

Datalab – The ZHAW Data Science Laboratory



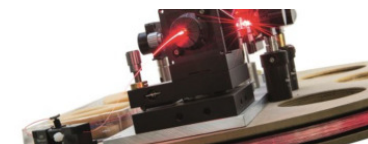


ZHAW Datalab

- Das **Datalab** ist eine **virtuelle Organization** an der ZHAW
- Gegründet **2013** vom InIT (Institut für Angewandte Informationstechnologie) und IDP (Institut für Datenanalyse und Prozessdesign), beide Departement T.
- Mitglieder sind aktuell **5 Institute** an **3 Departementen** sowie 4 assoziierte Institute
- Das Datalab bündelt Aktivitäten im Bereich **Data Science** von verschiedenen ForscherInnen an den jeweiligen Instituten
- Derzeit **59 Associates** auf Webseite gelistet
- Aktivitäten in **F&E, Lehre** und **Weiterbildung**

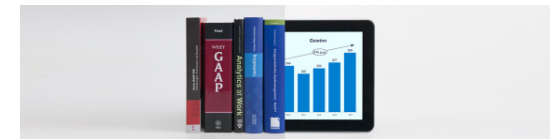
Beteiligte Institute

- **IDP**
Institut für Datenanalyse und Prozessdesign
Dept. T
- **InIT**
Insitut für Angewandte Informationstechnologie
Dept. T
- **ZSR**
Zentrum für Sozialrecht
Dept. W
- **IAS**
Institut für Angewandte Simulation,
Dept. N
- **IAMP**
Institut für Angewandte Mathematik und Physik
Dept. T



Assoziierte Institute

- **IDK**
Institut für Delinquenz und Kriminalprävention
Dept. S
- **IFI**
Insitut für Financial Management
Dept. W
- **InES**
Institute of Embedded Systems
Dept. T
- **IWI**
Institut für Wirtschaftsinformtaik
Dept. W



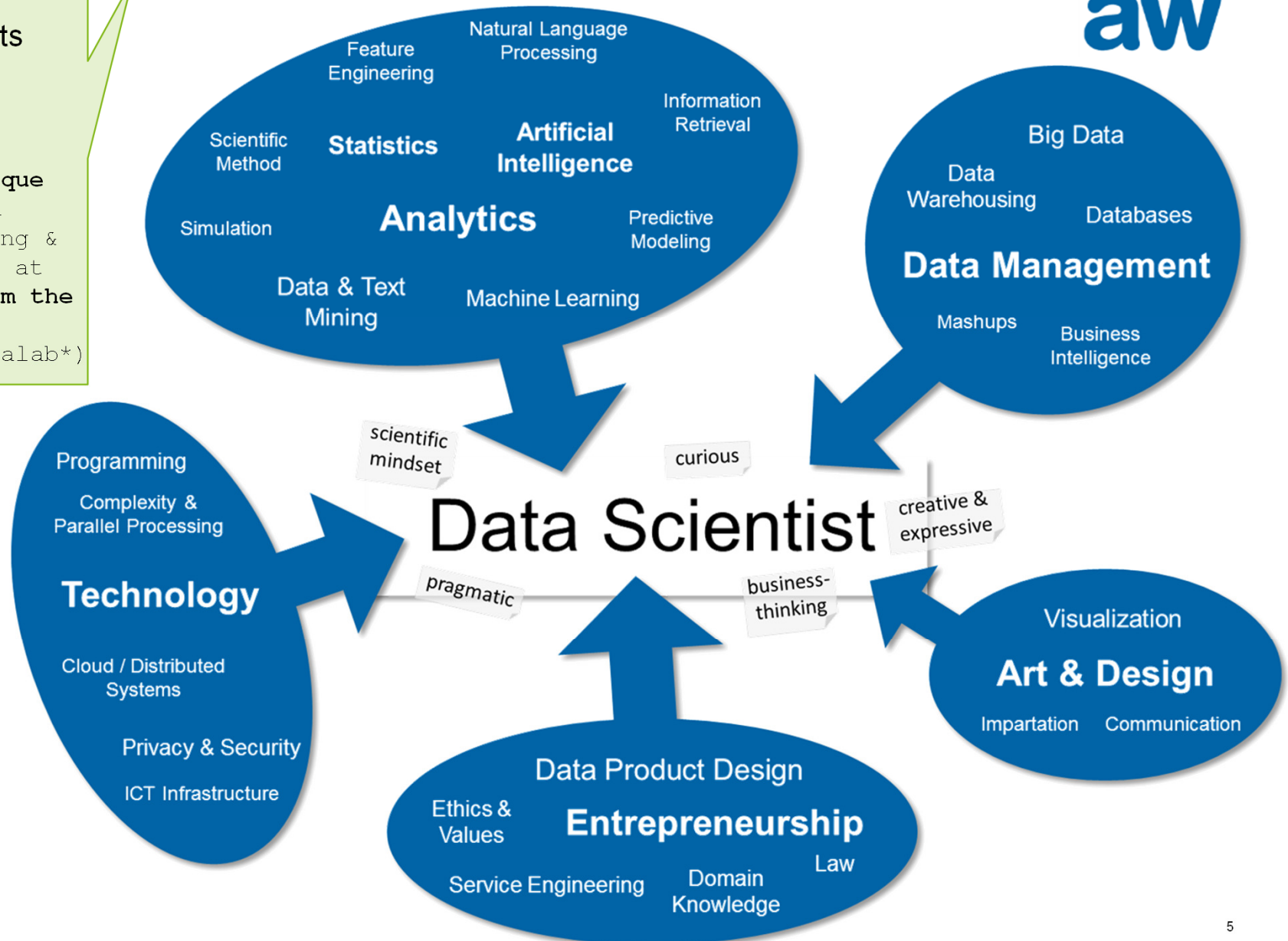
What is Data Science?

Enables Data Products

→ **Applied Science**

→ **Interdisciplinary**

Data Science := "Unique blend of skills from analytics, engineering & communication aiming at generating **value from the data** itself [...]"
(ZHAW Datalab*)



*) Stadelmann, Stockinger, Braschler, Cieliebak, Baudinot, Dürr and Ruckstuhl (2013). *Applied Data Science in Europe*. ECSS 2013, Amsterdam.

Research Agenda

- Database and Big Data technology
- Data Mining, Statistics and Predictive Modeling
- Machine Learning and Graph Analytics
- Information Retrieval and Natural Language Processing
- Business Intelligence and Visual Analytics
- Data Warehousing and Decision Support
- Communication and Visualization of Results
- Privacy, Security and Ethics
- Entrepreneurship and Data Product Design

Welche Rolle spielt die Statistik?

- Kombination von klassischen statistischen Methoden mit Ansätzen aus Machine Learning und AI
- Entwicklung von interpretierbaren, auf Sachwissen beruhenden Modellen
- Versuchdesign, Sampling
- Einbringen von Wissen über spezielle Datenprobleme (fehlende Werte, Ausreisser, stochastische Abhängigkeiten, zensierte Daten etc.)
- Vermeiden von Fehlschlüssen (Korrelation vs. Kausalität, Beobachtungsdaten vs. Experiment, Signifikanz vs. Effektstärke...)

Lehre

Neue Vertiefung “Data Science” im **MSE** (in der Informatik angesiedelt) mit vielen einschlägigen Modulen:

- Machine Learning
- Deep Learning
- Analysis of Text Data
- Analysis of Sequential Data
- Machine Learning in Computer Vision
- Predictive Modelling

Möglichkeiten zur **Promotion:**

- PhD Network in Data Science (ZHAW, SUPSI, UZH, UNINE)
- Weitere Kooperationen mit einzelnen Universitäten (Università Ca’Foscari, Venedig)

Weiterbildung

- Grosse Nachfrage nach **MAS “Data Science”** (4 von 5 CAS + Masterarbeit):

CAS Machine Intelligence

Machine Learning,
Deep Learning,
Text Analysis, Advanced topics
in Big Data

CAS Statistical Modeling

Information processing with R,
Advanced regression modeling,
Analysis of time to event data,
Network analysis

CAS Data Product Design

Data-specific Service Design,
Data-specific Business Models,
Practice workshop,
Security & Privacy

CAS Information Engineering

Scripting in Python,
Information Retrieval &
Text Analytics, Databases &
SQL, Data Warehousing,
Big Data

CAS Data Analytics

Data Description &
Visualization, Statistical
Foundations of Analytics,
Multiple Regression,
Time Series & Forecasting,
Clustering & Classification

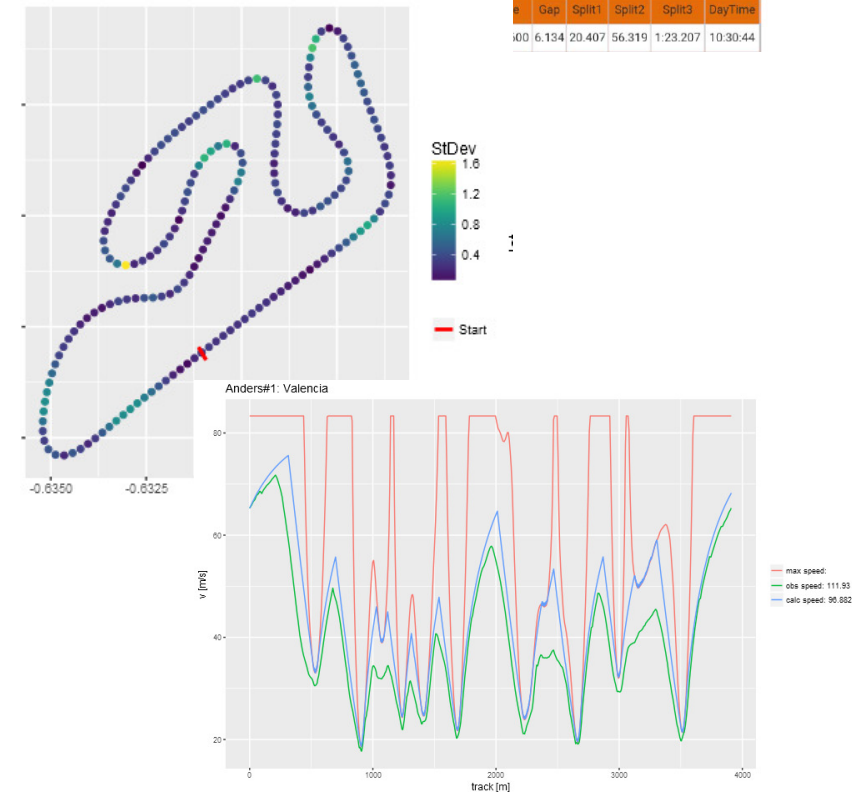
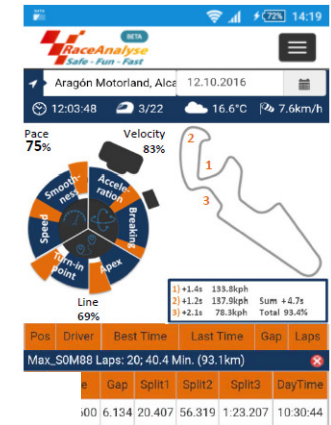
Weitere Aktivitäten

- **Datalab-Seminar** (ehemals “Brown-Bag-Seminar”) zum Austausch über Software, Methoden, Projekte und andere Data-Science-Themen (organisiert von Beate Sick, IDP)
- Organisation der **Swiss Conference on Data Science** 2014-2016 (seit 2017 organisiert von der Swiss Alliance for Data-Intensive Services)
- Konferenzen **Swiss Text** (jährlich seit 2016)
- **Deep Learning Day** 2016 - 2018 (14.9. in Winterthur), organisiert von Beate Sick (IDP), Oliver Dürr (ehem. IDP) und Elvis Murina (IDP)
- Nationales Thematisches Netzwerk **Swiss Alliance for Data-Intensive Services**
- **Buchprojekt** (Lead Thilo Stadelmann): Applied Data Science - Lessons Learned for the Data-Driven Business
- Blog: **Data Science Made In Switzerland** auf der ZHAW-Webseite



PROJEKTBEISPIELE

Virtueller Trainer für den Motorradsport

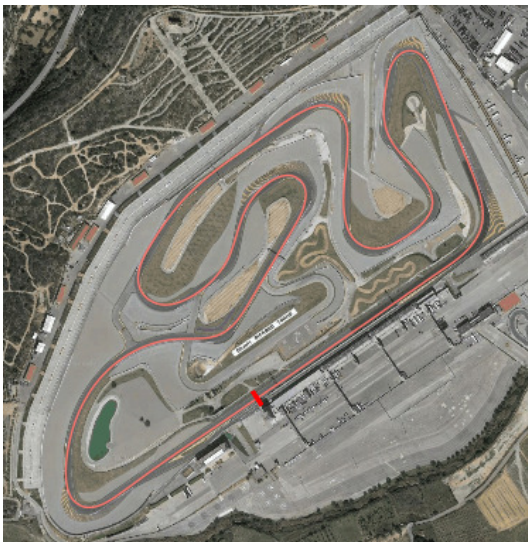


Situation: RaceAnalyse hat ein GPS-Device entwickelt, das das **Fahrverhalten im Amateur- und Profi Motorradrennsport** aufzeichnet und online in eine Cloud zu überträgt, wo die Daten für eine Webapplikation aufbereitet werden.

Ziele: Entwicklung von Indices: wie nahe kommt ein Fahrer der «perfekten Fahrt» (z.B. Fahrlinie, Schräglage oder Beschleunigung).

Involvierte Institute: IDP: fret, gome, rkst, gorn, heit; InIT; ZSN.
Kontakt: Martin Frey

Businesspartner: RaceAnalyse AG (www.raceanalyse.com)



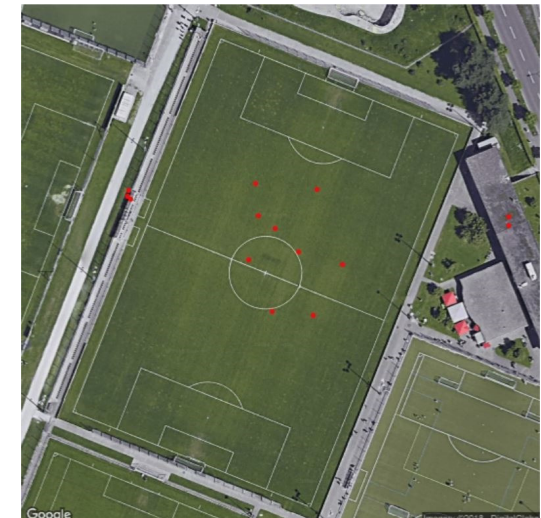
Algorithmen zur Analyse von Fussballspielern und Spielsituationen anhand von Bewegungsdaten

Situation: TRACKTICS hat das erste **Tracking-System für Fussballteams** entwickelt, das auch für ambitionierte Amateur-Mannschaften erschwinglich ist.

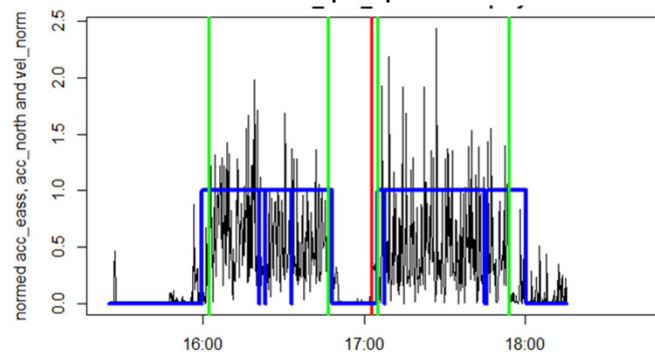
- Ziele:**
- Verbesserung der Messgenauigkeit und Entwicklung neuer Produktfeatures und **Auswertungsmetriken**.
 - Spezifische fussballerische Aktionen werden mittels **Machine Learning** aus Beschleunigungsdaten ermittelt.
 - **Automatisierung** von taktischen Analysen
 - neue **Kennzahlen** für Leistungsfähigkeit, Agilität und Antizipationsfähigkeit

Involvierte Institute: IDP: fret, dtli, murl, roha; ZSN. *Kontakt: Martin Frey*

Businesspartner: Tracktics GmbH (www.tracktics.com)

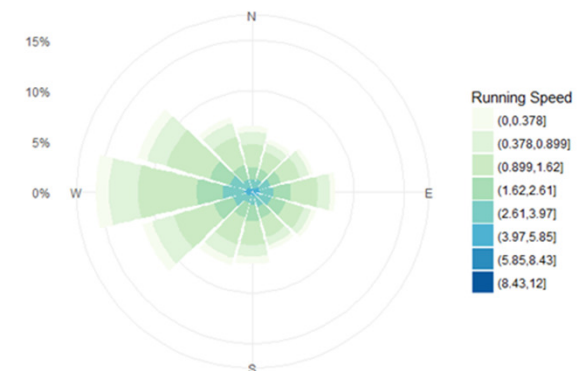


Bewegungsindex zur automatischen Detektion der Spielphasen



Mittlere Blickrichtung der Mannschaft zur Detektion der Spielrichtung

Ganzes Team während 2 HZ



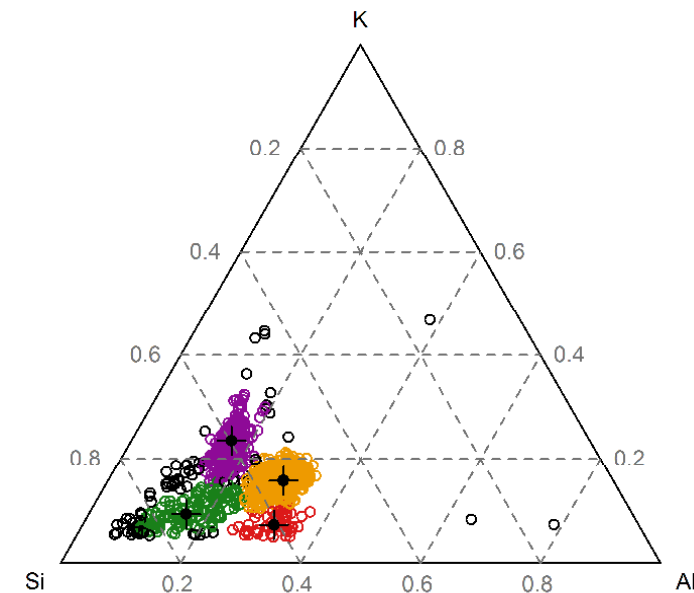
Partikelklassifikator

Situation: Mittels Elektronenmikroskopie können einzelne Partikel im μm -Bereich chemisch und morphologisch untersucht werden. Der Aufwand für manuelle Analyse der gewonnenen Daten ist dabei sehr hoch.

- Ziele:**
- **Einteilung** bislang gemessener Partikel aus verschiedenen Proben **in interpretierbare Klassen** (Training)
 - Berücksichtigung der speziellen **Datenstruktur** (compositional data)
 - Automatische **Klassifikation** neuer Partikel in die im Training festgelegten Klassen
 - Bisher **nicht bekannte Partikeltypen** werden als „unbekannt“ markiert

Involvierte Institute: IDP: lore, mild, rkst. *Kontakt: René Locher*

Businesspartner: Particle Vision GmbH (www.particle-vision.ch)



Cleverpendeln

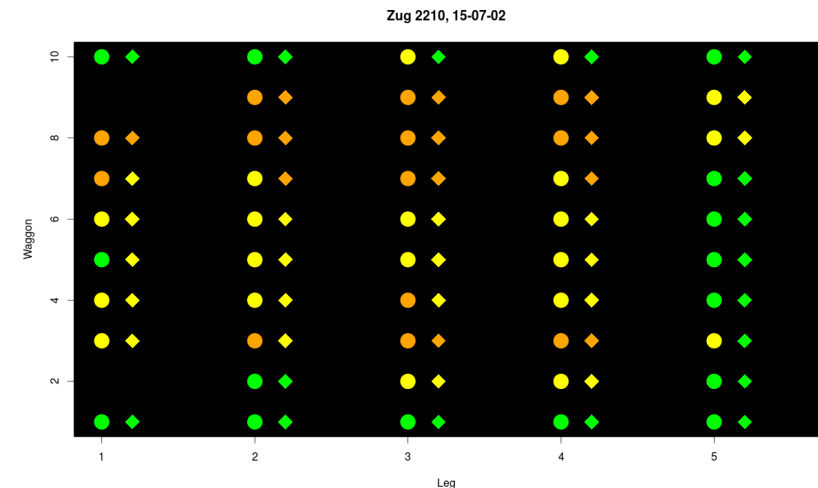
Situation: Es gibt (z.B. auf der SBB-Webseite oder -App) Auslastungsprognosen für Züge, jedoch nur für den gesamten Zug bzw. 1. und 2. Klasse.

Ziele:

- Wagengenaue Auslastungsprognosen für Nahverkehrs- und Fernzüge auf Basis verschiedener Daten (Zählungen durch Zugpersonal, automatische Türzählungen, Simulationen)
- Kombination verschiedener Prognosen

Involvierte Institute: IDP: gome, sine, mild, hofc.
Kontakt: Claudio Gomez

Businesspartner: one2seven (Shrebo)



Comprehensive Sales Forecasting

Projekt: KTI, 2014-2017






Ziele: Software-Lösung für Absatzprognosen und –Planung insbesondere für den Lebensmittelbereich (Verkauf und Produktion). Die Lösung vereint eine Bibliothek von ML- und Zeitreihenalgorithmen in einem Ensemble-Ansatz mit menschlichem Vorwissen (Priors). Das System „Maschine+Mensch“ wird als Einheit aufgefasst und die Verständlichkeit und Erklärbarkeit der Prognosen steht im Zentrum.

Involvierte Institute: IAS, IWI, IMM, (InIT, IDP)

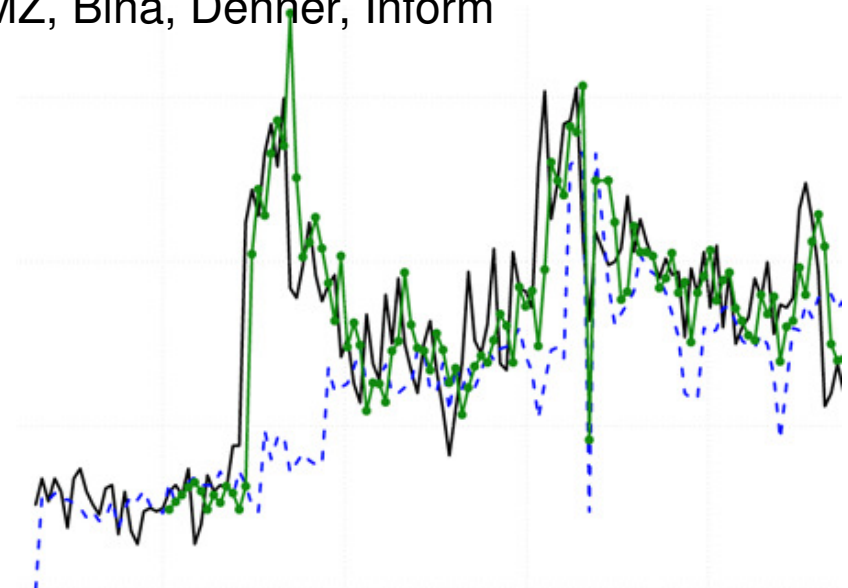
Businesspartner: Prognosix AG (ZHAW Spin-off), GMZ, Bina, Denner, Inform

Ware suchen

alle (6) tief (3) mittel (1) hoch (1)

Abbildung	Waren	Warengruppe	Menge	Warenwert	Relevanz
	Bananen	Südfrüchte	310kg	CHF 842	tief ●
	Kiwis	Südfrüchte	27kg	CHF 96	mittel ● ✓
	Orangen	Zitrusfrüchte	76kg	CHF 145	tief ●
	Papayas	Exotische Früchte	14kg	CHF 55	hoch ● ⚠
	Zitronen	Zitrusfrüchte	22kg	CHF 8	hoch ● ⚠

- Prognose rückblickend: Mittel ⓘ
- Prognose vorausschauend: Schlecht ⓘ
- Umsatzrisiko: Gering ⓘ



Dayzzi-Recommender: Next Generation

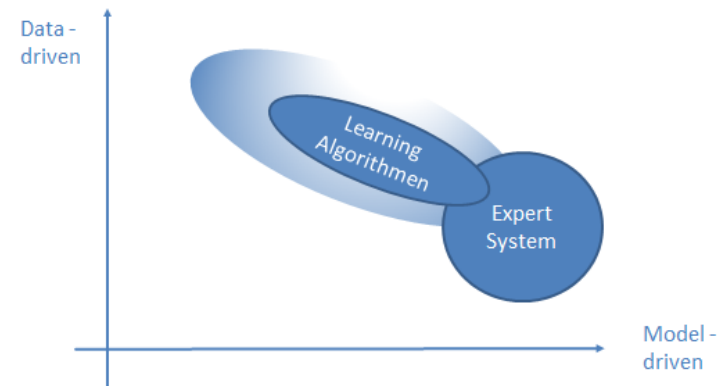
Projekt: KTI, 2016-2018

Ziele: Das Online-Experten-System von Dayzzi für das B2B-Geschäft wird in diesem KTI-Projekt zu einem neuartigen System weiterentwickelt, das die Stärken unterschiedlicher Empfehlungsalgorithmen kombiniert. Die Annäherung an ein reales Beratungsgespräch in Kombination mit hinterlegtem Expertenwissen und lernfähigen Algorithmen ermöglicht kundenspezifische, transparente Empfehlungen und sichert das Vertrauen in das Beratungssystem. Technologischer Kern der Lösung ist ein neuartiges hybrides Empfehlungssystem, das ein Expertensystem mit einem lernfähigen neuronalen Modul kombiniert und Online-Lernen ermöglicht.



Involvierte Institute: IAS

Businesspartner: Dayzzi AG



Deep Learning zur Erkennung von neuartigen Zelltypen

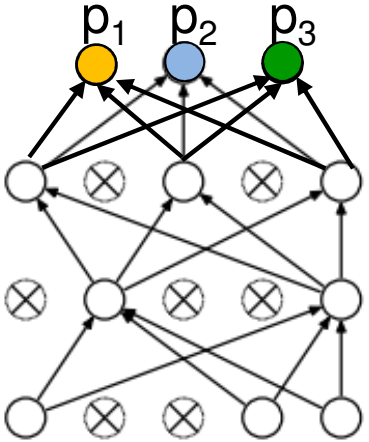
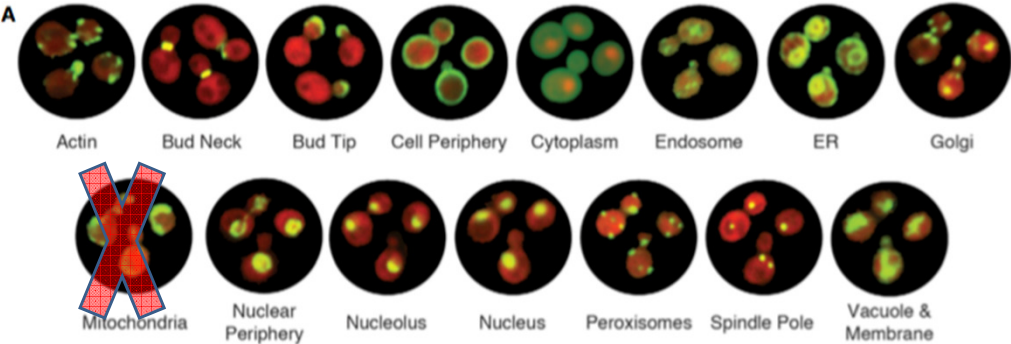
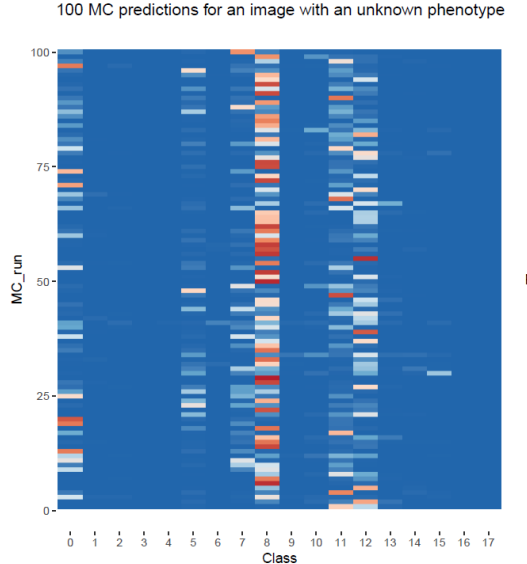
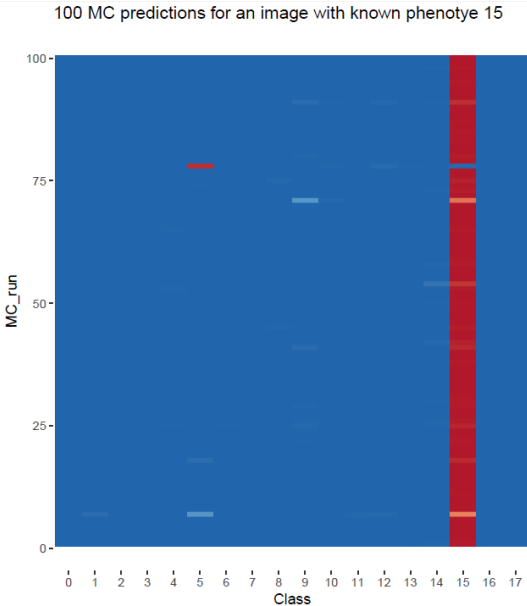


Situation: - Zur Klassifikation von Zelltypen in einem Pharmakologischen Screen wurde ein CNN-Modell auf den bekannten Zelltypen trainiert. Wenn das Bild eines neuen Zelltyps klassifiziert wird, so wird es einer der bekannten Klassen zugeteilt und oft sogar ohne dass die Probability über alle möglichen Klassen verschmiert ist.

Ziele: - Durch mehrfache Klassifikation mit MC-Dropout Varianten des CNNs kann durch die fehlende Stabilität der Vorhersagen erkannt werden, wenn ein neuartiger Zelltyp vorliegt.

Involvierte Institute: IDP: xdui, sick, murl, tolk. *Kontakt: Oliver Dürr*

Businesspartner: GeneData AG



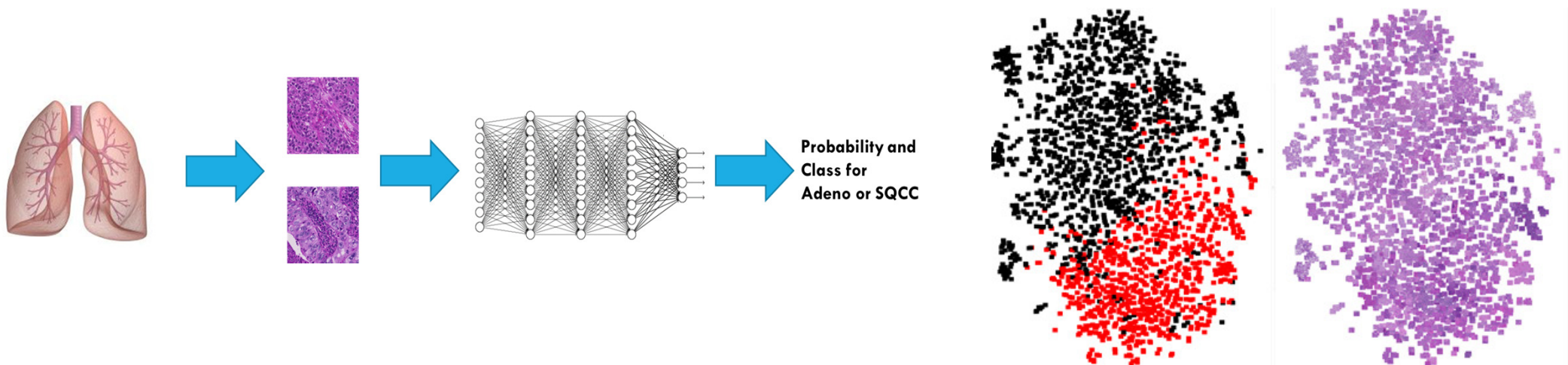
Deep Learning zur Klassifikation von Lungenkrebs

Situation: - Bei Lungenkrebs gibt es verschiedene Typen – die häufigsten sind Adeno und Squamous. Die **Erkennung des Typs anhand von histologischen Schnitten bestimmt die Behandlung.**

Ziele: - Entwicklung eines DL Tools, das im USZ von Pathologen eingesetzt wird, um ein Klassifikationsmodell zu trainieren und auf histologische Bilder anzuwenden.

Involvierte Institute: IDP: murl, casv, sick, xdui. *Kontakt: Beate Sick*

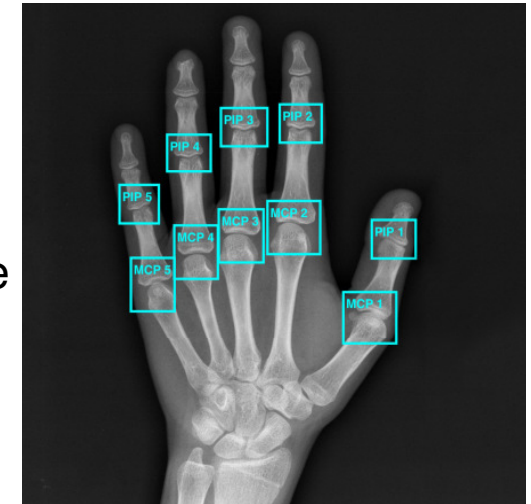
Businesspartner: Universitätsspital Zürich (www.usz.ch)



Arthritis-Scoring von Gelenken mittels Deep Learning

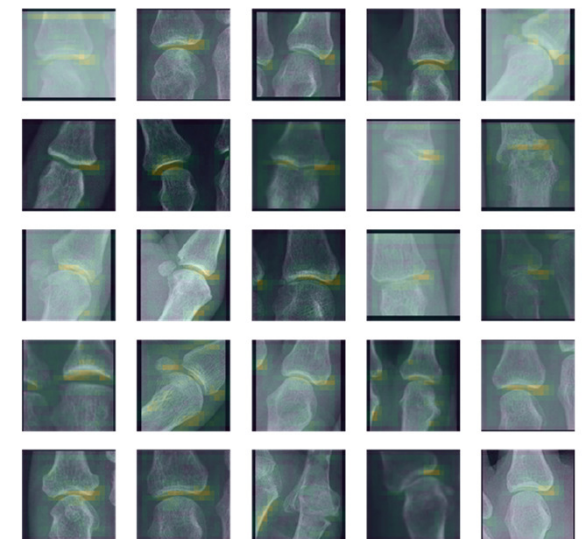
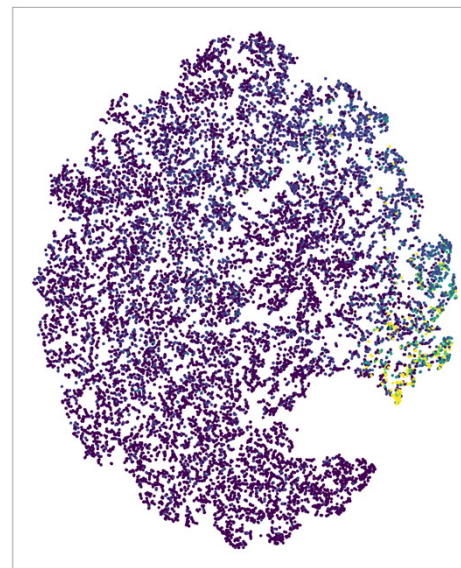
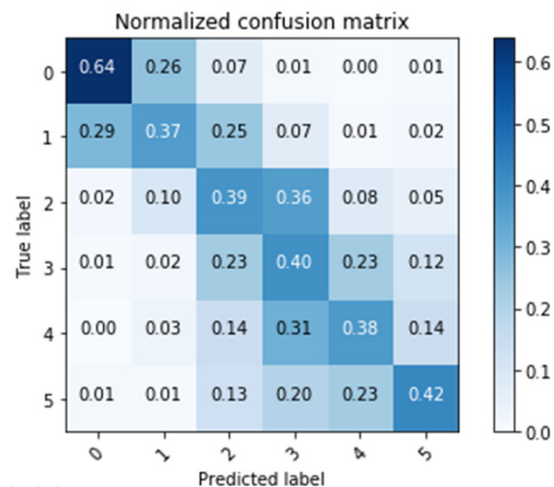
Situation: - Die Korrosion von Gelenken in Röntgenaufnahmen von Händen ist ein wichtiger Indikator für Arthritis. Die meisten Gelenke sind gesund, d.h. der Datensatz ist extrem unbalanciert.

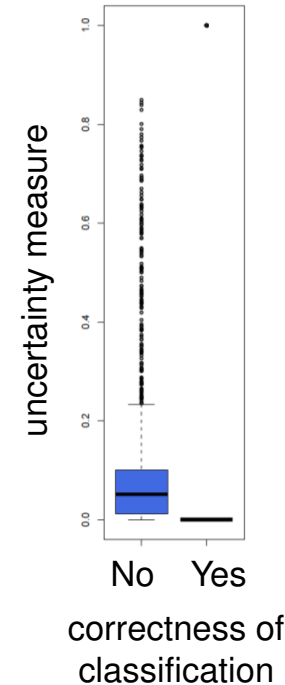
Ziele: - Ein Modell soll anhand von Bildern dasselbe Scoring der Gelenke vornehmen wie ein erfahrener Arzt. Es soll ausserdem visualisiert werden, wo welche Informationen im Bild wichtig für die Entscheidung waren. Es sollen ausserdem Gelenke identifiziert werden, die eventuell falsch gelabelt waren.



Involvierte Institute: IDP: roha, xdui, sick. *Kontakt: Beate Sick*

Businesspartner: Seantis GMBH





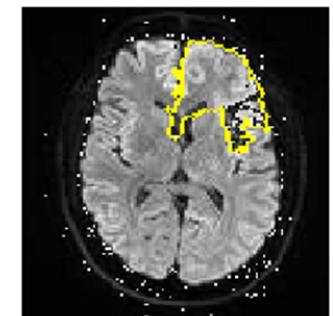
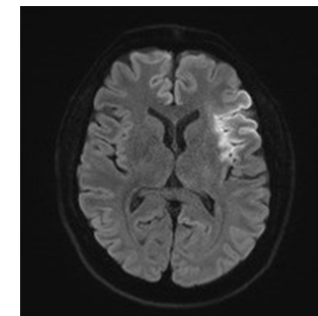
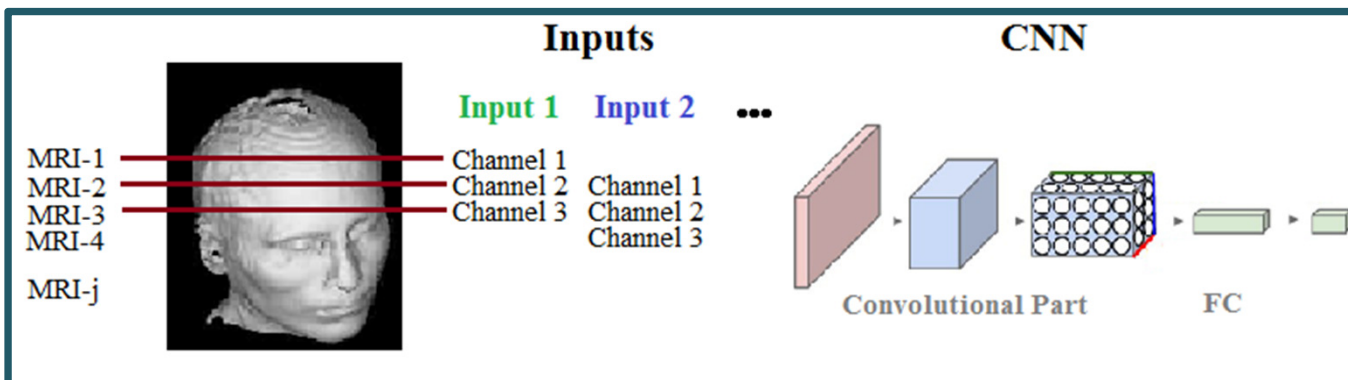
Einstufung von möglichen Schlaganfallpatienten anhand 3D MRT Bildern mittels Deep-Learning-Modellen

Situation: - Es ist wichtig, dass Schlaganfallpatienten schnell behandelt werden. Dafür ist eine schnelle Beurteilung von 3D-MRT-Bildern ein wichtiger Schritt

Ziele: -Wir wollen anhand von MRT-Bildern erkennen, ob es sich um einen Schlaganfall handelt und welche Behandlungsmethoden für den Patienten am Erfolgversprechendsten ist. Ausserdem soll gezeigt werden, warum die Entscheidung getroffen wurde und mit welcher Sicherheit (Bayesian-Dropout-Methode).

Involvierte Institute: IDP: hezo, murl, xdui, sick. *Kontakt: Beate Sick*

Businesspartner: Unispital Zürich, Susanne Wegener, SNF Projekt





data1ab

www.zhaw.ch/data1ab