

Vorhersage des peripheren Sehvermögens anhand von Fundusfotografien mittels Machine Learning

Motivation

Viele Augenerkrankungen führen zu Gesichtsfeldveränderungen, die sehr häufig mit einem irreversiblen Verlust des peripheren und zentralen Sehvermögens einhergehen. Dadurch werden einzelne Ausschnitte des Blickfeldes schwächer oder gar nicht mehr wahrgenommen. Mit der Gesichtsfeldmessung werden solche Veränderungen im Gesichtsfeld festgestellt und quantifiziert. Die Gesichtsfeldmessung ist allerdings sowohl zeitaufwändig als auch fehleranfällig, da sie von der Mitarbeit des Patienten abhängt. Eine weitere weniger aufwendige und häufig verwendete Diagnosemethode für Erkrankungen ist die Fundusfotografie. Methoden des Machine Learning, die eine genaue Vorhersage über den Zustand des Gesichtsfeld treffen können, würden in der Praxis dabei unterstützen, die Vorsorge und Kontrolle zu vereinfachen und zeitaufwändige Untersuchungen zu reduzieren.

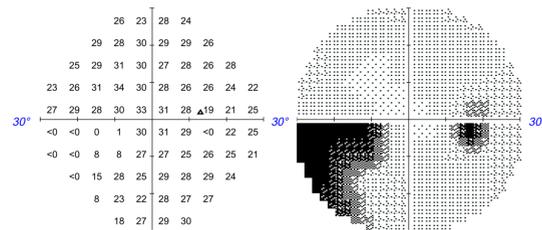


Figure 1. Auszug aus einer Gesichtsfeldmessung mit Informationen der retinalen Sensitivität. Links: Numerische Sensitivitätsgrafik. Rechts: Graustufen Karte.

In einer verwandten Studie zur Erkennung von glaukomatösen Gesichtsfelddefekten anhand der retinalen Nervenfaserschichtdicke, konnte bereits gezeigt werden, dass auf ähnlichen Daten die Defekte zuverlässig vorhergesagt werden können.[1] Ebenso konnte in einer weiteren Studie gezeigt werden, dass mit einem DCNN die glaukomatösen neuroretinalen Schäden anhand der minimalen Randsaumweite der Bruch'schen Membranöffnung (MRW) aus SD-OCT Bildern quantifiziert und vorhergesagt werden können.[2]

Ziele der Arbeit

Diese Masterarbeit befasst sich mit der Frage, inwieweit mit Methoden aus dem Bereich des Machine Learnings das periphere Sehen des menschlichen Auges anhand von Fundusfotos vorhergesagt werden kann. Des Weiteren werden Ergebnisse aus der optischen Kohärenztomographie (OCT) in die Experimente mit eingeschlossen. Hierfür wurden drei verschiedenen Machine Learning Modelle zur Erkennung, Quantifizierung und Verortung von Gesichtsfelddefekten entwickelt und ausgewertet. Die Ergebnisse der Arbeit sollen folgende Fragen beantworten:

- Können anhand von Fundusfotografien Gesichtsfelddefekte zuverlässig erkannt und quantifiziert werden?
- Haben die Werte der retinalen Nervenfaserschicht und der minimalen Randsaumweite einen Einfluss auf die Vorhersage von Gesichtsfelddefekten?

Quellen

- [1] Mark Christopher u. a. „Deep Learning Approaches Predict Glaucomatous Visual Field Damage from OCT Optic Nerve Head En Face Images and Retinal Nerve Fiber Layer Thickness Maps“. In: *Ophthalmology* 127.3 (März 2020), S. 346–356. doi: 10.1016/j.ophtha.2019.09.036. url: <https://doi.org/10.1016/j.ophtha.2019.09.036>.
- [2] Atalie C. Thompson, Alessandro A. Jammal und Felipe A. Medeiros, „A Deep Learning Algorithm to Quantify Neuroretinal Rim Loss From Optic Disc Photographs“. In: *American Journal of Ophthalmology* 201 (Mai 2019), S. 9–18. doi: 10.1016/j.ajo.2019.01.011. url: <https://doi.org/10.1016/j.ajo.2019.01.011>.
- [3] Klinik für Augenheilkunde des Universitätsklinikums Münster (<https://www.ukm.de/kliniken/augenklinik>)
- [4] Mingxing Tan und Quoc V. Le. „EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks“. In: (2019). doi: 10.48550/ARXIV.1905.11946. url: <https://arxiv.org/abs/1905.11946>.
- [5] R.K. Parrish und D.R. Anderson E. Hodapp. *Clinical Decisions in Glaucoma*. Mosby, 1993.

Methoden & Vorgehen

Für die Untersuchungen in dieser Arbeit wurden drei Machine Learning Modelle trainiert. Im ersten Schritt wurde ein Convolutional Neural Network anhand von Fundusfotos und den Ergebnissen aus der Gesichtsfeldmessung trainiert. Dieses Modell klassifiziert die Schwere von Gesichtsfelddefekten, die auf den Fundusfotos erkannt werden. Als Architektur wurde die EfficientNet B4 Architektur verwendet.[4] Im zweiten Schritt wurden die Ergebnisse des CNN mit den Daten aus der optischen Kohärenztomographie verknüpft. Auf dem erweiterten Datensatz wurde ein Gradient Boosted Tree Modell angewendet, um die Vorhersagegüte zu verbessern. Neben der quantitativen Erkennung von Gesichtsfelddefekten, wurde mit einem weiteren Modell die Lage von Gesichtsfelddefekten vorhergesagt. Dazu wurde ein XGBoost Regressionsmodell anhand der Daten aus der Perimetrie und der optischen Kohärenztomographie trainiert. In Abbildung 2 wird das Vorgehen und die in der Arbeit gewählten Verfahren grafisch veranschaulicht.

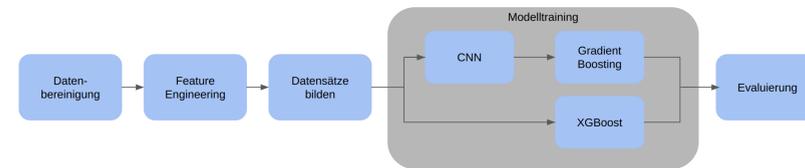


Figure 2. Schaubild über Methoden und Ablauf der Experimente.

Datengrundlage

Als Datengrundlage dienen Fundusfotos, Untersuchungsergebnisse aus der Perimetrie sowie der optischen Kohärenztomographie, die durch die Klinik für Augenheilkunde am Universitätsklinikum Münster (UKM) bereitgestellt wurden.[3] Aus diesen Datenquellen wurde für das Training eines CNN, zur Erkennung und Quantifizierung der Gesichtsfelddefekte, ein Datensatz mit 6126 Paare von Fundusfotografien und Gesichtsfeldmessung von 2265 Patienten erstellt. Mit einem weiteren Datensatz aus 1372 Paare von Gesichtsfeldmessungen und der optischen Kohärenztomographie, wurde ein XGBoost Modell für die orts aufgelöste Vorhersage von Defekten trainiert.

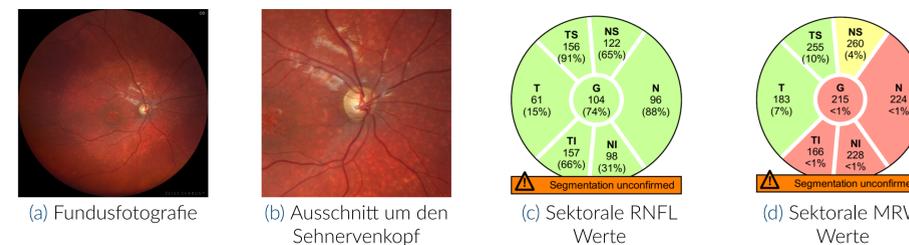


Figure 3. Beispiel einer Fundusfotografie und der Messwerte der RNFL und MRW.

Für die Erstellung der Datensätze wurden Methoden des Feature Engineerings angewendet. Da die Untersuchungsergebnisse aus der Perimetrie und den OCTs nur als PDF vorlagen wurden die relevanten Werte mittels *Regular Expression* und *Optical Character Recognition* extrahiert. Ebenso lagen für die Datensätze noch keine Label vor, daher wurden anhand der Werte aus der Perimetrie Feature für die Erstellung von Klassenlabels genutzt. Hierzu wurden alle Fundusfotos, anhand der gemessenen Mean Deviation aus der Perimetrie, nach den Hodapp-Stadien in Klassen eingeteilt.[5]

Ergebnisse

Durch die Experimente in dieser Arbeit konnte nicht bestätigt werden, dass sich Fundusfotografien alleine zur Erkennung und Quantifizierung von Gesichtsfelddefekten eignen. Das CNN, das zur Quantifizierung von Gesichtsfelddefekten trainiert wurde, erreicht eine Accuracy von 54 Prozent auf den Testdaten.

Evaluationsmetriken - Testdaten			
	Precision	Recall	F ₁ -Score
Stadium 1	0,74	0,69	0,71
Stadium 2	0,15	0,22	0,18
Stadium 3	0,45	0,4	0,42
Micro-Average	0,58	0,54	0,56
Macro-Average	0,45	0,44	0,44

Figure 4. Übersicht über Evaluationsmetriken auf den Testdaten des CNN Klassifikationsmodell zur Erkennung der Hodapp-Stadien.

Durch die Verknüpfung von Ergebnissen aus den optischen Kohärenztomographien konnte die Modellgüte mit Methoden des Gradient Boosting verbessert werden. Hier erreicht das Modell zur Quantifizierung von Gesichtsfelddefekten eine Accuracy von 86 Prozent auf den Testdaten.

Evaluationsmetriken - Testdaten			
	Precision	Recall	F ₁ -Score
Stadium 1	0,91	0,95	0,93
Stadium 2	0,58	0,61	0,59
Stadium 3	0,81	0,67	0,73
Micro-Average	0,85	0,86	0,85
Macro-Average	0,76	0,75	0,75

Figure 5. Übersicht über Evaluationsmetriken des Gradient Boosted Klassifikationsmodells.

Für die orts aufgelöste Vorhersage von Gesichtsfelddefekten erreichte das XGBoost Regressionsmodell ein mittleres R^2 von 0.56 bei einem mittleren MAE von 4.41 dB über alle Modelle.

XGBoost Regressions Evaluation - Testdaten				
	y_{pred} (dB)	y_{true} (dB)	R^2	MAE (dB)
Quadrant 1	21.75 (7.25)	21.39 (9.95)	0.538	4.711
Quadrant 2	21.68 (7.53)	21.78 (9.57)	0.561	4.497
Quadrant 3	22.88 (7.37)	22.55 (9.65)	0.563	4.333
Quadrant 4	22.86 (7.24)	22.98 (9.18)	0.574	4.100

Figure 6. Gegenüberstellung der durchschnittlichen retinale Sensitivität je Quadrant aus der Perimetrie und der durchschnittlichen Werte aus der Vorhersage des Modells auf den Testdaten. Zusätzlich werden R^2 und der MAE je Modell gezeigt.

Zusammenfassung

Die wichtigsten Ergebnisse dieser Arbeit sind:

- Nach den Erkenntnissen dieser Arbeit bieten Fundusfotos alleine, mittels der gewählten Machine Learning Modelle, noch keine zuverlässige Vorhersage von Gesichtsfelddefekten.
- Durch Einbezug der Ergebnisse aus der optischen Kohärenztomographie konnte die Vorhersagegüte des Modells verbessert werden. Entsprechend haben die RNFL und MRW einen positiven Einfluss auf die Vorhersage von Gesichtsfelddefekten.