigsten Bedürfnisse und Herausforderungen in der Gebäudeautomation identifiziert. Hierbei kamen Experteninterviews und eine groß angelegte Stakeholder-Befragung zum Einsatz, die wertvolle Erkenntnisse über den Status quo und die Erwartungen der Praxisakteure lieferten.

Nach der Spezifikation der technischen Anforderungen erfolgte die Erhebung und Verarbeitung von Sensordaten in Pilotprojekten. In zwei unterschiedlichen Bürogebäuden wurden Sensoren zur Messung von Luftqualität, Geräuschpegel, Lichtintensität und Raumtemperatur installiert. Diese Datensätze dienten als Grundlage für das Training der KI-Modelle. Zur Belegungsdetektion wurden verschiedene maschinelle Lernverfahren getestet und miteinander verglichen, darunter neuronale Netzwerke, Entscheidungsbäume und klassische statistische Verfahren. Besondere Aufmerksamkeit galt der Evaluierung der Modellgenauigkeit unter realen Bedingungen und der Optimierung der Algorithmen für den Einsatz auf ressourcenbeschränkten Edge-Geräten.

Parallel dazu wurde das Modell zur thermischen Komfortprognose entwickelt, das auf autoregressiven Verfahren und externen Einflussfaktoren wie Wetterdaten basiert. Die Modelle wurden in Echtzeit getestet und ihre Vorhersagegenauigkeit durch eine rollierende Validierung überprüft. Das Ziel war es, eine zuverlässige Methode zu schaffen, die sich dynamisch an veränderte Umweltbedingungen anpasst und praxisnahe Empfehlungen für Gebäudemanager liefert.

Ein wesentlicher Bestandteil des Projekts war zudem die Entwicklung eines interaktiven Dashboards, das die Sensordaten visualisiert und den Nutzer:innen ermöglicht, auf Basis der Modellprognosen Entscheidungen zu treffen. Usability-Tests mit verschiedenen Zielgruppen trugen dazu bei, die Benutzerfreundlichkeit des Dashboards zu optimieren und die praktische Anwendbarkeit der entwickelten Lösungen zu verbessern.

Ergebnisse und Schlussfolgerungen

Die entwickelten Modelle zur Belegungsdetektion erreichten eine Genauigkeit von bis zu 87 %, was eine präzise Steuerung von Heizungs- und Lüftungssystemen ermöglicht. Die prädiktiven Modelle zur thermischen Komfortbewertung konnten Temperaturverläufe mit einem mittleren Fehler (RMSE) von 0,158°C vorhersagen und bieten damit eine solide Grundlage für zukünftige Regelungsstrategien. Zusätzlich zeigte sich, dass Edge-KI für intelligente Gebäude geeignet ist und selbst mit leistungsschwacher Hardware zuverlässige Ergebnisse liefert. Die Nutzer:innen profitierten von reduzierten Latenzzeiten und höherer Datensicherheit.

Ausblick

Die Ergebnisse des OctoAI-Projekts legen das Fundament für weiterführende Entwicklungen im Bereich Edge-KI für Gebäude. Künftige Forschungen sollten sich mit der Integration in bestehende Smart-Building-Systeme, der Erweiterung der Sensorik sowie der Skalierung der Technologie für größere Gebäudekomplexe beschäftigen. Darüber hinaus bietet die Kombination von Edge-KI mit dezentralen Energieversorgungssystemen vielversprechende Perspektiven für energieeffiziente und resiliente Gebäude der Zukunft.

2 Abstract

Motivation and Research Question

The building stock in the European Union accounts for approximately 40% of final energy consumption and 36% of CO₂ emissions. Increasing urbanization, rising energy costs, and the growing demand for energy-efficient solutions require innovative approaches to reducing energy consumption and integrating renewable energy sources. At the same time, existing cloud-based systems pose challenges regarding data protection, latency, and availability.

The **OctoAI** research project explores the application of edge AI technologies for intelligent buildings to enable data-driven decisions efficiently and securely. The goal is to develop alternative computing models that improve both energy efficiency and comfort in buildings.

Initial Situation/Status Quo

Many current smart building applications rely on cloud computing, which presents significant challenges: high latency, dependence on a stable internet connection, and data protection and security risks. Edge computing offers an alternative where AI models are processed directly on end devices. This reduces response times, ensures data sovereignty for users, and increases reliability. Despite these advantages, edge AI adoption in buildings is still limited, and there is a lack of practical applications with validated use cases.

Project Content and Objectives

The OctoAI project focuses on the development and testing of edge AI technologies in intelligent buildings. Two key use cases have been identified that will be optimized using innovative machine learning techniques.

The first use case is occupancy detection. This examines how sensor data can be used to determine the presence of people in indoor spaces as accurately as possible. Data collected, including CO₂ concentration, noise levels, light intensity, and door opening status, were processed using various AI methods to enable precise room occupancy classification. The primary goal was to develop resource-efficient models that can run efficiently on edge devices without requiring cloud connectivity. In addition to improving building automation through intelligent heating and ventilation systems, the focus was also on minimizing energy consumption.

The second use case involves thermal comfort prediction. A model was developed to predict room temperature for the next 24 hours with high accuracy. By integrating weather forecasts and real-time sensor data, an adaptive climate control system was achieved. The model is based on the EN 16798-1 standard for comfort zones and provides a data-driven decision basis for efficient building heating and cooling management.

The project aimed to develop robust and scalable edge AI models that can be implemented in both new and existing buildings. Optimizing computing power, ensuring high data quality, and practically validating the developed models in real test environments were central to the project. Additionally,

the project sought to contribute to the general acceptance and dissemination of edge AI technologies by making the benefits of this technology tangible for building operations.

Methodical Approach

The OctoAI project followed an iterative and data-driven approach that encompassed the development, implementation, and validation of AI-based models in various real test environments. Initially, a comprehensive requirements analysis was conducted to identify the key needs and challenges in building automation. Expert interviews and a large-scale stakeholder survey provided valuable insights into the status quo and the expectations of industry participants.

After defining the technical requirements, sensor data were collected and processed in pilot projects. Sensors for measuring air quality, noise levels, light intensity, and room temperature were installed in two different office buildings. These datasets formed the basis for training the AI models. Various machine learning techniques were tested and compared for occupancy detection, including neural networks, decision trees, and classical statistical methods. Special attention was given to evaluating model accuracy under real conditions and optimizing the algorithms for use on resource-constrained edge devices.

At the same time, a model for thermal comfort prediction was developed based on autoregressive techniques and external influencing factors such as weather data. The models were tested in real-time, and their prediction accuracy was checked through rolling validation. The goal was to create a reliable method that dynamically adapts to changing environmental conditions and provides practical recommendations for building managers.

A key component of the project was also the development of an interactive dashboard that visualizes sensor data and enables users to make informed decisions based on model predictions. Usability tests with various target groups helped optimize the dashboard's user-friendliness and improve the practical applicability of the developed solutions.

Results and Conclusions

The developed occupancy detection models achieved an accuracy of up to 87%, allowing precise control of heating and ventilation systems. The predictive models for thermal comfort assessment were able to forecast temperature trends with a mean error (RMSE) of 0.158°C, providing a solid foundation for future control strategies.

Additionally, it was demonstrated that edge AI is suitable for intelligent buildings and can deliver reliable results even with low-performance hardware. Users benefited from reduced latency times and higher data security.

Outlook

The results of the OctoAI project lay the foundation for further developments in the field of edge AI for buildings. Future research should focus on integrating the technology into existing smart building systems, expanding sensor networks, and scaling the technology for larger building complexes. Furthermore, combining edge AI with decentralized energy supply systems offers promising perspectives for energy-efficient and resilient buildings of the future.

3 Ausgangslage

3.1. Ausgangslage (Status Quo, Motivation)

Der Gebäudebestand in der Europäischen Union ist nach wie vor energieintensiv und überwiegend ineffizient. Mit einem Anteil von 40 % am Endenergieverbrauch und 36 % der CO₂-Emissionen leisten Gebäude einen erheblichen Beitrag zu den Klimabelastungen. Zur Reduzierung des Energieverbrauchs und zur effektiven Integration erneuerbarer Energien sind innovative Systeme erforderlich. Diese müssen durch fortschrittliche Technologien wie Cyber-Physical Systems und neuartige Energy Services unterstützt werden.

Mit der steigenden Verfügbarkeit von großen Datenmengen wird die Anwendung von datengetriebenen Analysen, basierend auf diversen Algorithmen der Künstlichen Intelligenz (KI), zunehmend attraktiver. KI bietet das Potenzial, als Schlüsseltechnologie neue Energy Services wie modellprädiktive Regelung und Demand-Side-Management zu ermöglichen. Diese Anwendungen können den Energieverbrauch in Gebäuden signifikant senken und damit zur Nachhaltigkeit beitragen.

Aktuell wird die Mehrzahl der Internet-of-Things-Anwendungen (IoT) bzw. KI-Analysen durch Cloud-Computing-Technologien umgesetzt. Obwohl diese IoT-Anwendungen kosteneffizient und vielseitig einsetzbar sind, bestehen gravierende Schwachstellen in den Bereichen Verlässlichkeit, Vertrauenswürdigkeit und Datenschutz.

Eine vielversprechende Alternative zu Cloud-Lösungen ist Edge Computing. Durch die lokale Verarbeitung von Daten in Endgeräten werden Latenzzeiten verkürzt, die Verfügbarkeit verbessert und der Datenschutz erhöht. Gleichzeitig ermöglicht Edge Computing den Nutzerinnen und Nutzern eine direkte Kontrolle über ihre Daten, was insbesondere im Bereich intelligenter Gebäude von zentraler Bedeutung ist.

3.2. Vorprojekte, Stand der Technik und Zielsetzungen des Projekts

3.2.1. Vorprojekte

Das Projekt „OctoAI“ baut auf einer soliden Grundlage und Vorerkenntnissen aus Vorprojekten des Konsortiums auf, die entscheidende Beiträge zur Entwicklung von Edge-KI-Technologien für Gebäude und Nutzer:innenintegration geliefert haben. Projekte wie „GameOpSys“ (FFG-Förderung), „Smart2B“ (Horizon 2020 – Förderungsvertrag Nr. 101023666) und „ANSERS“ (Zukunftsfonds Steiermark Förderung, PN 1335) haben zahlreiche neue Forschungsfragen aufgezeigt, die in OctoAI adressiert wurden.

GameOpSys (Stadt der Zukunft, 2019-2022)

Das zentrale Ziel des Projekts GameOpSys bestand in der Entwicklung einer mobilen Anwendung, die durch die Einbindung der Nutzer:innen mittels Gamification relevante Daten und Informationen zur Optimierung von Energie- und Kostenaspekten (Strom und Wärme) bereitstellt. Durch die Kombination dieser Daten mit Smart-Home-Technologien und dem Internet of Things soll in Zukunft

eine sektorübergreifende Energieoptimierung sowie eine verbesserte Planung von Gebäuden, Quartieren und übergeordneten Energiesystemen ermöglicht werden. Zudem wurde weiterer Forschungsbedarf hinsichtlich der Integration weiterer externer Datenquellen identifiziert.

Smart2B (Horizon 2020, 2021-2024)

Das EU-geförderte Projekt Smart2B (Horizon 2020) zielte darauf ab, Bürgerinnen, Gebäudenutzer:innen und Energiegemeinschaften ins Zentrum der Energiewende zu stellen. Ein intelligentes Gebäudesystem bestehend aus IoT-Geräten, einer zentralen Datenplattform und diversen Analyseservices ermöglicht die Echtzeit-Interaktion zwischen Nutzer:innen, Gebäuden und dem Energienetz. Dies verbessert die Energieeffizienz und unterstützt lokale Flexibilitäten für den Energiemarkt zu erschließen. Bestehende Gebäude wurden durch smarte Technologien wie IoT-Anwendungen zu aktiven Elementen des Energiesystems weiterentwickelt.

ANSERS (Zukunftsfonds Steiermark, 2021-2023)

Die energetische Optimierung von Gebäuden spielt eine Schlüsselrolle für Nachhaltigkeit, indem Energieverbrauch reduziert, Effizienz gesteigert und der Wechsel zu erneuerbaren Energien gefördert wird. Dabei sind die aktive Einbindung und das Wohlbefinden der Nutzer:innen zentral.

Im Projekt ANSERS wurde diese Partizipation interdisziplinär untersucht. Drei Studien kombinierten psychologische Datenerhebungen und die Entwicklung von Apps: Eine App zeigte die Verfügbarkeit erneuerbarer Energien, ermöglichte Energieverbrauchsplanung und stellte Verbrauchervergleiche dar. Dabei wurden Nutzerverhalten, Datenbereitstellung und Systemrückmeldungen untersucht. Eine weitere App sammelte Feedback zum Wohlbefinden in Gebäuden mit automatischer Heizungs- und Kühlungsregelung, um Raumklima und Nutzerzufriedenheit zu optimieren. Die Studien zeigten, dass Normen, Ziele und Überzeugungen über kollektives Handeln nachhaltiges Verhalten fördern und die Partizipation an smarten Energiesystemen anregen. Die Ergebnisse liefern wichtige Erkenntnisse für die Entwicklung intelligenter Gebäuderegulierungssysteme

IoT-Plattform „Inframonitor“ an der Technischen Universität Graz

Zusätzlich wurde vom Institut für Softwaretechnologie im „Innovation District Inffeldgasse“ eine IoT-Plattform implementiert, die eine bidirektionale Kommunikation mit verschiedenen Geräten ermöglicht. Diese Plattform stellt eine Testumgebung bereit, um die Interaktion zwischen Sensoren, Geräten und KI-Modellen in realen Bedingungen zu evaluieren.

3.2.2. Zusammenfassung des nationalen und internationalen Wissensstands

Cloud Computing

Cloud Computing bietet eine Vielzahl von Vorteilen, darunter Kosteneffizienz, hohe Verfügbarkeit und Flexibilität. Dennoch stehen diesen Vorteilen erhebliche Herausforderungen in den Bereichen Sicherheit, Zuverlässigkeit und Vertrauen gegenüber. Insbesondere bei sicherheitskritischen IoT-Anwendungen, die eine verlässliche und sichere Datenverarbeitung sowie -kommunikation erfordern, können diese Schwachstellen die Integration und den Betrieb erheblich beeinträchtigen (Gala et al., 2021).

Darüber hinaus sind derartige Systeme anfällig für Cyberangriffe und Betriebsstörungen, die ihre Verfügbarkeit und Integrität gefährden können (Dekker et al., 2012; Alouffi et al., 2021).

Edge Computing

Im Vergleich zum Cloud Computing bietet Edge Computing deutliche Vorteile, darunter eine reduzierte Latenz, lokale Datenverarbeitung und ein höheres Maß an Datenschutz. Ein vielversprechender Ansatz in diesem Kontext ist Edge-ready Künstliche Intelligenz (KI), bei der KI-Modelle direkt auf Edge-Geräten ausgeführt werden – mit oder ohne Internetverbindung. Dennoch bleibt die begrenzte Rechenkapazität von Edge-Geräten eine zentrale Herausforderung, insbesondere im Hinblick auf rechenintensive KI-Modelle wie Deep Neural Networks (Jakovljevic et al., 2014).

Künstliche Intelligenz für Energy Services

Studien zeigen, dass KI-basierte Energy Services erhebliche Energieeinsparungen ermöglichen können. So lässt sich durch prädiktive Regelung der Energieverbrauch von Gebäuden um bis zu 28 % reduzieren, während automatische Fehlererkennung Einsparungen zwischen 15 % und 30 % ermöglicht. (Quelle?) Diese Erkenntnisse unterstreichen die Bedeutung von KI für die Optimierung der Energieeffizienz in Gebäuden.

Im maschinellen Lernen (ML) werden problembezogene Datenbanken genutzt, um einen Algorithmus zu trainieren. Dieser Algorithmus lernt iterativ aus den Eingabedaten und verbessert dabei kontinuierlich seine Leistung. Dabei wird ein internes Modell automatisch erstellt, das komplexe Muster in den Daten beschreibt und erkennt. Meistens werden hochdimensionale Daten verwendet, um Aufgaben wie Regression, Klassifikation oder Clustering auszuführen. (Janiesch et al., 2021)

ML-Algorithmen können wie folgt kategorisiert werden:

- **Überwachtes Lernen (Supervised Learning)**
Die Eingangsdaten sind mit den tatsächlichen Zielwerten gekennzeichnet. Diese Paare werden genutzt, um die Parameter des ML-Modells während des Trainings zu bestimmen. Typischerweise werden hierbei Aufgaben wie Regression und Klassifikation durchgeführt.
- **Unüberwachtes Lernen (Unsupervised Learning)**
Es werden keine Zielwerte vorgegeben. Das Ziel des Modells ist es, gemeinsame Eigenschaften und Muster im Datensatz zu identifizieren. Dies wird oft durch Dimensionsreduktion der Eingabedaten mithilfe statistischer Ansätze erreicht. Diese Technik wird z. B. beim Clustering eingesetzt.
- **Verstärkendes Lernen (Reinforcement Learning)**
Beim verstärkenden Lernen wird ein Ziel und eine Liste erlaubter Aktionen oder Operationen definiert. Das gewünschte Verhalten wird belohnt. Das gewünschte Verhalten wird durch ein Belohnungssystem verstärkt, sodass der ML-Algorithmus darauf trainiert wird, die Belohnung zu maximieren, um das vorgegebene Ziel effizient zu erreichen. Zur Optimierung des Lernprozesses ist eine Simulationsumgebung erforderlich, in der der Algorithmus verschiedene Strategien testen und durch iterative Experimente seine Leistung kontinuierlich verbessern kann.

Je nach zugrundeliegender Aufgabe können verschiedene ML-Algorithmen wie Support Vector Machines, k-nächste Nachbarn, Entscheidungsbäume oder künstliche neuronale Netze eine geeignete Wahl sein.

Neuronale Netzwerke (NN) sind in allen drei Typen des maschinellen Lernens anwendbar. Ihr Aufbau orientiert sich am menschlichen Gehirn: zwischen der Eingabe- und Ausgabeschicht liegen mehrere versteckte Schichten, die aus Neuronen bestehen. Die Neuronen der benachbarten Schichten sind miteinander verbunden, was den Datenfluss von der Eingabe zur Ausgabe symbolisiert (Abbildung 1).

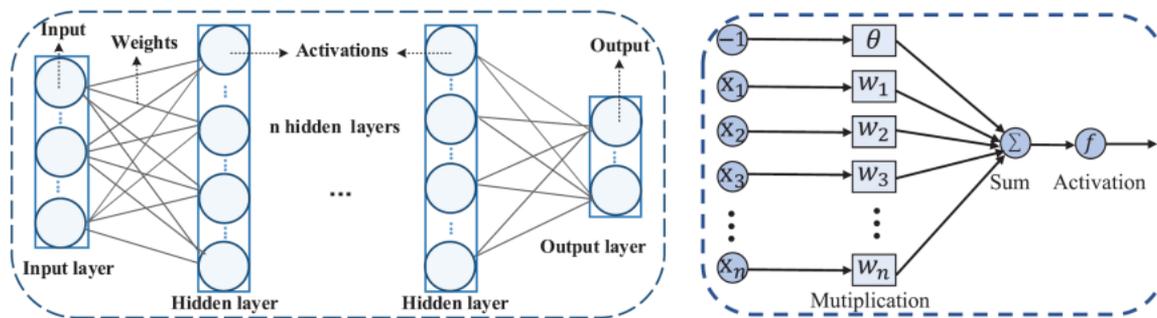


Abbildung 1: Grundstruktur eines neuronalen Netzwerks (links) und eines einzelnen Neurons (rechts), Chen et al (2020)

Ein einzelnes Neuron multipliziert seine Eingaben „ $x_1 - x_n$ “ mit Gewichten, addiert einen Bias-Wert θ und gibt die Summe durch eine nichtlineare Aktivierungsfunktion weiter. Das Ergebnis wird an die nächste Schicht weitergeleitet. Während des Trainings optimiert ein Lernalgorithmus die Gewichte und Bias-Werte, sodass ein gegebener Eingang die gewünschte Ausgabe erzeugt (überwachtes Lernen).

Deep Learning als Teilbereich von ML beschreibt fortgeschrittene neuronale Netzwerke, sogenannte Deep Neural Networks. Diese tief verschachtelten Architekturen nutzen komplexere Operationen wie Convolution oder Zustandsgedächtnisse (State Memory) sowie mehrere Aktivierungsfunktionen in einem Neuron. Dadurch entstehen spezialisierte Modelle wie Convolutional Neural Networks (CNNs) oder Long-Short-Term-Memory (LSTMs). (Janiesch et al., 2021).

ML-Anwendungen sind vom Grundprinzip Blackbox-Systeme, die zwar leistungsstarke Vorhersagen treffen, jedoch eine Ergebnisinterpretation erschweren. Aufgrund dessen etabliert sich zunehmend die Anwendung von sogenannten Explanatory Frameworks in den ML-Entwicklungsumgebungen. Diese Frameworks ermöglichen eine nachvollziehbare Analyse der Entscheidungsprozesse, wodurch das Vertrauen in ML-Modelle erhöht und deren Anwendung insbesondere in sicherheitskritischen oder regulierten Bereichen erleichtert wird.

Lundberg & Lee (2017) haben mit SHapley Additive exPlanations (SHAP) ein Framework entwickelt, welches Entwickler:innen dabei unterstützt, das Verhalten ihrer ML-Modelle zu analysieren. Dabei werden jedem Merkmal sogenannte „Shapley-Werte“ zugewiesen, die den Einfluss des Merkmals auf das Modellergebnis für ein bestimmtes Eingabebeispiel beschreiben. Durchschnittliche Shapley-Werte aus mehreren Eingaben erlauben eine Bewertung der Merkmalsbedeutung.

SHAP basiert auf Konzepten der kooperativen Spieltheorie und bewertet den individuellen Beitrag eines „Spielers“ (Features) zum Gesamtergebnis, indem sein Einfluss durch Anwesenheit oder Abwesenheit simuliert wird. In ML-Modellen wird ein „Nicht-Beitritt“ durch das Ersetzen eines Merkmals mit einem festen Referenzwert dargestellt.

Da die exakte Berechnung der Shapley-Werte für alle Merkmale NP-hart ist, wird sie in der Praxis häufig auf eine Teilmenge der Merkmalskombinationen beschränkt, um die Berechnungseffizienz zu verbessern (Lundberg, 2018).

Das Framework bietet spezifische Analyse-Tools für verschiedene Modelltypen, darunter Ensemble-Modelle, Deep-Learning-Modelle und lineare Modelle. Für neuronale Netzwerke kann der sogenannte „Kernel-Explainer“ verwendet werden, der den Einfluss einzelner Merkmale simuliert, indem fehlende Merkmale durch typische Werte (z. B. Mittelwerte) ersetzt werden. Die SHAP-Werte werden dann durch ein spezielles gewichtetes lineares Regressionsmodell berechnet.

3.2.3. Zielsetzungen und Innovationsgehalt von OctoAI

Das zentrale Ziel des Projekts „OctoAI“ ist die Entwicklung hochleistungsfähiger Edge-KI-Technologien, die speziell auf die Anforderungen intelligenter Gebäude zugeschnitten sind. Der Schwerpunkt liegt auf der Implementierung effizienter KI-Methoden, die auf rechenressourcenbeschränkten Geräten wie IoT-Endgeräten ausgeführt werden können – ohne dabei auf eine permanente Cloud-Anbindung angewiesen zu sein. Dies ermöglicht eine autarke, datenschutzfreundliche und latenzoptimierte KI-Verarbeitung direkt an der Quelle der Datenerzeugung.

Das Projekt „OctoAI“ geht über den aktuellen Stand der Technik hinaus und verfolgt drei wesentliche Innovationsansätze und Forschungsfragen:

1. Skalierbare und ressourcenschonende KI-Modelle

Es werden neue KI-Ansätze entwickelt, die speziell für Edge-Geräte mit begrenzter Rechenleistung optimiert sind. Unter anderem wird der Einsatz von „TinyML“ untersucht, das leistungsstarke Modelle mit minimalem Ressourcenverbrauch ermöglicht.

2. Transfer-Learning-Methoden

Die Einführung von Transfer-Learning-Ansätzen ermöglicht die Übertragbarkeit von KI-Modellen auf unterschiedliche Anwendungsbereiche. Dadurch wird die Skalierbarkeit erhöht und der Implementierungsaufwand verringert.

3. Nutzerzentrierte Energy Services

Im Mittelpunkt stehen Energy Services, die durch eine bidirektionale Kommunikation mit den Nutzern interagieren können. Über intuitive Dashboards und Apps erhalten Nutzer nicht nur Einblicke in den Energieverbrauch, sondern können aktiv an der Steuerung teilnehmen.

Mit diesen Innovationen leistet „OctoAI“ einen wichtigen Beitrag zur Weiterentwicklung intelligenter Gebäude und ebnet den Weg für ressourceneffiziente und nutzerzentrierte Technologien.

Darüber hinaus verfolgt das Konsortium mit dem Projekt das mittelfristige Ziel, die Technologieführerschaft im Bereich Edge-KI für intelligente Gebäude in Österreich und Europa zu übernehmen. Durch die Kombination von innovativer Forschung und praktischer Implementierung soll ein nachhaltiger Fortschritt in der modernen Gebäudetechnologie erzielt werden.

4 Projektinhalt

Das Projekt OctoAI folgte einem systematischen Ansatz, der in mehrere aufeinander aufbauende Arbeitspakete unterteilt war. Diese gliederten sich in die Anforderungsanalyse, die Entwicklung von Modellen und Technologien, die prototypische Implementierung sowie die Validierung der KI-Modelle auf diversen Edge-Devices.

4.1. Anforderungsanalyse

Zu Beginn des Projekts wurden die Anforderungen für Edge-KI-Lösungen systematisch erhoben und präzisiert. Dies geschah durch 15 detaillierte Interviews mit Gebäudemanagern, Ingenieuren, gemeinnützigen Wohnbauträgern und weiteren relevanten Stakeholdern wie Gebietskörperschaften oder Energieberatungsunternehmen. Die Interviews konzentrierten sich auf mehrere zentrale Themen, um ein umfassendes Verständnis der technischen und organisatorischen Anforderungen an Edge-KI-Lösungen zu gewinnen. Ein Schwerpunkt lag auf der Einschätzung der technischen Voraussetzungen in bestehenden Gebäuden sowie der Verbreitung und Nutzung von IoT-Sensorik. Zudem wurden die spezifischen Anforderungen der Nutzer:innen an intelligente Gebäude erfasst. Ein weiterer wichtiger Aspekt war das Feedback zu projektrelevanten Use Cases, darunter Warmwasserbedarfsprognosen, Anwesenheitserkennung und Behaglichkeitsbewertung. Darüber hinaus wurden potenzielle Sicherheitsbedenken im Zusammenhang mit Cloudlösungen und Cybersecurity thematisiert.

Zentrale Erkenntnisse waren:

- Die Belegungserkennung kann für einen effizienten Gebäudebetrieb auf der Grundlage des tatsächlichen Bedarfs von großem Nutzen sein - sie ist besonders in großen, kommerziell genutzten Gebäuden wie Büros von Bedeutung.
- Die Luftqualität in Innenräumen und der thermische Komfort der Gebäudenutzer haben in den letzten Jahren zunehmend an Bedeutung gewonnen.
- In der Regel kein großes Nutzen-/Anwendungspotenzial für Warmwasserbedarfsprognosen, außer bei Mehrfamilienhäusern mit ~60-90 Wohnungen, die mit Wärmepumpen, Warmwasserspeichern und PV-Systemen/Solaranlagen ausgestattet sind. Voraussetzung für eine erfolgreiche Implementierung ist jedoch ein Lastverschiebungspotenzial sowie die technische Kompatibilität des Heizsystems mit Prognosemodellen.
- Die meisten Interessenvertreter haben weder einschlägige Erfahrungen mit IoT-Geräten oder drahtlosen Sensoren, noch sind sie sich der potenziellen Vorteile von Edge Devices bewusst. Am häufigsten genannte IoT-Anwendungen sind intelligente Zähler und Luftqualitätsmessungen in Innenräumen.
- Bei den Präferenzen für das Software-Hosting (On-Premise, Cloud, hybride Lösungen) gibt es keine einheitliche Tendenz- dies hängt meist von den internen IT-Richtlinien ab. Einige Stakeholder bevorzugen jedoch Cloud-Lösungen, da sie den internen IT-Abstimmungsaufwand reduzieren.

- Die meisten Beteiligten konnten keine konkreten Aussagen dazu treffen, ob spezielle Cybersicherheitsrichtlinien für ihre Gebäude existieren oder ob bestimmte IoT-Geräte und -Protokolle auf einer Blacklist stehen.

Um die wichtigsten Erkenntnisse aus den Interviews zu vertiefen und zu ergänzen, wurde eine Stakeholder-Befragung zum Thema „Anwendbarkeit von IoT-Infrastrukturen und Edge-Ready AI-Modellen für Smart Buildings“ entwickelt und vom Konsortium sowie vom Innovationslabor „Digital Findet Stadt“ an über 4.000 Empfänger verschickt. Schlussendlich nahmen 46 branchenspezifische, qualifizierte Teilnehmer:innen an der Umfrage teil. (Abbildung 2)



Abbildung 2: Stakeholder-Survey

Die Ergebnisse der Umfrage (Abbildung 3) zeigten, dass Gebäudeautomationssysteme (BAS) bzw. Gebäudeleittechnik (GLT) besonders in kommerziellen Gebäuden und großen Wohngebäuden verbreitet ist, die Umfrageteilnehmer:innen aber mit Hindernissen wie hohen Kosten, mangelnder technischer Expertise und Interoperabilitätsproblemen konfrontiert sind. Etablierte Tools und Protokolle wie BACnet und KNX sind weit verbreitet, während das Interesse an KI-basierten und prädiktiven Analysetools eine Verlagerung hin zu fortschrittlicherer Automatisierung widerspiegelt. Die Umfrage unterstreicht Möglichkeiten, die Akzeptanz von BAS/GLT zu verbessern, einschließlich finanzieller Anreize und Schulungsprogramme, um Kosten- und Qualifikationslücken zu schließen. Die Förderung offener Standards könnte zudem Interoperabilitätsprobleme abmildern und so eine breitere Akzeptanz fördern.

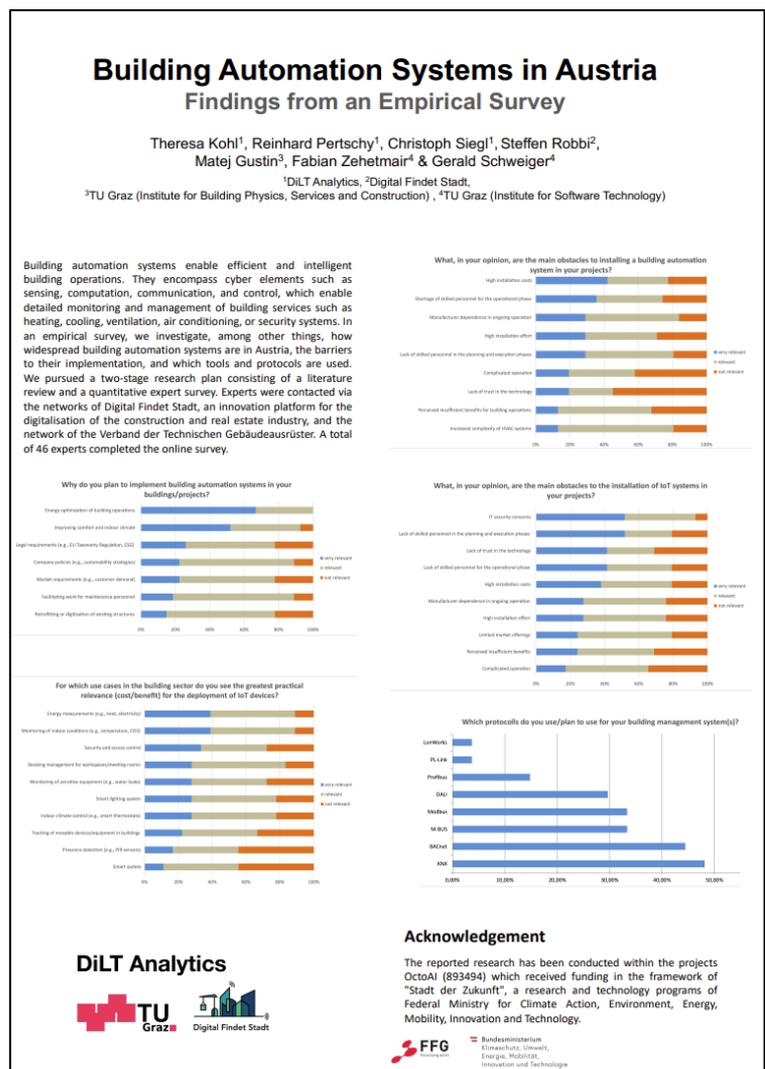


Abbildung 3: publiziertes Poster mit den Umfrageergebnissen

Die Analyse führte zur Identifikation von zwei Use-Cases:

- **Belegungsdetektion:** Ziel war die präzise Erkennung der Anwesenheit von Personen in Räumen auf Basis von Raummessdaten und maschinellem Lernen. Dabei sollte die Belegung eines Büros mit sechs Arbeitsplätzen klassifiziert werden. Die Kategorien umfassten leer, geringe Belegung (1–2 Personen) und hohe Belegung (>3 Personen). Zur Umsetzung wurde Sensorfusion verwendet, womit verschiedene Sensordaten kombiniert wurden. Als Testumgebung wurde das Büro von DiLT Analytics in Graz genutzt.
- **Vorhersage des thermischen Komforts:** Hier lag der Schwerpunkt auf der Entwicklung von Modellen zur Prognose der Raumtemperatur bis zu 24 Stunden im Voraus. Die ermittelten Werte wurden in Komfortzonen eingeteilt, basierend auf gängigen Normen und Wetterprognosen. Die Tests erfolgten in einem Büro der TU Graz.

4.2. Entwicklung von Modellen und Technologien

Ein Kernelement war, die Algorithmen so zu gestalten, dass sie Echtzeitfähigkeit gewährleisten und gleichzeitig Datenschutzanforderungen erfüllen, indem die Datenverarbeitung lokal auf den Edge-Geräten erfolgt.

Auf Basis der identifizierten Anforderungen für die beiden Use Cases wurden verschiedene Modelle für maschinelles Lernen entwickelt, darunter Feedforward Neural Networks (FFNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) und ARX-Modelle. Diese Modelle wurden speziell für die Anwendung auf ressourcenbeschränkten Geräten, wie beispielsweise Raspberry Pi, optimiert.

4.2.1. Belegungsdetektion

Studien zeigen, dass durch belegungsbasierte Heiz- und Kühlsysteme Energieeinsparungen von bis zu 30 % bei der Beleuchtung und 20 % bei HVAC-Systemen (Heizung, Lüftung, Klimatisierung) möglich sind (Schweiger et al. 2020).

Ansätze zur Belegungsdetektion lassen sich in direkte und indirekte Methoden unterteilen. Direkte Methoden bieten hohe Genauigkeit, greifen jedoch stark in die Privatsphäre ein, da sie auf Kameras oder Smartphones basieren. Indirekte Methoden erfassen Umgebungsdaten und ermöglichen eine datenschutzfreundlichere Belegungsdetektion. Sie lassen sich in drei Hauptkategorien unterteilen (Jin et al. 2016; Rueda et al. 2020):

1. **Analytische Methoden** modellieren die Interaktion zwischen Personen und ihrer Umgebung mathematisch. Sie basieren auf physikalischen oder statistischen Modellen, die beispielsweise Temperatur-, CO₂- oder Bewegungsdaten auswerten, um die Belegung abzuleiten.
2. **Wissensbasierte Methoden** nutzen regelbasierte Systeme, die auf vordefinierten Heuristiken oder Expertensystemen beruhen. Diese Methoden nutzen vorgefertigte Regeln zur Abschätzung der Belegung, beispielsweise basierend auf Raumfunktionen oder Zeitplänen.
3. **Datenbasierte Methoden** setzen maschinelles Lernen ein, um Muster und Zusammenhänge in Sensordaten zu erkennen. Durch den Einsatz von KI-Algorithmen können diese Modelle kontinuierlich aus den Daten lernen und sich an veränderte Bedingungen anpassen, wodurch sie eine hohe Genauigkeit und Flexibilität bieten.

Eine Evaluierung von 15 relevanten Publikationen zwischen 2014-2023 in diesem Bereich hat gezeigt, dass die häufigsten Merkmale („Features“), die in datenbasierten Ansätzen verwendet werden CO₂-Konzentration in Innenräumen, den Status von Türen/Fenstern, Energieverbrauch, Lichtintensität, Bewegungserkennung durch PIR-Sensoren, Luftdruck, Lautstärke, relative Luftfeuchtigkeit, Raumtemperatur, Gehäusetemperatur der Arbeitscomputer und flüchtige organische Verbindungen (volatile organic compounds VOC) in der Luft umfassen. Die Untersuchung hat bestätigt, dass datenbasierte Methoden gut für die Belegungserkennung geeignet sind. Allerdings bestehen weiterhin Herausforderungen hinsichtlich der Skalierbarkeit und Übertragbarkeit dieser Ansätze. Die Beschaffung hochwertiger Datensätze für das Training dieser Modelle ist schwierig und erfordert oft zusätzliche Sensoren, wie Kameras, um eine verlässliche Referenzbasis (Ground-Truth) zu schaffen. Dies steht jedoch im Widerspruch zum Ziel der Nicht-Interferenz und wirft Datenschutzbedenken auf. Um diesen Problemen entgegenzuwirken, wurde zusätzlich das Potenzial von Transfer Learning untersucht. Für die entwickelten KI-Modelle von OctoAI wurden die in Abbildung 4 dargestellten Features verwendet. Die Belegungserkennung wurde als ein Klassifikationsproblem mit drei Kategorien definiert:

- leer (0 Personen)
- geringe Belegung (1–2 Personen) und
- hohe Belegung (>3 Personen)

Name	Unit	Range	Accuracy
Date + Time	[1]	yyyy:mm:dd hh:mm	-
Motion Counter	[1]	0-255	-
Temperature	[°C]	0-50	±0.2
Relative Humidity	[%]	0-85	±2
Light	[lux]	4-2000	±10
CO2 Conc.	[ppm]	0-2000	±50
Average Loudness	[dBspl]	35-70	±5
Peak Loudness	[dBspl]	65-99	±5
#Open Windows	[1]	0-2	-
#Open Doors	[1]	0-1	-
IR-based Occupancy	[1]	0 = Unocc. 1 = Pending 2 = Occupied	-

Abbildung 4: verwendete Features für die OctoAI-KI-Modelle zur Belegungserkennung

4.2.2. Komfortvorhersage

Im Use-Case der Komfortvorhersage wurden die prognostizierten Raumtemperaturen des folgenden Tages in sechs Komfortzonen kategorisiert. Diese Zonen repräsentierten die Bandbreite der thermischen Bedingungen, die von den Nutzern subjektiv wahrgenommen werden könnten. Die Bewertung des thermischen Komforts basiert üblicherweise auf Standards wie ISO 7730, ASHRAE 55 und EN1521. Der thermische Komfort wird dabei aus drei Hauptperspektiven beschrieben:

- **Physiologisch:** Der Zustand des Wohlbefindens, der die Zufriedenheit mit der thermischen Umgebung ausdrückt.
- **Psychologisch:** Die thermische Wahrnehmung des Menschen, die durch Nervensensoren und thermische Rezeptoren beeinflusst wird.
- **Rational:** Das thermische Empfinden, das vom Wärmehaushalt des menschlichen Körpers abhängt.

Zur Bewertung des thermischen Komforts stehen zwei Ansätze zur Verfügung, wobei für den OctoAI-Use Case letzterer Ansatz gewählt wurde:

1. **Steady-State-Ansätze** gehen von einer konstanten Innenraumumgebung aus (z. B. Temperatur und Luftfeuchtigkeit), ohne Variationen in der Aktivität oder Kleidung der Nutzer zu berücksichtigen.
2. **Adaptive Ansätze** basieren auf der Theorie, dass der menschliche Körper sich an Außen- und Innenklima anpasst. Thermischer Komfort ist subjektiv und von individuellen Faktoren wie Stoffwechselrate, Kleidung und Aktivitätsniveau abhängig.

Ein neuartiger Aspekt des gewählten Ansatzes ist die Einbindung von thermischen Komfortzonen gemäß der EN 16798-1-Norm, die adaptive Komfortkriterien für Innenräume definiert. Für diese Studie wurde Kategorie II gewählt, da sie üblicherweise für bestehende Bürogebäude empfohlen wird. Sie stellt einen ausgewogenen Kompromiss zwischen Nutzerkomfort und realistischen Gebäudeanforderungen dar.

Die oberen und unteren Temperaturgrenzen für den thermischen Komfort in diesem adaptiven Modell werden anhand der gleitenden Durchschnittstemperatur im Außenbereich bestimmt. Diese wird nach der vereinfachten Methode aus EN 16798-1 berechnet, die die Außentemperaturen der letzten Tage berücksichtigt.

Das gewählte Modell unterteilt die prognostizierten Raumtemperaturen für die nächsten 24 Stunden in sechs Komfortzonen, die unterschiedliche thermische Bedingungen widerspiegeln:

- Heiß
- Warm
- Angenehm warm
- Angenehm kühl
- Kühl
- Kalt

Durch die Vorhersage der Raumtemperatur für die nächsten 24 Stunden und deren Zuordnung zu den Komfortzonen ermöglicht das Modell proaktive Maßnahmen zur Optimierung des Gebäudebetriebs. Gebäudemanager erhalten dadurch wertvolle Handlungsempfehlungen, um potenzielle Komfortprobleme frühzeitig zu erkennen und gezielt entgegenzuwirken. Diese datenbasierte Komfortbewertung und -vorhersage trägt wesentlich zur Energieeffizienz und Nutzerzufriedenheit in intelligenten Gebäuden bei (siehe Abbildung 5).

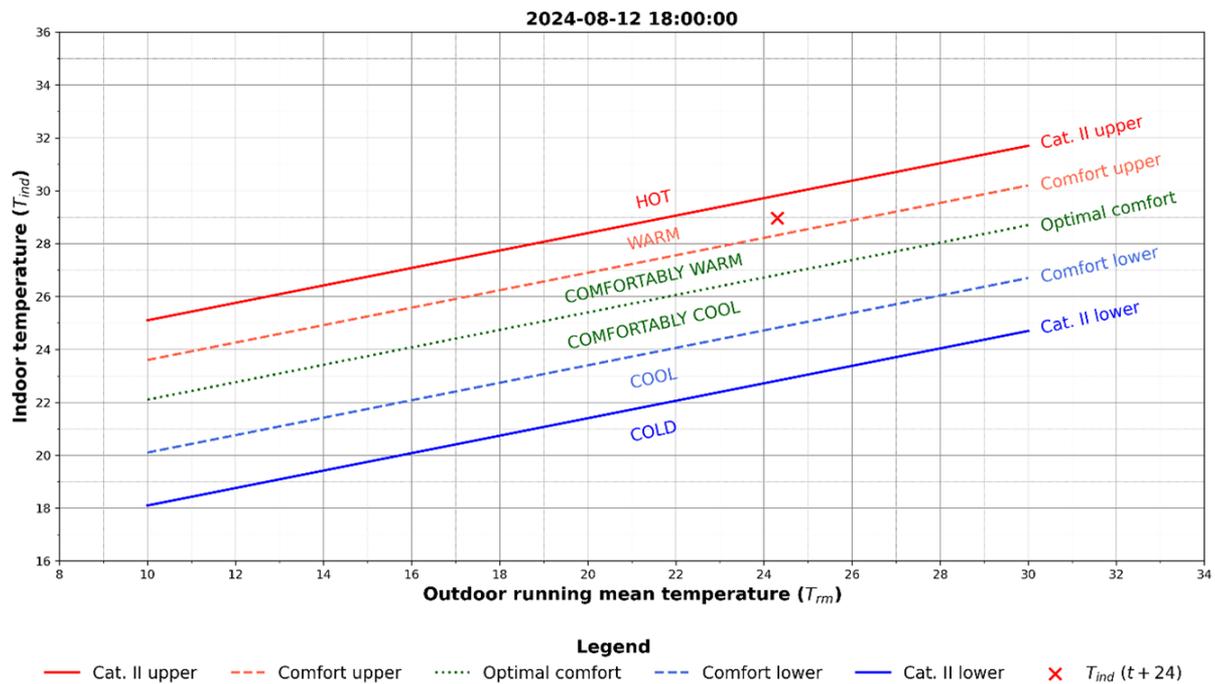


Abbildung 5: Beispielhaftes Ergebnis der Komfortvorhersage für 24h im Voraus: es wird die Komfortzone „WARM“ mit den erwarteten Bedingungen für den 12. August 2024 um 18:00 Uhr prognostiziert

4.3. Prototypische Implementierung, verwendete Datensätze und Erhebung

Für die Feldexperimente wurden drei Testaufbauten an verschiedenen Standorten implementiert, um die Anwendbarkeit von IoT-Infrastrukturen und Edge-fähigen KI-Modellen unter realen Bedingungen zu untersuchen. Für die Datenerfassung wurden Funk-basierte Sensoren eingesetzt, die Parameter wie CO_2 -Konzentration, Lichtintensität, Schallpegel, Bewegung und Temperatur messen konnten. Zusätzlich wurden Sensoren zur Erkennung des Tür- und Fensterstatus installiert, um weitere relevante Daten für die Modelle bereitzustellen. (siehe Abbildung 6, Abbildung 7 & Abbildung 8)

Für den Use Case der Belegungserkennung wurden zwei unterschiedliche Büroräume mit entsprechender Sensorik ausgestattet, um verschiedene Szenarien zu untersuchen. Einerseits wurde das Büro von DiLT Analytics in Graz, das über sechs Arbeitsplätze verfügt, für die primäre Belegungserkennung genutzt. Andererseits wurde ein benachbartes Büro mit zwei Arbeitsplätzen gezielt für die Untersuchung von Transfer-Learning eingesetzt. Das neu errichtete Bürogebäude, in dem die Tests durchgeführt wurden, ist mit einer Lüftungsanlage, einem bauteilaktivierten Kühlsystem und einer Fußbodenheizung ausgestattet.

Die Datenerfassung begann im August 2023 und umfasste den Einsatz kabelloser LoraWan-Innenraummessgeräte zur Erfassung verschiedener Umgebungsparameter. Dazu gehörten Raumtemperatur, Luftfeuchtigkeit, Bewegungserkennung, Fenster- und Türkontakte, CO_2 -Konzentration, Lichtintensität, Geräuschpegel. Ergänzend wurde eine automatische Personenzählung

implementiert, um die Belegung in Echtzeit zu erfassen. Zur Validierung der Modelle und als Ground Truth diente ein Kamerasystem, das mit den Sensordaten abgeglichen wurde. Die Mitarbeiter:innen in beiden Büros wurden im Vorfeld über die temporäre Kamerainstallation sowie die Bildaufzeichnung informiert. Zur Einhaltung der Datenschutzrichtlinien unterzeichneten alle Beteiligten entsprechende Einverständniserklärungen, die ihre Zustimmung zur Nutzung der erfassten Daten für die Forschung und Modellevaluierung bestätigten. Abbildung 9 zeigt zwei Bilder aus Raum A und B, die mit den installierten Reolink-Kameras aufgenommen wurden. Zur manuellen Erfassung der Belegung wurden die Bilder durchgesehen, und die Anzahl der Personen im Büro für die jeweilige Uhrzeit in einer Excel-Tabelle notiert. Die Excel-Dateien wurden anschließend als CSV-Dateien exportiert, die als Ground Truth für das Modelltraining verwendet wurden.



Abbildung 6: Raum A (DiLT Büro) mit diverser Sensorik - 3 unterschiedliche Elsys-Sensoren (Mitte), People Counter System (Links) und Kamera (rechts) zur Personenzählung

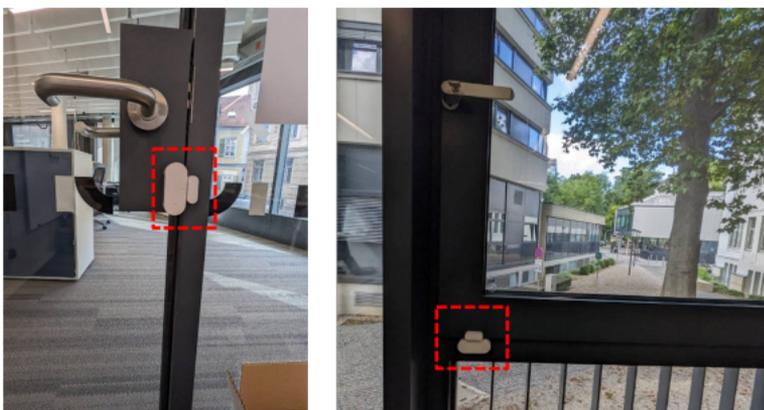


Abbildung 7: Tür- und Fensterkontakte der Fa. Dragina in Raum A (DiLT Büro)

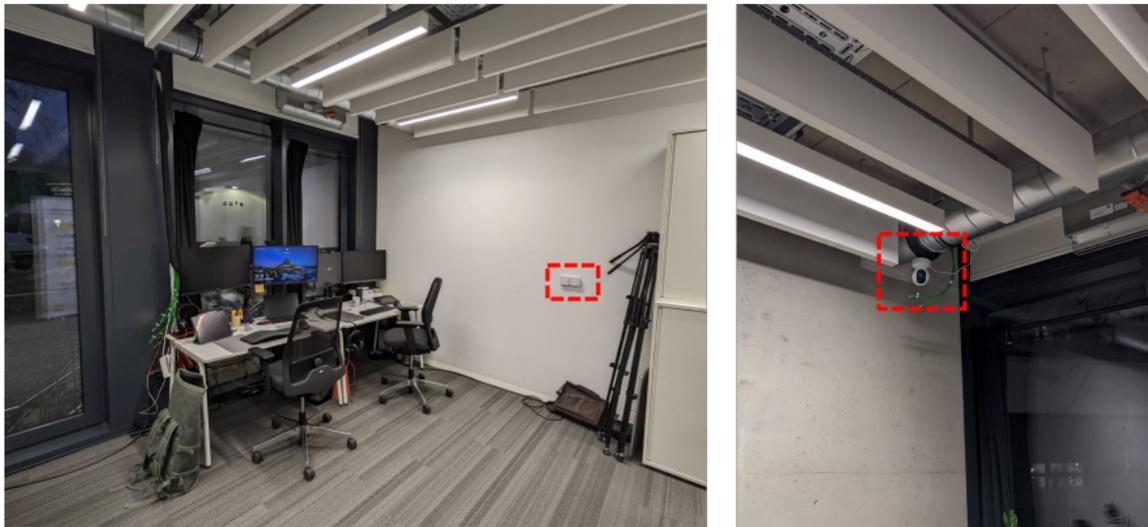


Abbildung 8: Raum B mit installierter Sensorik und Kamera zur Untersuchung von Transfer-Learning

Ein zweiter Messaufbau wurde im Sommer 2024 im Bürogebäude der TU Graz installiert. Die erfassten Daten umfassen Messwerte der Innenraumumgebung, die entscheidende Eingaben für das Vorhersagemodell darstellen. Die Raumtemperaturdaten wurden mithilfe von LoRaWAN-Sensoren erhoben, die in den Büroräumen installiert wurden. Ergänzend dazu wurden Wetterdaten über Meteoblue für den jeweiligen Standort abgefragt, um externe Einflüsse auf das Raumklima zu berücksichtigen.

Zudem wurde ein benutzerzentriertes Dashboard entwickelt das eine intuitive Interaktion mit den gesammelten Daten und den Modellen ermöglichte. Es stellte den Nutzern Rückmeldungen zur Verfügung und unterstützte eine datengestützte Entscheidungsfindung im Gebäudemanagement.



Abbildung 9: beispielhafte Kamerabilder in Büro A (links) und Büro B (rechts)

4.4. Validierung und Deployment

Wie im vorherigen Abschnitt beschrieben, konnten die entwickelten Edge-KI-Modelle unter realen Bedingungen getestet werden beispielsweise während der Hitzeperioden in den Jahren 2023 und 2024. Diese Tests ermöglichten eine umfassende Evaluierung der Belastbarkeit der Modelle sowie eine Analyse ihrer Vorhersagegenauigkeit unter extremen klimatischen Bedingungen.

Abbildung 10 zeigt den implementierten grundlegenden Aufbau der Edge-KI-Entwicklungsumgebung. Als LoRa-Gateway wurde eine Kerlink iFemtoCell verwendet, auf der zwei ChirpStack-Softwarekomponenten laufen: ChirpStack Concentrator und ChirpStack MQTT Forwarder. Ersterer verwaltet die ein- und ausgehenden Messdaten aus den Büros, während Letzterer diese via MQTT auf das verwendete Edge-Device weiterleitet, welches über LAN angebunden ist. Am Edge-Device erfolgt anschließend die Datenverarbeitung mittels unterschiedlicher KI-Modelle sowie die Datenaufbereitung für das Dashboard.

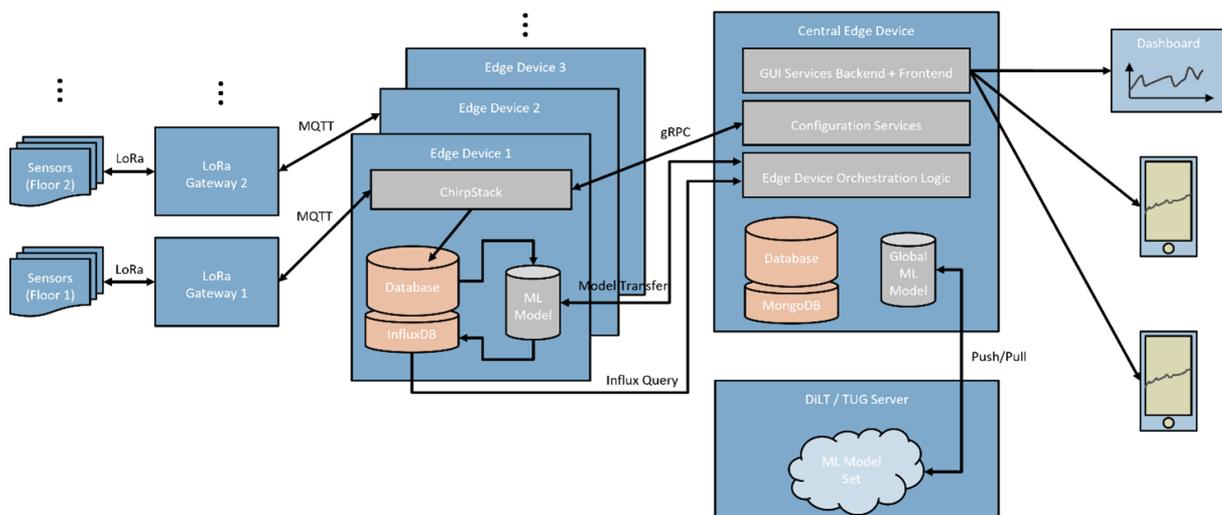


Abbildung 10: Prinzipskizze des implementierten Testset-Ups mit Datenerfassung und Edge-Devices

Das entwickelte Dashboard soll eine bidirektionale Kommunikation mit den Nutzer:innen ermöglichen und eine intuitive Schnittstelle für die Interaktion mit den Ergebnissen der eingesetzten Modelle bieten. Ziel ist es, den Nutzern nicht nur Einblicke in die erfassten Daten und Modellprognosen zu geben, sondern auch eine aktive Rückmeldung zu ermöglichen. In der ersten Entwicklungsphase wurden verschiedene Design-Mock-ups erstellt, um unterschiedliche Konzepte und Benutzerführungen zu testen. Diese Mock-ups wurden innerhalb des Projektteams als interaktive Klickdummies erprobt und evaluiert (Abbildung 11).

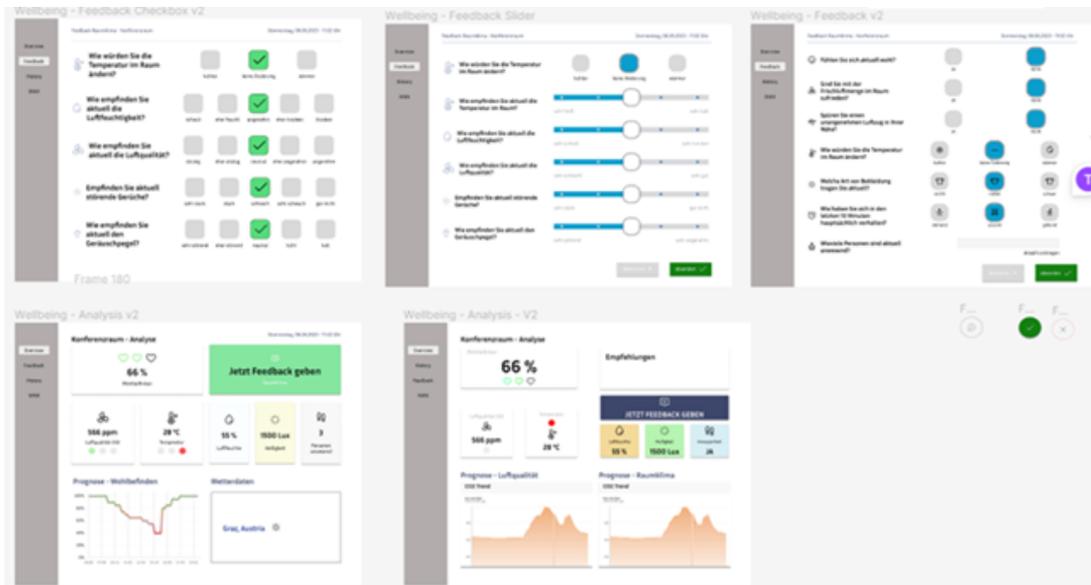


Abbildung 11: unterschiedliche Mock-Ups für das OctoAI-Dashboard

Der bereits in Vorprojekten entwickelte Technologiestack, der auf Tailwind für das Design und JavaScript für die Frontend-Entwicklung basierte, erwies sich als herausfordernd in Bezug auf Skalierbarkeit und Implementierung. Aus diesem Grund wurde das Dashboard-Framework vollständig neu aufgebaut und ein robusteres, wartungsfreundlicheres System umgestellt. Der modernisierte Stack nutzt TypeScript für das Backend und MaterialUI für die Frontend-Komponenten, wodurch eine höhere Wiederverwendbarkeit, Skalierbarkeit und Effizienz bei der Entwicklung und Bereitstellung erreicht wird. Dieser technologische Wechsel stellte sicher, dass das Dashboard nicht nur den aktuellen Projektanforderungen gerecht wird, sondern auch für zukünftige Projekte leicht erweiterbar und anpassbar bleibt..



Abbildung 12: Getestetes Edge-Device zur Modellentwicklung und für das Modelldeployment

5 Ergebnisse

Das Projekt OctoAI verfolgte das Ziel, die nächste Generation von Edge-KI-Technologien für intelligente Gebäude zu entwickeln und erfolgreich zu implementieren. Durch den Einsatz maschineller Lernmodelle und innovativer Ansätze wurden energieeffiziente und nutzerzentrierte Lösungen geschaffen, die eine optimierte Gebäudeautomation ermöglichen.

Das OctoAI-Projekt trägt direkt zu den Zielen des österreichischen Programms „**Stadt der Zukunft**“ bei, das eine nachhaltige und zukunftsfähige Stadtentwicklung anstrebt.

OctoAI unterstützt insbesondere die Themenfelder **Digitales Planen, Bauen und Betreiben** sowie **Plus-Energie-Quartiere**. Durch den Einsatz von Edge-KI zur Optimierung von Heizungs-, Lüftungs- und Klimasystemen leistet das Projekt einen Beitrag zur Energieeffizienz und Reduktion des Ressourcenverbrauchs.

Darüber hinaus trägt das Projekt zu den strategischen Zielen des Programms bei:

- **Nachhaltiges Energiesystem:** Die entwickelten KI-Modelle ermöglichen eine energieeffiziente Gebäudeautomation.
- **Reduktion der Klimawirkung:** Optimierte Regelungssysteme reduzieren den Energieverbrauch und damit verbundene CO₂-Emissionen.
- **Erhöhung der Wettbewerbsfähigkeit:** OctoAI stärkt Österreichs Innovationskraft im Bereich Umwelt- und Energietechnologie.
- **Verbesserung der F&E-Qualität:** Die Integration modernster KI-Technologien in Smart-Building-Systeme fördert den wissenschaftlichen und technischen Fortschritt.

Die erzielten Ergebnisse und gewonnenen Erkenntnisse sind sowohl wissenschaftlich als auch praxisorientiert von großer Relevanz. Sie liefern wertvolle Impulse für die Weiterentwicklung von Edge-KI und schaffen eine solide Grundlage für zukünftige technologische Fortschritte in diesem Bereich. Das Projekt trägt somit zur Umsetzung der österreichischen Klima- und Energiestrategie bei und stellt eine praxisnahe Lösung für die Modernisierung und Effizienzsteigerung von Städten dar.

5.1. Use Case 1 - Belegungserkennung

Für die Klassifizierung der Personenanzahl im Testraum wurden unterschiedliche Modelle angewandt und deren Performance miteinander verglichen. Abbildung 13 zeigt eine übersichtliche Zusammenfassung der gewählten Vorgehensweise bei der Entwicklung der Edge-KI-Modelle.

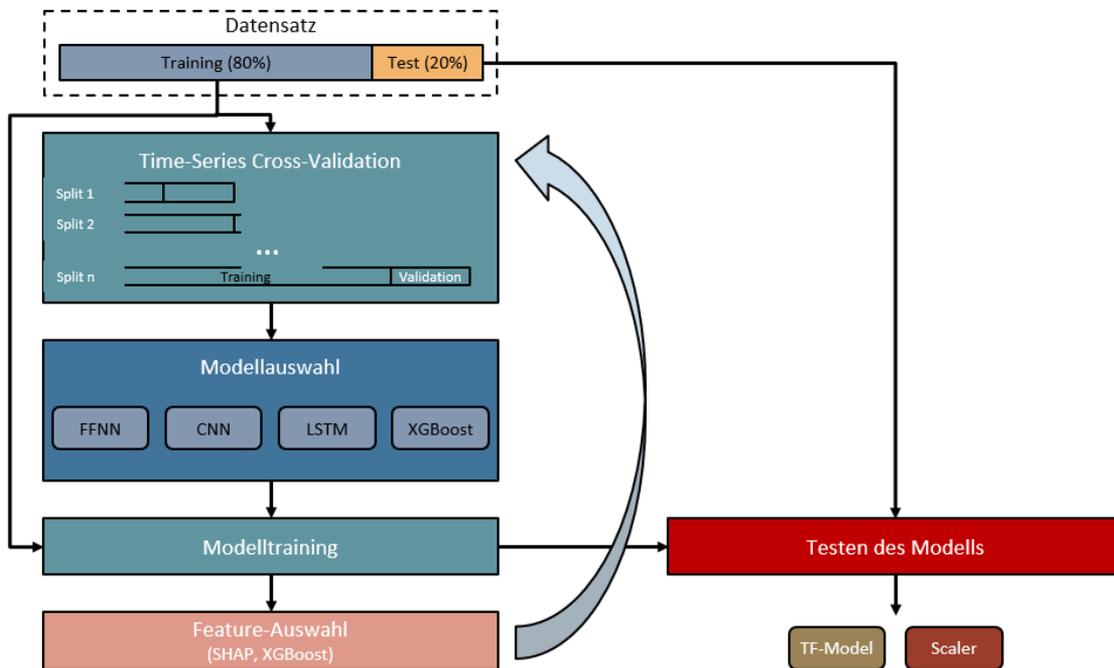


Abbildung 13: Vorgehensweise bei der Modellentwicklung

Zur Leistungsbewertung der Modellarchitekturen wurde eine Zeitreihen-Cross-Validation mit fünf Splits verwendet. Ein Min-Max-Scaler wurde in jedem Split auf dem Trainingsdatensatz trainiert, um die Eingabedaten zu normalisieren.

Die untersuchten Modelltypen waren FFNN, CNN, LSTM und XGBoost-Modelle, wobei zwei Arten von Eingabeschemata getestet wurden:

1. **Tabellarische Verarbeitung:** Tabulare Modelle verarbeiten strukturierte Daten, die in Tabellenform vorliegen. Dabei wird jeder Datenpunkt als eine einzelne Zeile mit verschiedenen Merkmalen (Spalten) dargestellt. Diese Modelle verwenden einen einzigen Wert pro Feature, der häufig ein Mittelwert oder eine aggregierte Kennzahl über einen bestimmten Zeitraum ist.
2. **Sequenzielle Verarbeitung:** Hier besteht jede Merkmalsvariable aus einer Reihe von Werten in chronologischer Reihenfolge, wobei die Reihenfolge der Werte entscheidend bei der Verarbeitung ist. Dies ermöglicht die Erfassung von Trends und Mustern über die Zeit.

Da zustandsbehaftete Modelle wie LSTM auf die Reihenfolge der Daten angewiesen sind, ist die Berücksichtigung der zeitlichen Abfolge für deren Leistung entscheidend. Im Gegensatz dazu sind FFNN, CNN und XGBoost zustandslose Modelle, die keine explizite Abhängigkeit von der Datenreihenfolge haben.

Abbildung 14 einen Überblick über die verwendeten Modellarchitekturen und den gewählten Parametern.

Sequenziell			
Model	#Hidden Layers	#Units Per Layer	Aktivierungsfunktion
FFNN_0	2	24:24	ReLu
FFNN_1	3	32:32:32	ReLu
1D_CNN	3	24:24:24	ReLu
2D_CNN	5	32:32:32:32:32	ReLu
LSTM_0	2	24:24	ReLu
LSTM_1	3	32:32:32	ReLu
Tabellarisch			
FFNN_0	1	6	Sigmoid
FFNN_1	1	10	Sigmoid
FFNN_2	1	16	Sigmoid
FFNN_3	2	6:4	Sigmoid
FFNN_4	2	12:6	Sigmoid
FFNN_5	2	18:18	Sigmoid

Abbildung 14: Verwendete Modellarchitekturen, Anzahl der Hidden Layers und verwendete Aktivierungsfunktionen

Zusätzlich wurden die Trainingsdaten mit verschiedenen Zeitfensterlängen von 1 bis 12 (1 = 5 Min., 12 = 60 Minuten) verarbeitet, um die optimale Fensterlänge zu identifizieren. Dabei blieb die zeitliche Reihenfolge der Stichproben erhalten, und alle Stichproben sind mindestens einmal Teil des Trainings- oder Validierungssplits, mit Ausnahme des letzten Validierungssplits. Abbildung 15 stellt die mittleren F1-Cross-Validation-Werte für unterschiedliche Zeitfensterlängen dar. Eine Fensterlänge von sechs Zeitschritten (30 Minuten) zeigte eine gute Balance zwischen Leistung und Zeitauflösung und wurde daher für weitere Analysen gewählt. Längere Fensterlängen führten lediglich zu geringfügigen Verbesserungen.

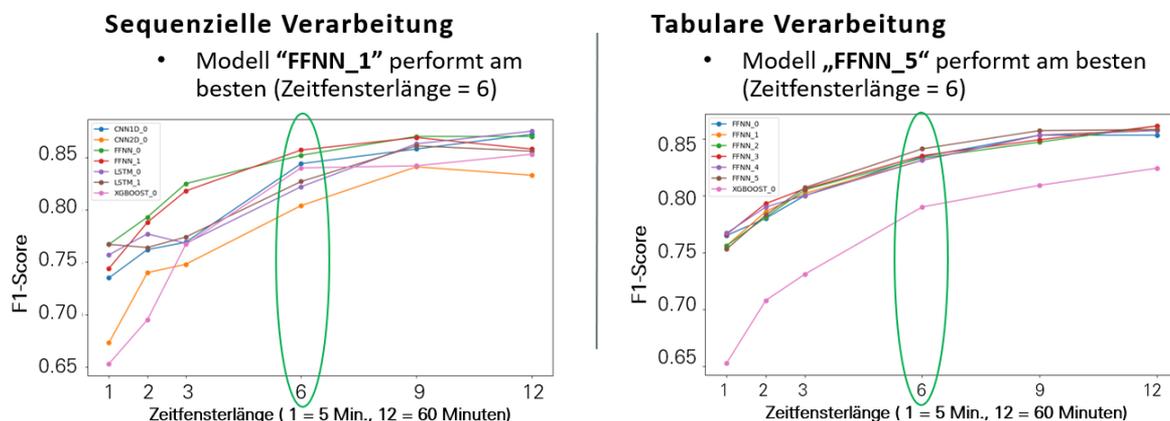


Abbildung 15: F1-Score für unterschiedliche Zeitfensterlängen der sequenziellen und tabularen Modelle

Abbildung 16 zeigt die Performanceevaluierung der tabellarischen und sequenziellen Modelle für die gewählte Zeitfensterlänge. Als Leistungsmetriken wurden die durchschnittliche Genauigkeit (Accuracy), der F1-Score (macro), der Recall (macro) und der Precision-Score (macro) über alle Splits hinweg erfasst.

Die Leistungsstreuung zwischen den sequenziellen Modellen war höher als bei den tabellarischen Modellen. Die tabellarischen FFNN-Modelle übertreffen dabei die tabellarischen XGBoost-Modelle deutlich. Das leistungsstärkste tabellarische Modell, FFNN_5, besteht aus zwei versteckten Schichten mit je 18 Neuronen. Die sequenziellen Modelle wiesen generell eine leicht höhere Recall-Rate als Precision auf. Umgekehrt hatten die tabellarischen Modelle eine höhere Precision als Recall. Insgesamt erreichten die sequenziellen Modelle leicht bessere Validierungswerte als die tabellarischen Modelle.

Model	Accuracy	Precision	Recall	F1
Tabular Models				
FFNN_0	0.849	0.840	0.837	0.834
FFNN_1	0.847	0.839	0.838	0.831
FFNN_2	0.848	0.842	0.836	0.833
FFNN_3	0.849	0.847	0.842	0.835
FFNN_4	0.846	0.845	0.839	0.831
FFNN_5	0.855	0.848	0.843	0.841
XGBoost	0.813	0.795	0.807	0.790
Sequential Models				
CNN1D	0.858	0.848	0.849	0.844
CNN2D	0.826	0.812	0.815	0.804
FFNN_0	0.867	0.856	0.864	0.852
FFNN_1	0.873	0.860	0.875	0.857
LSTM_0	0.839	0.829	0.838	0.822
LSTM_1	0.844	0.830	0.829	0.827
XGBoost	0.856	0.847	0.855	0.840

Abbildung 16: Leistungsmetriken der verwendete Modelle

Für das bestperformende „FFNN_5“-Modell wurde eine Feature-Importance-Analyse mithilfe von SHAP und XGBoost-Importance-Vektoren durchgeführt, um anschließend kleinere Modelle mit einem reduzierten Merkmalsdatensatz weiterzubearbeiten.

Abbildung 17 zeigt die SHAP-Werte der einzelnen Merkmale, die deren durchschnittlichen Einfluss auf die Modellvorhersagen beschreiben. Die wichtigsten Einflussfaktoren für das Modell waren:

- CO₂-Konzentration
- Geöffnete Türen
- Geräuschpegel (Durchschnittlicher Schalldruckpegel)
- Geräusch-Peaks (Maximaler Schalldruckpegel)

Basierend auf der SHAP-Analyse wurde festgestellt, dass die aktuelle Raumtemperatur, die aufgezeichneten Werte der Bewegungssensoren und relative Luftfeuchtigkeit den geringsten Einfluss auf die Modellvorhersagen haben. Demgegenüber zeigte eine Analyse der Featurerelevanz der XGBoost-Modelle, dass der maximale Schalldruckpegel das wichtigste Merkmal darstellt.

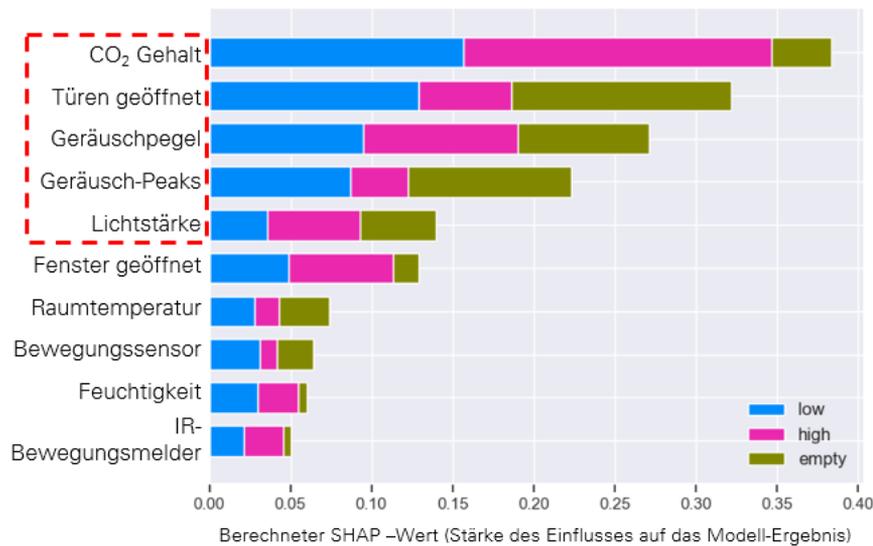


Abbildung 17: SHAP-Werte für das „FFNN_5“-Modell

5.1.1. Experimente zur Verringerung der Modellgrößen

Basierend auf diesen Erkenntnissen wurde ein reduzierter Merkmalsdatensatz, bestehend aus CO₂-Konzentration, Lichtintensität, Geräuschpegel und Geräuschpeaks sowie Türöffnungsstatus, ausgewählt, um die Modellgrößen noch weiter zu verringern.

Die Bewertungsmatrix (Abbildung 18) zeigt, dass in diesem Fall das Modell „FFNN_4“ die besten Ergebnisse hinsichtlich der relevanten Parameter liefert.

Model	Accuracy	Precision	Recall	F1
FFNN_0	0.848	0.847	0.845	0.832
FFNN_1	0.844	0.844	0.839	0.830
FFNN_2	0.851	0.849	0.846	0.837
FFNN_3	0.847	0.848	0.843	0.833
FFNN_4	0.851	0.851	0.849	0.837
FFNN_5	0.847	0.844	0.842	0.832
XGBoost	0.820	0.815	0.816	0.801

Abbildung 18: Bewertungsmatrix für die retrainierten Modelle mit reduziertem Featureset

Die Modellleistung hat sich trotz Eliminierung der Hälfte der ursprünglichen Merkmale nur geringfügig verändert. Precision und Recall haben sich sogar im Vergleich zu FFNN_5 mit allen Features verbessert.

5.1.2. Transfer Learning

Im nächsten Schritt wurde das Modell „FFNN_4“ für die Transfer-Learning-Experimente ausgewählt. Die Anordnung der Sensorik ist in Abbildung 19 dargestellt. Für die Experimente wurden erneut die Sensordaten in fünfminütigen Intervallen verwendet, um eine konsistente Datenbasis für das Training und die Evaluierung sicherzustellen. Die Erfassungszeiträume waren wie folgt definiert:

- Raum A: 12. September bis 20. Dezember 2023
- Raum B: 16. November bis 20. Dezember 2023

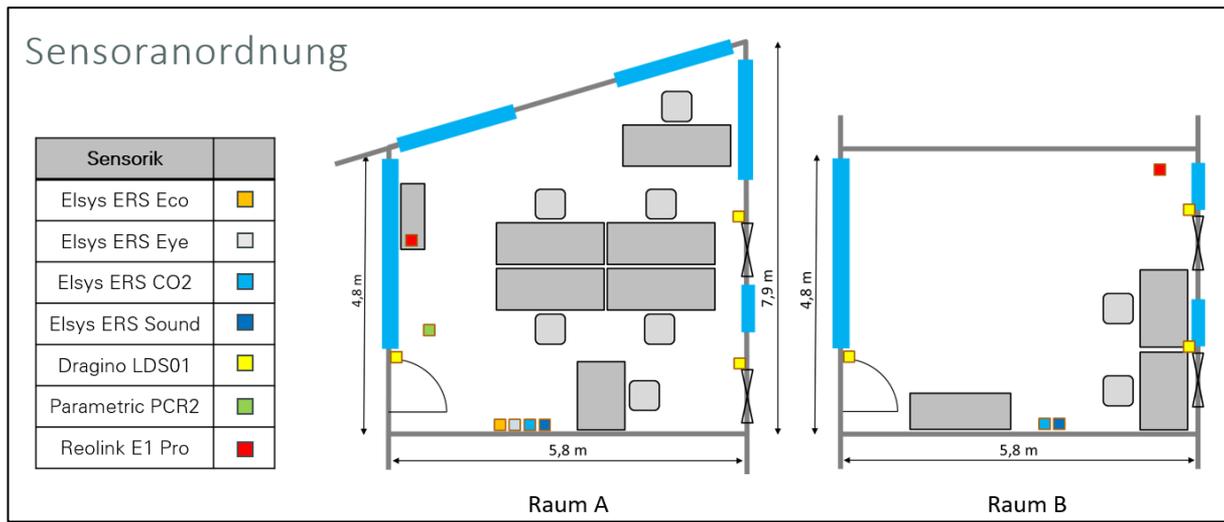


Abbildung 19: Anordnung der Innenraumsensorik in Raum A und Raum B

Das Modell wurde zunächst mit den Daten von Raum A trainiert (Lernrate $\eta_1 = 0,001$ für 20 Epochen) und anschließend dasselbe Modell auf Raum B übertragen (Lernrate $\eta_2 = (200^{-1}) * \eta_1$ für 30 Epochen). Eine kleinere Lernrate war erforderlich, um das Modell an die neue Umgebung anzupassen, während das zuvor erlernte Wissen erhalten blieb. Alle versteckten Schichten von FNN_4 sind dabei trainierbar.

Die Zusammenfassung des Transfer-Learning-Experiments ist in **Fehler! Verweisquelle konnte nicht gefunden werden.** visualisiert. Für Raum A erkannte das gewählte Modell (Zeitfensterlänge: 30 Minuten) die Belegungsstufen mit folgender Genauigkeit (F1-Score):

- **Leerer Raum:** 95 %
- **Geringe Belegung (1-2 Personen):** 41 %
- **Hohe Belegung (> 3 Personen):** 100 %

Nach der Anpassung des Modells durch Transfer Learning erkannte dasselbe Modell in Raum B folgende Belegungsstufen (F1-Score):

- **Leerer Raum:** 92 %
- **Geringe Belegung:** 86 %
- **Hohe Belegung:** 35 %

Obwohl die Stichprobe mit nur zwei Räumen sehr klein war, zeigen die Ergebnisse, dass Transfer Learning einen vielversprechenden Ansatz zur Skalierung von Belegungsmodellen darstellt.

5.2. Use Case 2 - Komfortbewertung- und Vorhersage

Für den zweiten Use Case wurde das ARX-Modell („AutoRegressive mit eXogenen Eingaben“) ausgewählt, da es eine klare, interpretierbare lineare Struktur bietet und gleichzeitig recheneffizient bleibt. Dieses Modell wurde in Python entwickelt und nutzt sowohl vergangene Raumtemperaturen (autoregressive Komponente) als auch externe Einflussfaktoren wie Außentemperatur und Globalstrahlung, um präzise 24-Stunden-Vorhersagen zu erstellen. Die gewählte Modellstruktur ermöglichte nicht nur eine effiziente Berechnung, sondern liefert auch wertvolle Einblicke in die Abhängigkeit der Raumtemperatur von externen Wetterbedingungen.

Um eine hohe Vorhersagegenauigkeit zu gewährleisten, setzt das Modell auf einen rollierenden Trainingsfenster-Ansatz. Dabei wird zunächst eine einstufige Vorhersage erstellt, die anschließend iterativ als autoregressiver Eingabewert für die nächste Prognose verwendet wird. Durch diese rekursive Strategie entsteht eine vollständige 24-Stunden-Prognose.

Dieser Ansatz stellte sicher, dass das Modell kontinuierlich an neue Daten angepasst wird. Nach jeder abgeschlossenen 24-Stunden-Prognose verschiebt sich das Trainingsfenster um eine Stunde nach vorne. Das Fenster umfasst stets die vergangenen 21 Tage, um eine ausreichende Datenhistorie bereitzustellen. Anschließend werden die Regressionskoeffizienten neu kalibriert, bevor die nächsten Vorhersagen berechnet werden. Durch diese rollierende Cross-Validation bleibt das Modell stets aktuell und kann flexibel auf Veränderungen in den Wetterbedingungen und Innenraumparametern reagieren.

Nach der Definition der Modellparameter wurde ein automatisierter Auswahlprozess gestartet, um die optimale Modellstruktur zu bestimmen. In der Statistik wird das Akaike-Informationskriterium (AIC) bevorzugt, da es sowohl die Vorhersagegenauigkeit bewertet als auch eine zu hohe Modellkomplexität bestraft. Das beste Modell wurde durch die Kombination von Eingangsvariablen identifiziert, die den niedrigsten AIC-Wert erzielte und somit optimal für die Vorhersage war.

Die maximale Verzögerung (Lag) der Prädiktorvariablen wurde auf 5 Zeitschritte festgelegt, was insgesamt 131.072 mögliche Modellkombinationen aus den 17 verfügbaren Eingangsparametern ergab. Zur schnellen Identifikation eines nahezu optimalen Modells wurde eine schrittweise rückwärtsgerichtete Regressionsauswahl angewandt, die darauf basiert, das AIC zu minimieren. Die verzögerten Eingaben mit dem niedrigsten AIC-Wert wurden automatisch für das endgültige Prognosemodell ausgewählt.

Nach der Validierung wurde das Modell auf einem ausgewählten Edge Device (Raspberry Pi) erprobt. Das Edge-Gerät wurde mit LoRaWAN-Sensoren und dem MeteoBlue-Wetterdienst über eine API integriert. Diese API sammelt Wetterdaten der vergangenen drei Wochen sowie Prognosen für die nächsten 24 Stunden und stellt die erforderlichen Eingabedaten für das Modell bereit, um präzise Vorhersagen zu erstellen. Zudem sorgt die API für eine korrekte Formatierung und Übertragung der Daten an das Edge-Gerät, sodass das Modell seine Vorhersagen kontinuierlich stündlich mit den neuesten Informationen aktualisieren kann.

Die Einsatzphase zeigte die praktische Anwendung des Modells für Echtzeit-Temperaturvorhersagen vor Ort und demonstrierte sein Potenzial für den kontinuierlichen Betrieb in freien (nicht klimatisierten) Büroumgebungen. Durch die Kombination von Echtzeit-Sensordaten mit lokalen Wetterprognosen konnte das Edge-Gerät genaue, zeitnahe Vorhersagen generieren, die Gebäudemanager dabei unterstützten, Überhitzung während Hitzewellen zu reduzieren. Zusätzlich wurden die Echtzeitdaten und Prognosen in einem benutzerfreundlichen Dashboard dargestellt, das von DiLT Analytics entwickelt wurde und eine einfache Überwachung und Entscheidungsfindung ermöglichte.

Die Validierungsphase umfasste Tests des Vorhersagemodells während einer Hitzewelle in Graz, um die Leistungsfähigkeit des Modells unter extremen Bedingungen zu bewerten. Das Trainingsfenster wurde auf 21 Tage festgelegt. Der initiale Trainingsdatensatz umfasste den Zeitraum vom 14. Juli bis 3. August 2024 und deckte verschiedene Wetterbedingungen vor der Hitzewelle ab. Anschließend generierte das Modell iterativ 24-Stunden-Prognosen über einen Zeitraum von zwei Wochen, vom 4. August bis zum 17. August 2024.

Zur **Bewertung der Modellgenauigkeit** wurden gängige Fehlermaße berechnet, darunter:

- **Mean Bias Error (MBE):** Bewertet den durchschnittlichen Vorhersagefehler.
- **Mean Absolute Error (MAE):** Gibt die durchschnittliche Fehlergröße an.
- **Root Mean Squared Error (RMSE):** Hebt größere Fehler hervor, indem diese quadriert werden.

Diese Fehlermaße wurden für verschiedene Prognosehorizonte berechnet – von 1 Stunde bis 24 Stunden im Voraus –, um eine detaillierte Analyse der Modellleistung zu ermöglichen.

Die Ergebnisse zeigten eine gute Modellgenauigkeit, wobei die Präzision mit zunehmendem Vorhersagehorizont leicht abnahm. Beispielsweise betrug der RMSE für die 1-Stunden-Vorhersage 0,158°C, während er für die 24-Stunden-Vorhersage auf 0,681°C anstieg. Diese Genauigkeit gilt als ausreichend, um den thermischen Komfort in Büroumgebungen zu steuern, insbesondere bei extremen Wetterbedingungen (siehe Abbildung 20).

Neben der Vorhersage von Raumtemperaturen wurde das Edge-Gerät so konzipiert, dass es für den nächsten Tag auch thermische Komfortzonen prognostiziert. Dabei werden die vorhergesagten Temperaturen den entsprechenden Komfortzonen zugeordnet. Diese Funktion liefert Gebäudemanagern praxisnahe Handlungsempfehlungen, um mögliche Unannehmlichkeiten frühzeitig zu erkennen und eine angenehme Raumumgebung zu erhalten.

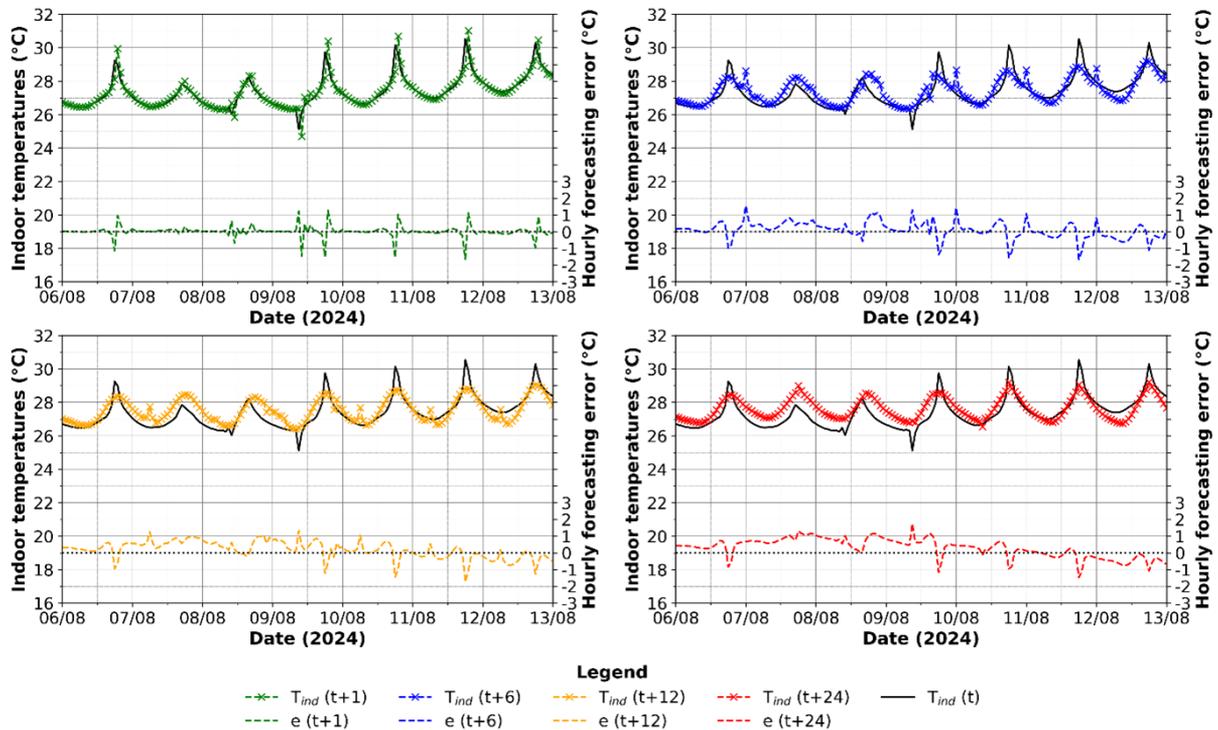


Abbildung 20: Darstellung der Modellgenauigkeit

5.3. Nutzerfreundliche Dashboards und Usability Tests

Zur Bewertung der Nutzerfreundlichkeit wurden unterschiedliche Varianten des Dashboards entwickelt sowie Usability-Tests durchgeführt. Sechs Teilnehmer mit unterschiedlichen Backgrounds (Gebäudetechnik, Sozialwissenschaft, Data Science) testeten die Anwendung und gaben Feedback zur Optimierung der Benutzeroberfläche.

Nach der Ausprogrammierung der finalen Frontend-Umgebung (Login-Bereich, Feedbacksystem, Datenvisualisierung – Beispiel siehe Abbildung 21) wurden Usability-Test unter Verwendung des „Think-Aloud-Protokolls“ durchgeführt. (Abbildung 21) Die 6 Teilnehmer:innen wurden gebeten, ihre Gedanken zu verbalisieren und die Workflows zu bewerten, während sie verschiedene Aufgaben am Dashboard erledigten. Ziel war es zu bewerten, wie gut das Dashboard Sensordaten (z. B. CO₂-Werte, Raumtemperatur, Luftfeuchtigkeit) darstellt und den Nutzer:innen ermöglicht, Feedback zu Komfort etc. zu geben. Verbesserungsvorschläge waren unter anderem das Hinzufügen von Erklärungen oder Tooltips zu Symbolen und Anzeigen im Dashboard für eine bessere Verständlichkeit der visualisierten Messdaten, erweiterte Diagrammfunktionen wie Zoomen oder Exportfunktionen der Messdaten.

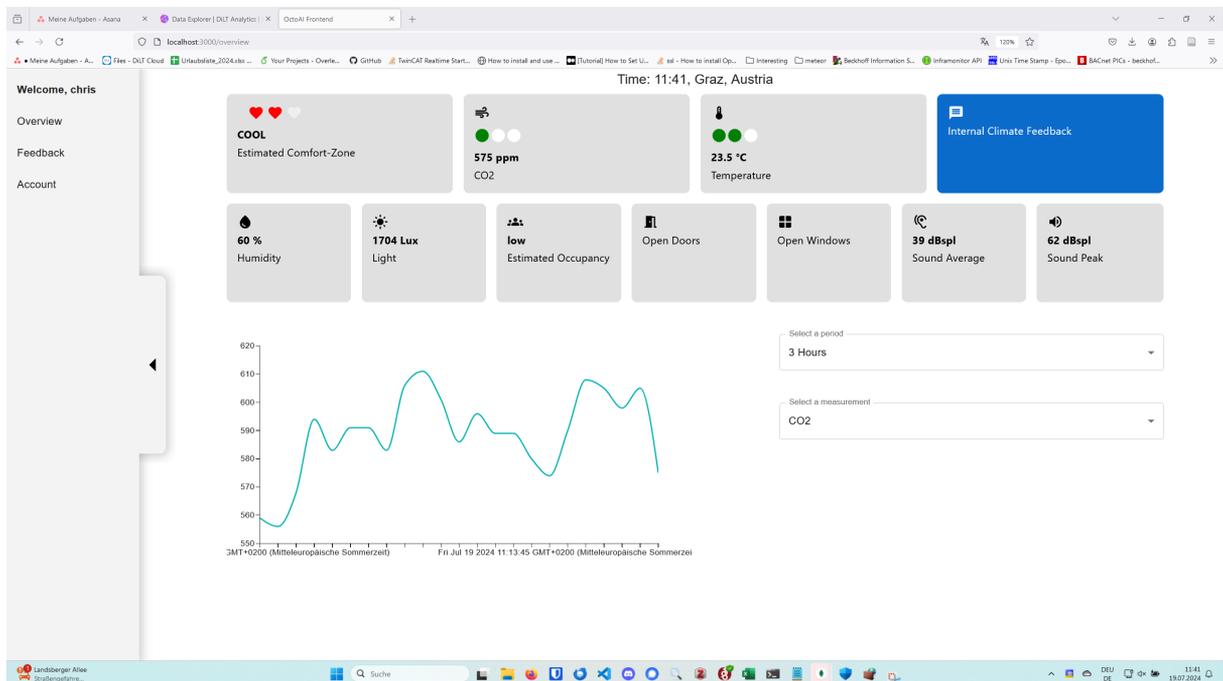


Abbildung 21: ausprogrammiertes Dashboard des Test-Set-Ups

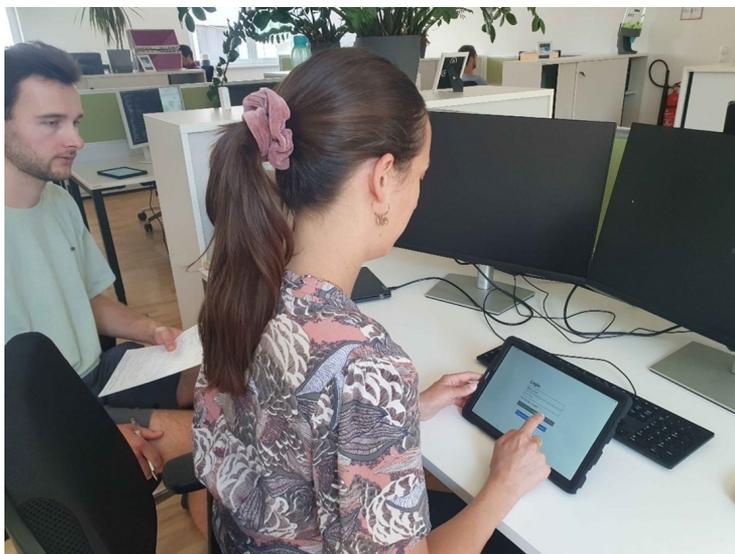


Abbildung 22: Beispielhafter Ablauf des Usability-Tests

6 Schlussfolgerungen

Das Projekt OctoAI hat gezeigt, dass Edge-KI ein erhebliches Potenzial für die energieeffiziente und datenschutzfreundliche Gebäudeautomation bietet. Die entwickelten Technologien sind skalierbar, ressourcenschonend und benutzerzentriert, wodurch sie sowohl für die Forschung als auch für die praktische Anwendung von großem Wert sind. Die gewonnenen Erkenntnisse werden in zukünftigen Projekten weiter genutzt, um die Lösungen zu optimieren und die Markteinführung voranzutreiben.

6.1. Belegungserkennung in Büros

Es wurde festgestellt, dass Modelle, die Eingaben in Form von Zeitsequenzen vergangener Daten nutzen, leicht bessere Ergebnisse erzielen als die entsprechenden tabellarischen Modelle, die nur Einzelwertrepräsentationen (Mittelwerte) vergangener Daten verwenden.

Zum Zeitpunkt der Sensorbeschaffung für die Nachrüstung in den Büros hat der Vergleich unterschiedlicher Sensortechnologien gezeigt, dass im Handel erhältliche LoRa-Sensorik anderen Technologien wie Bluetooth-basierten Geräten überlegen ist, insbesondere aufgrund der längeren Batterielaufzeit und größeren Reichweite. Bei den getesteten Edge-KI-Modellen haben sequentielle Modelle insgesamt bessere Ergebnisse als tabellarische Modelle gezeigt. Bei einer Zeitfensterlänge für Vorhersagen von 30 Minuten übertraf das FFNN-Modell sowohl CNN, LSTM als auch XGBoost. Dennoch wurde für eine minimale Büro-Lösung der tabellarische Ansatz gewählt, da synthetisches Upsampling und SHAP-Analysen standardmäßig nicht für sequentielle Datenformate unterstützt werden.

Die wichtigsten Merkmale für die Belegungserkennung wurden identifiziert und für eine minimale Büro-Lösung, z.B. in Demonstrationsprojekten, vorgeschlagen:

- CO₂-Konzentration
- Lichtintensität
- Durchschnittlicher Schalldruckpegel
- Maximaler Schalldruckpegel
- Anzahl geöffneter Türen

Die Experimente mit Transfer Learning haben gezeigt, dass dieser Ansatz eine vielversprechende Methode darstellt, um neuronale Netze an neue Umgebungen anzupassen. Insbesondere konnte durch den Einsatz von synthetischem Upsampling mittels SMOTE die Erkennungsrate der Minderheitsklasse in Raum B verbessert werden.

Allerdings war die Anzahl der verfügbaren Stichproben für diese Klasse zu gering, um das volle Potenzial von Transfer Learning umfassend zu bewerten. Dies zeigt, dass trotz der positiven Effekte auf die Modellanpassung zukünftig weitere Optimierungen und zusätzliche Datensätze erforderlich sind, um eine robuste Übertragbarkeit des Modells auf unterschiedliche räumliche Kontexte sicherzustellen.

6.2. Komfortzonen

Durch die Integration der Komfortzonen basierend auf der EN 16798-1-Norm bietet das System eine wertvolle Unterstützung zur Sicherstellung des Wohlbefindens der Gebäudenutzer, selbst unter extremen Wetterbedingungen.

Ein wesentlicher Vorteil dieses Ansatzes liegt in der nahtlosen Integration des Prognosemodells mit dem Edge-Gerät, den Raumklimasensoren und dem Wetterdienst, sodass Echtzeit-Vorhersagen direkt vor Ort erfolgen können. Die erfolgreiche Demonstration des Modells auf einem Raspberry Pi unterstreicht seine praktische Anwendbarkeit in realen Szenarien.

Dieses System stellt eine skalierbare und kosteneffiziente Lösung dar, die sich leicht in bestehende Gebäudemanagementsysteme integrieren lässt und somit einen wertvollen Beitrag zur Optimierung des Raumklimas leisten kann.

6.3. Zielgruppen und Relevanz der Projektergebnisse

Die Projektergebnisse sind für unterschiedliche Zielgruppen relevant bzw. bilden die Basis für unterschiedliche Entwicklungsrichtungen (Zusammenfassung siehe Tabelle 1).

Das Potenzial der Edge-AI für die Erkennung der Anzahl von Personen bzw. einer binären Klassifizierung von Anwesenheit (JA, NEIN) in Privathaushalten, Bürogebäuden oder sonstigen Gebäuden ist groß. Allerdings erfordert die Verallgemeinerbarkeit auf unterschiedliche Gebäudetypen und Raumnutzungen sowie technische Ausstattungsgrade spezifische Anpassungen. Die erfolgreiche Anwendung von Transfer-Learning-Methoden in OctoAI zeigt, dass Edge-AI-Modelle effektiv auf verschiedene Kontexte zugeschnitten werden können und die Erkennung von Belegungsmustern und energieeffizienten Lösungen ermöglichen.

Bei der Komfortvorhersage bestechen die Vorteile des verwendeten ARX-Modells (geringer Rechenaufwand, Echtzeitverarbeitungsfähigkeiten) und machen es sehr anpassungsfähig für verschiedene Anwendungen, insbesondere in Szenarien, in denen natürliche Belüftung und minimale mechanische Eingriffe im Vordergrund stehen. Das Modell eignet sich gut für Haushalte oder Bürogebäude mit einfachen Grundrissen und mit natürlicher Belüftung (z.B. Fensterlüftung). Durch die Einhaltung der EN 16798-1 ist diese Methode ideal für Büros, insbesondere für ältere Gebäude mit passiven Kühlsystemen.

Zielgruppen	Relevanz der Projektergebnisse
Bürogebäude	Optimierung der Gebäudesteuerung, Reduzierung von Energiekosten, Verbesserung des Arbeitskomforts
Wohngebäude und private Haushalte	Datenschutzfreundliche Hausautomatisierung, optimierte Klimasteuerung, Energieeffizienzsteigerung
Industrie & Produktionsstätten	Effizientes Flächenmanagement, Optimierung von Lüftung und Energieverbrauch, Reduktion von Betriebskosten
Smart Cities & urbane Infrastruktur	Skalierbare Nutzung für Verkehrs- und Umweltmonitoring, Integration in intelligente Stadtentwicklungsprojekte
Forschungs- & Entwicklungszentren	Reproduzierbare, offene Daten- und Modellansätze zur weiteren Entwicklung innovativer Lösungen

Tabelle 1: Kurzfassung Zielgruppen und Anwendungsbereiche

6.4. Bisherige Verwertungs- und Verbreitungsaktivitäten

Eine effektive Kommunikation spielte eine zentrale Rolle bei der Sensibilisierung für OctoAI und der Einbindung der Stakeholder im gesamten Projektverlauf. Dazu wurden ein Logo entworfen sowie eine Projektwebsite erstellt: <https://www.dilt.at/octoai/>.

Zur Erweiterung der Reichweite wurde ein professionelles Projektvideo produziert und auf LinkedIn veröffentlicht. Das Video vermittelt eine ansprechende Übersicht über die Projektziele und zentrale Erfolge, um die Sichtbarkeit zu erhöhen und eine breitere Zielgruppe anzusprechen.

Insbesondere LinkedIn wurde für regelmäßige Projektupdates genutzt (Abbildung 24) und die Projektergebnisse, insbesondere zum Thema „Transfer Learning“ beim „Alumni Tech Talk“ der Hochschule Burgenland im Jänner 2025 von DiLT Analytics einem breiten Publikum vorgestellt. (siehe Abbildung 23)

Projekt OctoAI: Neuronale Netze

Indirekte Anwesenheitserkennung aus Raummessdaten & Transfer Learning

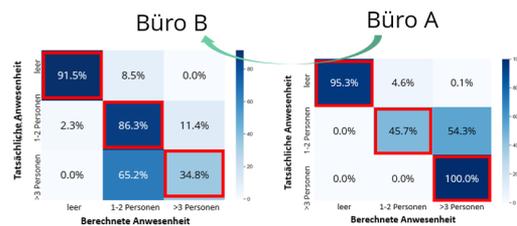
Messdaten: CO₂-Gehalt, Tür- und Fensterkontakte, Sound, Temperatur, Feuchtigkeit, Motion

Vergleich von unterschiedlichen **neuronalen Netzen** (tiny models, 2-5 hidden layers):

- Feed Forward Neural Networks
- Convolutional Neural Networks
- Long Short Term Memory
- XGBoost models



Deployment auf Edge Devices vor Ort



Source: Siegl, C., Hirsch, T., Kohl, T., Wotawa, F., & Schweiger, G. (2024). Neural network-based occupancy detection on the edge. In *Proceedings of BauSim Conference 2024, 10th Conference of IBPSA-Germany and Austria* (Vienna University of Technology, Wien, September 2024).

FFG OctoAI (FFG #893494)

Abbildung 23: Kurzzusammenfassung der Ergebnisse des Transfer-Learning Experiments (vorgestellt auf einer Veranstaltung im Jänner 2025)

Auf akademischer Seite präsentierte das Projektteam auf der BauSIM 2024 das Paper „Neural Network-Based Occupancy Detection on the Edge“. In diesem Beitrag wurden die im Projekt erarbeiteten Erkenntnisse, insbesondere die Leistung von neuronalen Netzwerkmodellen für die Erkennung von Belegung mittels Umgebungssensordaten und Edge-Geräte, vorgestellt. Zusätzlich wurden in einer Posterpräsentation die Ergebnisse der Umfrage zu Gebäudeautomationssystemen in Österreich zusammengefasst und Einblicke in die Barrieren und Möglichkeiten der Einführung gegeben. Die Veröffentlichung und das Poster sind unter <https://bausim2024.conf.tuwien.ac.at/> verfügbar:

- Siegl, C., Hirsch, T., Kohl, T., Wotawa, F. & Schweiger, G. (2024), 'Neural Network-based Occupancy Detection on the Edge', in *Proceedings of BauSim Conference 2024, 10th Conference of IBPSA-Germany and Austria, Vienna University of Technology, Wien, September 2024*
- Siegl, C., Hirsch, T., Kohl, T., Pertschy, R., Gustin, M., Robbi, S., Zehetmaier, F. & Schweiger, G. (2024) *Building Automation Systems in Austria: Findings from an Empirical Survey, in Companion Proceedings, BauSIM 2024, 10te Konferenz von IBPSA-DACH, TU Wien, Wien, September 2024*

Aktuell befindet sich eine weitere Publikation in Vorbereitung. Ein Abstract wurde für die Konferenz CLIMA 2025 in Mailand, Italien, eingereicht und bereits akzeptiert. Die Konferenz befasst sich mit dem Thema „Dekarbonisierte, gesunde und energiebewusste Gebäude in zukünftigen Klimazonen“.

Der zugehörige Konferenzbeitrag mit dem Titel „Real-Time Indoor Temperature Forecasting Using Edge Computing in Free-Running Offices“ wurde im Januar 2025 eingereicht. Diese Arbeit vertieft die

Erkenntnisse aus dem Projekt und hebt die Relevanz von Edge Computing für energieeffiziente Gebäude hervor.

Im Rahmen des Projekts wurde einer der weltweit größten offenen und qualitativ hochwertigen Datensätze für die Anwesenheitserkennung in Gebäuden erstellt. Dieser umfasst mehr als drei Monate an Innenraumsensordaten, ergänzt durch eine Ground Truth, die die tatsächliche Anwesenheit von Personen erfasst. Die Daten und Algorithmen wurden über Zenodo öffentlich zugänglich gemacht und leisten damit einen bedeutenden Beitrag zur Forschungsgemeinschaft. Die Veröffentlichung stellt eine wertvolle Ressource für die Entwicklung und Validierung von ML-Modellen zur Belegungserkennung dar. (siehe auch Siegl et al., 2024: 'Data and Sourcecode from: Neural Network-based Occupancy Detection on the Edge', Zenodo. URL: <https://doi.org/10.5281/zenodo.11573783>)

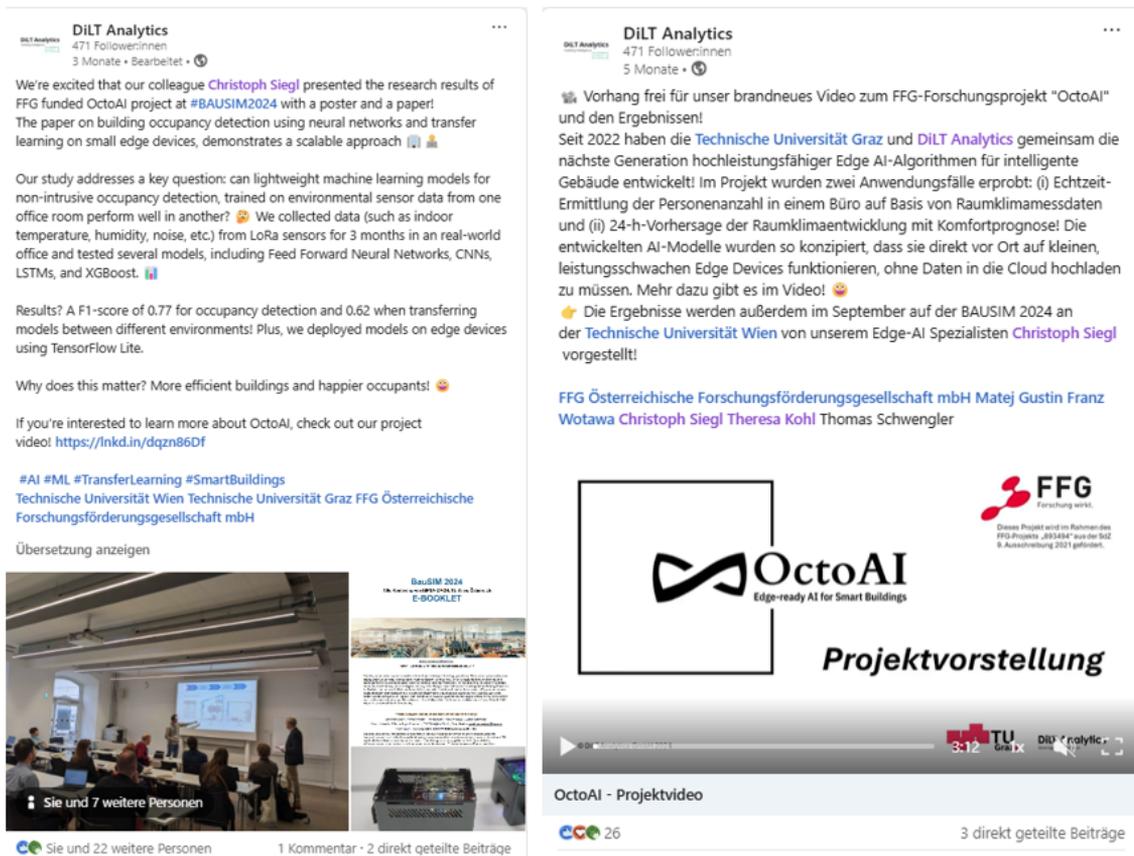


Abbildung 24: Beispielhafte Posts zum Projektfortschritt via LinkedIn

6.5. Marktpotenzial

Die Entwicklungen im Rahmen des Projekts OctoAI haben ein erhebliches Marktpotenzial, insbesondere im stark wachsenden globalen Smart-Building-Markt, der zwischen 2023 und 2030 mit einer durchschnittlichen jährlichen Wachstumsrate (CAGR) von 27 % expandieren wird. Dieser Markt wird maßgeblich durch technologische Fortschritte in den Bereichen Künstliche Intelligenz, Internet of Things, Cloud-Computing und Datenanalytik vorangetrieben und soll bis 2030 eine Marktgröße von 416 Milliarden Euro erreichen.

Der Bereich Energiemanagement nimmt innerhalb dieses Marktes eine immer größere Rolle ein und wird seinen Anteil bis 2030 voraussichtlich von 18 % auf 26 % steigern. Dies unterstreicht die zunehmende Bedeutung intelligenter Energie- und Gebäudemanagementsysteme für die Optimierung von Energieeffizienz und Betriebskosten. (Grand View Research, 2023)

Mit einem Marktanteil von rund 29 % (121 Milliarden Euro) ist Europa der zweitgrößte Markt nach Nordamerika. Besonders Nichtwohngebäude – darunter Büro- und Einzelhandelsflächen, Industriegebäude und andere gewerblich genutzte Objekte – stehen im Fokus, da sie den Markt für Energiemanagement-Softwarelösungen und -Dienstleistungen dominieren und auch künftig die treibende Kraft in diesem Sektor bleiben werden. (Coherent Market Insights, 2023)

Parallel dazu wächst der Building-Analytics-Markt, angetrieben durch die steigende Nachfrage nach energieeffizienten Lösungen und die Notwendigkeit zur Betriebskostenoptimierung. Bis 2030 wird dieser Markt mit einer CAGR von 15 % expandieren (Market Research, 2023), was die Relevanz intelligenter und datengetriebener Gebäudetechnologien weiter verstärkt.

7 Ausblick und Empfehlungen

Die entwickelten Edge-AI-Technologien bieten eine energieeffiziente, datenschutzfreundliche und intelligente Lösung für die Gebäudeautomation. Durch die Bereitstellung offener Daten und Methoden wird eine einfache Replikation und Anpassung in verschiedenen Anwendungsbereichen ermöglicht. Diese Ansätze bilden eine solide Grundlage für die breitere Implementierung in zukünftigen Smart Systems und leisten einen wertvollen Beitrag zur nachhaltigen Stadtentwicklung.

Ein vielversprechender Ansatz ist Transfer Learning, das die Entwicklung skalierbarer, datengetriebener Belegungserkennungssysteme ermöglicht. Diese Systeme bieten eine hohe Modellleistung, während sie gleichzeitig den Datenschutz der Nutzer gewährleisten. Ein Set an Basis-Modellen, optimiert für kleine, mittlere und große Räume, könnte flexibel an verschiedene Umgebungen angepasst und in unterschiedlichen Gebäudetypen eingesetzt werden.

Die lokale Ausführung auf Edge-Geräten reduziert den Netzwerkverkehr und verhindert die Übertragung sensibler Nutzerdaten in die Cloud. Gleichzeitig ermöglicht die drahtlose Sensortechnologie eine einfache Nachrüstung in bestehenden Gebäuden, sodass auch ältere Bauwerke von intelligenter Gebäudetechnologie profitieren können. Durch die direkte Anbindung an Gebäudemanagementsysteme (BEMS) lassen sich Energieverbrauch und -kosten optimieren, während der Nutzerkomfort in Nicht-Wohngebäuden sichergestellt wird.

Langfristig kann die Integration von Belegungserkennungssystemen zur Dekarbonisierung beitragen und eine nachhaltigere Zukunft unterstützen. Tabelle 2 zeigt eine Übersicht der Skalierungspotenziale der beiden Anwendungsfälle, ihre Übertragbarkeit auf verschiedene Gebäudetypen sowie die Herausforderungen für zukünftige Demonstrationsprojekte und weiterführende Forschungs- und Entwicklungsarbeiten.

	Use Case 1 – Belegungserkennung 	Use Case 2 – Komfortbewertung 
 Wohngebäude	Unterschiedliche Raumgrößen, Gebäudetypen und Nutzungsmuster machen es schwierig, Modelle, die in Büroumgebungen trainiert wurden, direkt zu übertragen. Haushalte haben im Vergleich zu Büros oft kleinere und weniger standardisierte Umgebungen, was zu einer Heterogenität der Daten führt.	
Multiplizierbarkeit	Transfer-Learning wurde im Projekt erfolgreich auf verschiedene Räume innerhalb eines Bürogebäudes angewendet, was eine gewisse Verallgemeinerbarkeit zeigt. Für Haushalte müssten die Modelle aufgrund der spezifischen Umgebungsbedingungen und Nutzungsmuster weiter angepasst werden.	Wohngebäude sind in der Regel auf natürliche Belüftung (Fensterlüftung) angewiesen, was sich gut mit dem Design des ARX-Modells deckt. Die geringen Rechenanforderungen des ARX-Modells machen es für den Einsatz auf kostengünstigen, verbraucherfreundlichen Geräten geeignet. Die Haushalte unterscheiden sich in Größe, Gebäudehüllqualität und technischer Ausstattung, was eine Anpassung des Prognosemodells erfordert.
Potential	IoT-Anwendungen wie smarte Thermostate, Beleuchtungssteuerung und Sicherheitssysteme könnten davon profitieren. Relevante Messwerte wie CO ₂ -Gehalt und Bewegungssensoren (z.B. basierend auf LoRaWAN-Protokoll) könnten leicht in Haushalte implementiert bzw. nachgerüstet werden.	Durch die Vorhersage der Innenraumtemperaturen können die Bewohner:innen die natürliche Belüftung und Beschattung optimieren, um den Bedarf an Klimaanlage zu minimieren. Die Vorhersagen ermöglichen es den Hausbewohnern, sich auf Temperaturschwankungen im Laufe des Tages einzustellen und entsprechende Gegenmaßnahmen zu ergreifen, sofern die technischen Möglichkeiten gegeben sind.
Herausforderungen	Datenschutzbedenken können mit Edge-AI entschärft werden, da die Verarbeitung lokal erfolgt. Die Erhöhung der Datengenauigkeit durch kalibrierte Sensoren auf kleinerem Raum ist jedoch entscheidend.	Nicht alle Haushalte haben Zugang zu genauen Wettervorhersagen wie der im Projekt verwendeten MeteoBlue API.
 Bürogebäude	Standardisierte Umgebungen mit größeren Räumen und klar definierten Nutzungszeiten. Sensordaten wie CO₂, Bewegung und Schalldruck sind besonders aussagekräftig für die Belegung. Ähnliche Sensordaten, wie z. B. Temperatur, Bewegung und CO₂, sind allgemein im Gewerbegebäudektor relevant.	
Multiplizierbarkeit	Edge-AI kann in Bürogebäuden effektiv eingesetzt werden. Die getesteten Modelle erreichten einen durchschnittlichen F1-Score von 0,77 und waren nach Anpassung durch Transferlernen auf einen zweiten Raum verallgemeinerbar. Einschränkungen wurden bei der Erkennung von Grenzfällen, wie z. B. einer mittleren Belegung, festgestellt.	Die Fähigkeit, mehrere externe Datenquellen wie Wetterdaten zu integrieren, ermöglicht eine breite Anwendbarkeit für verschiedene Gebäudetypen und Klimabedingungen. Der geringe Rechenaufwand des ARX-Modells ermöglicht Echtzeit-Vorhersagen vor Ort an abgelegenen oder ressourcenbeschränkten Industriestandorten.
Potential	Energieeinsparungen durch vorausschauende Steuerung von Heizung, Lüftung und Beleuchtung auf Basis der Anwesenheitserkennung aus schon vorhandenen Raummessdaten. Optimierung der Raumnutzung und des Arbeitsplatzkomforts durch Nutzung gängiger Messdaten sind zukünftige Entwicklungsfelder.	Das Modell entspricht der Norm EN 16798-1, was die Kompatibilität mit dem thermischen Komfortmanagement in österreichischen und anderen europäischen Bürogebäuden gewährleistet. Ein vorausschauendes „Management“ von sommerlicher Überwärmung kann die Abhängigkeit von mechanischen Kühlsystemen verringern und so den Energieverbrauch und die Betriebskosten senken. Die Vorhersagen können zur Anpassung industrieller Energiesysteme verwendet werden, z. B. zur Steuerung der Kühllast oder der Belüftung für die Sicherheit und den Komfort der Mitarbeiter:innen.
Herausforderungen	Die Skalierung auf größere Gebäude mit mehreren Stockwerken oder Gebäudekomplexen erfordert möglicherweise eine weitere Feinabstimmung. Die Synchronisierung und Verarbeitung von Daten über verteilte Sensornetze hinweg ist unerlässlich. Die in OctoAI getesteten Modelle wurden für kleinere, kontrollierte Umgebungen entwickelt. Anwendungen in groß angelegten industriellen oder öffentlichen Räumen erfordern möglicherweise komplexere Modelle und zusätzliche Sensoren.	Größere Bürogebäude mit unterschiedlichen Umgebungsbedingungen in verschiedenen Stockwerken oder Abschnitten erfordern möglicherweise mehrere, lokalisierte Modelle. Innere Lasten können die Innentemperaturen beeinflussen und erfordern eine sorgfältige Kalibrierung der Modelle. Stark schwankende Belegungsgrade in industriellen Umgebungen könnten die Vorhersagen zusätzlich erschweren. In Industriegebäuden gibt es möglicherweise besondere Wärmequellen (z. B. Maschinen), die das Modell berücksichtigen muss, was eine Anpassung erforderlich macht.

Tabelle 2: Übersicht zu Ausblick und Empfehlungen aus dem Projekt

Bevor eine flächendeckende Implementierung für die beiden Use Cases erfolgen kann, sind weitere Untersuchungen erforderlich bzw. offene Forschungsfragen zu klären:

- Größere Datensätze aus mehreren Räumen sollten analysiert werden, um die Praxistauglichkeit von Transfer Learning zu validieren.
- Bisher wurden Daten über 2-3 Monate hinweg gesammelt. Eine Erfassung über alle Jahreszeiten ist notwendig, um saisonale Schwankungen zu berücksichtigen.
- Einfluss bestimmter Merkmale je nach Jahreszeit sollte analysiert werden, da einige Faktoren möglicherweise nur in bestimmten Perioden relevant sind.
- Bei dem Use Case der Belegungserkennung sollte die Trennung zwischen niedriger und hoher Belegung weiter verbessert werden – dies könnte durch individuelles Hyperparameter-Tuning oder optimierte Datenverarbeitungstechniken erreicht werden.

Darüber hinaus bildeten Projektergebnisse und Entwicklungen von OctoAI die Basis für relevante Nachfolgeprojekte der Konsortialpartner. Ein Beispiel hierfür ist das Forschungsprojekt multiSENSE (FFG Nr. 904614), das am 1.1.2024 mit einem Konsortium der Technischen Universität Graz, Universität Graz und DiLT Analytics gestartet wurde:

Das Ziel von multiSENSE ist die Entwicklung und kleinmaßstäbliche Erprobung eines nachrüstbaren Multisensors, der auf chipbasierter Radartechnologie basiert und eine anonymen, berührungslosen Vitaldatenmessung ermöglicht. Dazu gehören die Erfassung der Herzfrequenz (HRV), Atemfrequenz und Belegungserkennung, kombiniert mit Sensoren für Innenraumqualität wie Temperatur, Luftfeuchtigkeit, CO₂ sowie Feinstaubpartikel (PM10 und PM2.5) mit hoher Zeitauflösung.

Ein weiteres zentrales Ziel ist die Konzeption eines Umwelt- und Gesundheitsinformationssystems für Einrichtungen des Gesundheitswesens darunter Krankenhäuser und Altenheime. Dabei sollen nachrüstbare Sensoren und Edge Devices in Gebäudeautomationssysteme integriert werden, um die Raumklimaüberwachung und das Vitaldaten-Monitoring zu optimieren.

Die Ergebnisse von OctoAI liefern dabei wertvolle Grundlagen für das Projekt multiSENSE, insbesondere in den Bereichen Raumklimasensorik, Zuverlässigkeit und Netzwerktechnologien, Belegungserkennung sowie KI-Modelle für Smart-Building-Anwendungen. Diese Forschungsansätze werden im Rahmen von multiSENSE weiterentwickelt, um neue Anwendungen im Bereich Gesundheitswesen und Gebäudetechnologie zu ermöglichen.

8 Verzeichnisse

Abbildungsverzeichnis

Abbildung 1: Grundstruktur eines neuronalen Netzwerks (links) und eines einzelnen Neurons (rechts), Chen et al (2020)	15
Abbildung 2: Stakeholder-Survey	18
Abbildung 3: publiziertes Poster mit den Umfrageergebnissen	18
Abbildung 4: verwendete Features für die OctoAI-KI-Modelle zur Belegungserkennung	20
Abbildung 5: Beispielhaftes Ergebnis der Komfortvorhersage für 24h im Voraus: es wird die Komfortzone „WARM“ mit den erwarteten Bedingungen für den 12. August 2024 um 18:00 Uhr prognostiziert	22
Abbildung 6: Raum A (DiLT Büro) mit diverser Sensorik - 3 unterschiedliche Elsys-Sensoren (Mitte), People Counter System (Links) und Kamera (rechts) zur Personenzählung	23
Abbildung 7: Tür- und Fensterkontakte der Fa. Dragina in Raum A (DiLT Büro)	23
Abbildung 8: Raum B mit installierter Sensorik und Kamera zur Untersuchung von Transfer-Learning	24
Abbildung 9: beispielhafte Kamerabilder in Büro A (links) und Büro B (rechts)	24
Abbildung 10: Prinzipskizze des implementierten Testset-Ups mit Datenerfassung und Edge-Devices	25
Abbildung 11: unterschiedliche Mock-Ups für das OctoAI-Dashboard	26
Abbildung 12: Getestetes Edge-Device zur Modellentwicklung und für das Modelldeployment	26
Abbildung 13: Vorgehensweise bei der Modellentwicklung	28
Abbildung 14: Verwendete Modellarchitekturen, Anzahl der Hidden Layers und verwendete Aktivierungsfunktionen	29
Abbildung 15: F1-Score für unterschiedliche Zeitfensterlängen der sequenziellen und tabularen Modelle	29
Abbildung 16: Leistungsmetriken der verwendete Modelle	30
Abbildung 17: SHAP-Werte für das „FFNN_5“-Modell	31
Abbildung 18: Bewertungsmatrix für die retrainierten Modelle mit reduziertem Featureset	31
Abbildung 19: Anordnung der Innenraumsensorik in Raum A und Raum B	32
Abbildung 20: Darstellung der Modellgenauigkeit	35
Abbildung 21: ausprogrammiertes Dashboard des Test-Set-Ups	36
Abbildung 22: Beispielhafter Ablauf des Usability-Tests	36
Abbildung 23: Kurzzusammenfassung der Ergebnisse des Transfer-Learning Experiments (vorgestellt auf einer Veranstaltung im Jänner 2025)	40
Abbildung 24: Beispielhafte Posts zum Projektfortschritt via LinkedIn	41

Tabellenverzeichnis

Tabelle 1: Kurzfassung Zielgruppen und Anwendungsbereiche	39
Tabelle 2: Übersicht zu Ausblick und Empfehlungen aus dem Projekt.....	43

Literaturverzeichnis

- Alouffi B., Hasnain M., Alharbi A., Alosaimi W., Alyami H., Ayaz M.: *A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies*. IEEE Access, 9, 2021, S. 57792–57807. doi: 10.1109/ACCESS.2021.3073203.
- Chen Y., Zheng B., Zhang Z., Wang Q., Shen C., Zhang Q.: *Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions*. ACM Comput. Surv., 53(4), 2020. doi: 10.1145/3398209.
- Coherent Market Insights: *Smart Building Market*. Report ID: CMI5733, 2023.
- Dekker M.A.C.: *Critical Cloud Computing - A CIIP Perspective on Cloud Computing Services*. European Union Agency for Cybersecurity, Report/Study, 2012. Online verfügbar unter: <https://www.enisa.europa.eu/publications/critical-cloud-computing> (abgerufen am 23. Mai 2023).
- Gala G., Castillo Rivera J., Fohler G.: *Work-in-Progress: Cloud Computing for Time-Triggered Safety-Critical Systems*. In: 2021 IEEE Real-Time Systems Symposium (RTSS), Dez. 2021, S. 516–519. doi: 10.1109/RTSS52674.2021.00054.
- Grand View Research: *Smart Building Market Size, Share, & Trends Analysis Report By Solution (Safety & Security Management, Energy Management), By Service (Consulting, Implementation), By End Use, By Region, And Segment Forecasts, 2023–2030*. Report ID: GVR-4-68039-964-, 2023.
- Jakovljevic M., Insaurralde C. C., Ademaj A.: *Embedded cloud computing for critical systems*. In: 2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC), Okt. 2014, S. 4A5-1-4A5-9. doi: 10.1109/DASC.2014.6979465.
- Janiesch C., Zscheck P., Heinrich K.: *Machine learning and deep learning*. Electronic Markets, 31(3), 2021, S. 685–695. doi: 10.1007/s12525-021-00475-2.
- Jin M., Bekiaris-Liberis N., Weekly K., Spanos C. J., Bayen A. M.: *Occupancy detection via environmental sensing*. IEEE Transactions on Automation Science and Engineering, 15(2), 2016, S. 443–455.
- Lundberg S. M., Lee S.-I.: *A Unified Approach to Interpreting Model Predictions*. In: Advances in Neural Information Processing Systems, 30. Curran Associates, Inc., 2017. Online verfügbar unter: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (abgerufen am 22. Dez. 2023).
- Lundberg S.: *Welcome to the SHAP Documentation — SHAP latest documentation*. 2018. Online verfügbar unter: <https://shap-lrjball.readthedocs.io/en/latest/index.html> (abgerufen am 16. Jan. 2024).
- MarketResearch: *Building Analytics Market - Industry Segment Outlook, Market Assessment, Competition Scenario, Trends, And Forecast 2023-2032*. Report ID: 24585, 2023.
- Rueda L., Agbossou K., Cardenas A., Henao N., Kelouwani S.: *A comprehensive review of approaches to building occupancy detection*. Building and Environment, 180, 2020, S. 106966.

- Schweiger G., Eckerstorfer L. V., Hafner I., Fleischhacker A., Radl J., Glock B., Wastian M., Rößler M., Lettner G., Popper N. et al.: *Active consumer participation in smart energy systems*. Energy and Buildings, 227, 2020, S. 110359.

Abkürzungsverzeichnis

ML	Machine Learning
CAGR	Compound Annual Growth Rate

