

## Skin Cancer Detection

(28. November 2017)

### Beschreibung

Hautkrebs ist eine der am häufigsten vorkommenden Krebsarten [1]. Eine frühe Erkennung und Unterscheidung zwischen einem malignen oder benignem Hautkrebs ist hierbei von großer Wichtigkeit, da eine Behandlung in einem spätem Stadium sehr viel schwieriger ist. Da diese Unterscheidung aber selbst für Ärzte nicht immer einfach ist, wurde 2017 von Esteva et. al ein Maschinelles Lernverfahren zur Erkennung von Hautkrebs vorgestellt [2]. Auf 129.450 klinischen Bildern trainierten sie ein *Deep convolutional neural networks* (CNNs) [3], das schließlich in etwa gleich gut zwischen maligne und benigne unterscheiden konnte.

In dieser Arbeit werden wir auf Grundlage des Papers von Esteva et. al ein Verfahren entwickeln, das schließlich zwischen benignem und malignem Hautkrebs unterscheiden und einfach in einer Android App verpackt werden kann. Die App kann natürlich keine ärztliche Diagnose ersetzen, jedoch durch einen gut eingestellten Klassifikator eine Hilfe sein Hautkrebs früh zu erkennen und zum Arzt zu gehen.

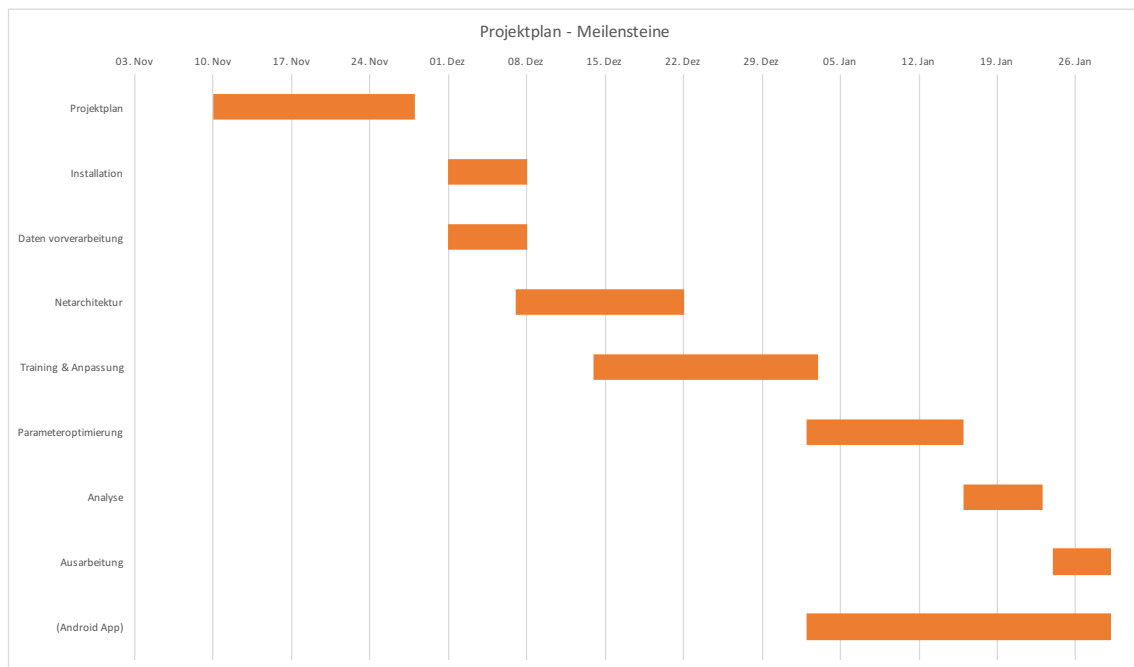


Abbildung 1: Meilensteine des Projekts als GANTT-Chart

## Tools

### Daten

Die Autoren des Papers nutzen in ihrer Arbeit einen Datensatz von insgesamt 129.450 klinischen Bildern, die 2.032 verschiedene Krankheiten abbilden. All diese Krankheiten wurden in drei Unterklassen unterteilt: maligne, benigne und nicht-neoplastische, also nicht-tumoröse Krankheiten bzw. Läsionen. Aus diesem gesamten Datensatz wurden 127.463 Bilder für das Training und die Validierung genutzt und die restlichen 1.942 Bilder wurden zum Testen gebraucht. Die Daten stammten aus 18 verschiedenen Quellen, unter anderem aus dem ISIC Dermoscopic Archive [4], welches ein freier Datensatz ist, der durch digitale Hautbilder dazu dienen soll die Sterblichkeitsrate bei Melanomen zu reduzieren. In unserer Ausarbeitung des Projekts werden wir unsere Daten ausschließlich aus diesem Datensatz beziehen, da sie frei erhältlich und für unsere Zwecke ausreichend sind. Somit verkleinert sich unsere Datenmenge auf 13.786 Bilder. Weiterhin werden wir zuerst lediglich eine Unterscheidung zwischen malignen und benignen Läsionen vornehmen und keine Einzelunterscheidungen zwischen den verschiedenen Krankheiten.

Die Bilder in dem ISIC Datensatz bestehen ausschließlich aus Aufnahmen von melanozytischen Hautläsionen, die anhand von Biopsien als maligne oder benigne Läsionen annotiert sind:

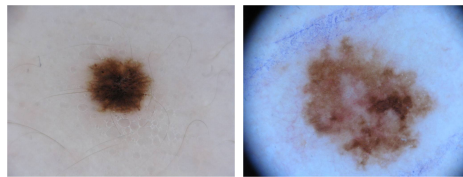


Abbildung 2: Beispiele einer benignen (links) und einer malignen (rechts) Läsion.

Esteva et. al (2017) [2] sortierten bei ihrer Datenverarbeitung alle unscharfen Bilder und solche, die von zu weit weg aufgenommen aus. Zusätzlich beinhaltete ihr Datensatz auch Bilder der gleichen Läsion aus mehreren verschiedenen Blickwinkeln, was dem Training des Klassifizierers positiv beeinflusst hat. In unserem Fall tritt jede Läsion genau einmal auf, weswegen wir bei der Aufteilung des Datensatzes in Trainings- und Validierungsdaten, anders als die Autoren, auch keinen gesonderten Trenn-Algorithmus anwenden müssen, der dafür sorgt, dass zusammengehörige Bilder auch in das gleiche Subset kommen. Allerdings werden wir zur Verbesserung der Performanz unseres Klassifizierers alle Bilder sowohl in originaler Darstellung als auch in einer rotierten Darstellung präsentieren. Wir erhoffen uns dadurch einen stabileren Klassifizierer, der auch für die Anwendung in einer mobilen Android App geeignet wäre.

Zusammengefasst gestaltet sich unsere Datengewinnung und -verarbeitung etwas anders als die im originalen Paper von Esteva et. al (2017). Wir streben eine Vereinfachung des Klassifizierungsverfahrens an, indem wir nur zwischen zwei verschiedenen Klassen (maligne vs. benigne) unterscheiden und dafür insgesamt weniger Daten nutzen. Dies hat den Vorteil, dass wir anders als im Paper nicht auf die Mithilfe ausgebildeter Dermatologen angewiesen sind, die sich alle Bilder des Datensatzes einzeln ansehen müssten, um die Daten den 2.032 verschiedenen Krankheiten zuzordnen. Unser Ziel ist es einen allgemeingültigeren Klassifizierer zu finden, der dann für die jeweiligen Anwendungsfälle spezialisiert werden kann.

## Methode

### Optional: Android App

Damit der Klassifikator auch für jedermann nutzbar ist, soll falls wir noch genügend Zeit haben eine intuitiv bedienbare Android App entwickelt werden mit der ganz einfach über ein Foto gecheckt werden kann ob man zum Arzt gehen sollte oder der potentielle Hautkrebs harmlos ist.

## Kontakt

Name	Matrikelnummer	E-Mail Adresse
Florence Lopez	3878792	florence.lopez@student.uni-tuebingen.de
Jonas Einig	XXXXXXX	jonas.einig@student.uni-tuebingen.de
Julian Späth	3938726	julian.spaeth@student.uni-tuebingen.de

## Literatur

- [1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2016. CA: a cancer journal for clinicians, 66(1):7–30, 2016.
- [2] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639):115–118, 2017.
- [3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436–444, 2015.
- [4] Memorial Sloan Kettering Cancer Center. Isic dermoscopic archive, 2017. Eingesehen am 26.11.2017.