

# Maschinelles Lernen für Physiker\*innen -Master Physik-

## Klassifizierung durch Verfahren basierend auf Maschinellern Lernen von Retina Aufnahmen mittels Optischer Kohärenztomographie

Kevin Sedlacek

kevin.sedlacek@tu-dortmund.de

Björn Wendland

bjoern.wendland@tu-dortmund.de

Projektbericht vorgelegt von Björn Wendland

Datum der Abgabe: 29.07.2018

TU Dortmund - Fakultät Physik

# Inhaltsverzeichnis

1. Einleitung	1
2. Datensatz	2
3. Faltendes neuronales Netz zur Klassifizierung von OCT Aufnahmen der Retina	3
4. Flaches tiefes neuronales Netz zur Klassifizierung von OCT Aufnahmen der Retina	9
5. Zusammenfassung und Schlussfolgerungen	12
A. Anhang	14

## 1. Einleitung

Die Optische Kohärenztomographie (OCT) basiert auf dem Prinzip des Michelson-Interferometers [1]. Kohärentes Licht, welches sich aus Wellenlängen einer festgelegten Bandbreite zusammensetzt, wird auf einen semitransparenten Spiegel gesendet. Dabei wird ein Teil des Lichtes transmittiert und der andere Teil reflektiert. Einer dieser entstandenen Teilstrahlen wird auf einen Spiegel gesendet, an dem er reflektiert und als Referenzstrahl benutzt wird. Der andere Teilstrahl wird auf die zu untersuchende Probe gesendet. Dringt dieser Teilstrahl in die Probe ein, so tritt er mit den Atomen und Molekülen des Probenmaterials in Wechselwirkungen. Er kann dabei gestreut, absorbiert oder reflektiert werden. Die Reflektion kann dabei in unterschiedlichen Schichttiefen geschehen.

Der reflektierte Referenzstrahl und der von der Probe reflektierte Strahl treffen an dem semitransparenten Spiegel wieder zusammen, wo beide in die gleiche Richtung transmittiert beziehungsweise reflektiert werden. Dabei interferieren die beiden Teilstrahlen und werden auf einen Detektor geleitet. Dieser analysiert die Intensität des Signals als Funktion der Wellenlänge. Aus der Form des Signals kann dabei ermittelt werden bei welcher Probentiefe die Reflexion des auf die Probe treffende Strahl vorliegt. Während der Analyse überlagern sich jedoch mehrere Signale, die durch Reflektion in verschiedenen Probentiefen entstehen, sodass diese zunächst getrennt werden müssen. Dazu kann die Fourier-Transformation genutzt werden, welche diese Signale trennt und somit direkt die komplette Information über das Tiefenprofil der Probe liefert. Durch Verschieben der Apparatur entlang der Probe können somit 2 oder 3 dimensionale Bilder der Probe erstellt werden.

Eines der Hauptanwendungsgebiete der OCT findet sich in der Medizin wieder. Es kann dazu genutzt werden die menschliche Retina auf Krankheiten zu untersuchen. Die Retina besitzt eine komplizierte, mehrschichtige Struktur und Erkrankungen der Retina können in jeder Schicht auftreten. Daher ist die Aufnahme eines Tiefenprofils der Retina von großem medizinischen Interesse.

In der vorliegenden Projektarbeit werden OCT Aufnahmen der Retina analysiert, wobei der Datensatz von Ref. [2] verwendet wird. Die darin enthaltenen Aufnahmen lassen sich in 4 Klassen unterteilen. Während drei Klassen Aufnahmen beinhalten, welche Erkrankungen der Retina beinhalten, zeigt die vierte Klasse Retina Aufnahmen, die keiner der drei Erkrankungen aufweist und im Folgenden als NORMAL bezeichnet wird. Eine der Krankheiten ist die choroidale Neovaskularisation (CNV), wobei sich irreguläre Blutgefäße in der Netzhaut bilden, aus welchen Flüssigkeiten treten können, welche zu Schwellungen der Netzhaut führen. Drusen, DRUSEN im Folgenden, bezeichnen Abfallprodukten aus der zentralen Netzhaut, die sich unterhalb der Netzhaut ansammeln. Das diabetische Makulaödem (DME) ist eine Schwellung der Makula, die im Zuge der Bildung von kleinen Gefäßausbuchtungen in der Netzhaut entsteht.

In dem Projektbericht wird eine Methode basierend auf überwachtem, maschinellem Lernen vorgestellt, die es ermöglicht diese OCT Aufnahmen zu klassifizieren. Diese Methode verwendet dabei ein tiefes faltendes neuronales Netz (CNN) und wird mit einer Methode verglichen, die auf ein neuronales Netz welches aus flachen dichten Lagen besteht (NN).

## 2. Datensatz

Tabelle 1: Zusammensetzung des untersuchten Datensatzes aufgeteilt nach den im Datensatz vorhandenen Krankheiten, CNV, DME und DRUSEN, und Aufnahmen, die keine dieser Krankheiten aufweisen und mit NORMAL gekennzeichnet werden.

Erkrankung	Anzahl
NORMAL	26315
CNV	37205
DME	11348
DRUSEN	8616
Total	83484

Der verwendete Datensatz besitzt die in Tabelle 1 dargestellte Zusammensetzung und in Abbildung 1 sind Beispielaufnahmen für jede der Klassen dargestellt. Die Aufnahmen sind in Graustufen aufgezeichnet, wodurch die Pixelwerte den Wertebereich  $[0, 255]$  abdecken, wobei schwarze Pixel dem Wert 0 und weiße Pixel dem Wert 255 entsprechen. Die Erkrankung CNV ist in Abbildung 1(b) abgebildet und durch eine deutliche Schwellung der Retina zu erkennen. Abbildung 1(c) zeigt eine Retina, die von einem DME befallen ist, was an der ründlichen, schwarzen Färbung, die einer Gefäßausbuchtung entspricht, innerhalb einer deutlichen Schwellung der Retina. Auch die DRUSEN lassen sich in Abbildung 1(d) gut erkennen, da unterhalb der Netzhaut deutliche Ausbuchtungen zu sehen sind, die im Vergleich zu der NORMAL Aufnahme in Abbildung 1(a) nicht erwartet werden. Es sollte jedoch hier herausgestellt werden, dass die Aufnahmen in Abbildung 1 ideale Beispiele darstellen, bei denen die Klassifizierung auch durch unerfahrene Betrachter vorgenommen werden kann. Bei viele Aufnahmen ist dies jedoch nicht der Fall, sodass zur Klassifizierung eine andere Methodik oder Erfahrung in der Auswertung dieser Aufnahmen notwendig ist.

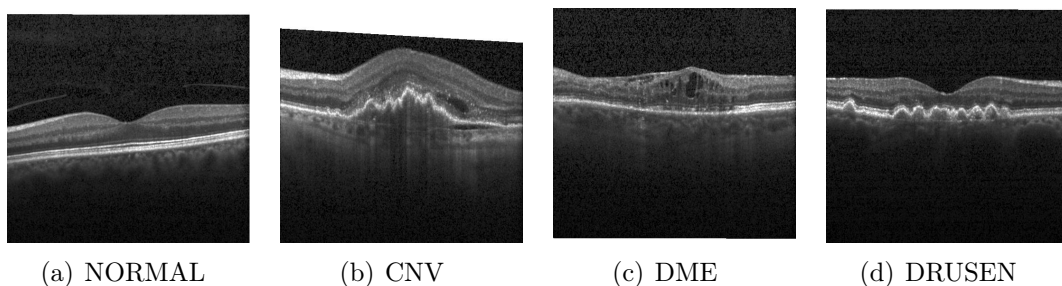


Abbildung 1: OCT Aufnahmen der Retina, die beispielhaft ein Aufnahme jeder untersuchten Klasse im Datensatz zeigen. Dabei zeigt Abbildung 1(a) eine OCT Aufnahme der Klasse NORMAL, die von keiner der Krankheiten, CNV, DME und DRUSEN, befallen ist, welche in den Abbildungen 1(b), 1(c) repektive 1(d) dargestellt sind.

### 3. Faltendes neuronales Netz zur Klassifizierung von OCT Aufnahmen der Retina

Die Klassifizierung von OCT Aufnahmen der Retina stellt eine Bildklassifizierung dar. Für diese Aufgabenstellungen haben sich in der Praxis tiefe faltende neuronale Netze (CNN) bewährt. Das bekannteste Beispiel stellt das Erkennen von handgeschriebenen Zahlen dar, wobei eine nahezu menschliche Genauigkeit erreicht werden kann [3]. Daher wird auch im Rahmen dieser Projektarbeit auf diesen Typ der tiefen neuronalen Netze zurückgegriffen. Die Aufnahmen besitzen unterschiedliche Dimensionen und werden vor der Übergabe an das CNN auf eine einheitlich Größe von  $(400 \times 400)$  skaliert. Pixel, die bei diesem Vorgang zu der Aufnahme hinzugefügt werden, werden weiß eingefärbt. Zudem werden die Pixelwerte auf den Wertebereich  $[0, 1]$  eingeschränkt, indem der minimale Pixelwert einer Aufnahme auf 0 und der maximale Pixelwert auf 1 skaliert wird und die übrigen Pixel entsprechend angepasst werden.

Demzufolge erhält das CNN in diesem Fall die Pixelwerte in Form einer 2 dimensionale Liste der Dimension  $(400 \times 400)$  als Eingangswerte. Das CNN besteht in der Regel hauptsächlich aus drei Bestandteilen, den sogenannten faltenden Lagen, welche im Folgenden als Conv2D Lagen bezeichnet werden, den Aggregationsschichten, die mehrere Neuronen einer Conv2D Lage zu einem Neuron zusammenfassen und welche im Folgenden als Pooling Lagen bezeichnet werden, und flache vollständig vernetzte dichte Lagen. In den faltenden Lagen wird die Dimension einer Faltungsmatrix (Kernel) definiert, welche die 2 dimensionale Neuronenstruktur in festgelegten Schrittweiten abrastert. Bei jeder Schritt wird über eine diskrete Faltung wird die Aktivität eines jeden Neurons, die innerhalb des Kernels liegen berechnet. Die Addition der Aktivitäten der einzelnen Neuronen ergibt den Ausgangswert für jeden Schritt. Anschaulich bedeutet dies, dass eine Neuronenstruktur, die beispielsweise in einem  $(400 \times 400)$  Format vorliegt und mit einer  $(2 \times 2)$  Faltungsmatrix in Schrittweiten der Form  $(2, 2)$  abgerastert wird, eine Ausgangsstruktur der Form  $(200 \times 200)$  generiert. Die Anzahl der Faltungsmatrizen, die zur Abrasterung benutzt werden, wird ebenfalls in jeder Conv2D Lage festgelegt.

Für die Pooling Lage wird die Dimension eines Fensters festgelegt, welches die Probe abfährt und bei jedem Schritt die Neuronen innerhalb dieses Wertes zu einem Neuron zusammenfasst. Wird beispielsweise ein  $(2 \times 2)$  Fenster definiert, welches eine Neuronenstruktur der Form  $(200 \times 200)$  abfährt, wobei keine Überlappung der einzelnen Schritte entsteht, bedeutet dies, dass je vier Neuronen zu einem zusammengefasst werden, sodass sich eine  $(100 \times 100)$  Ausgangsstruktur ergibt. Häufig wird bei jedem Schritt das Neuron mit der höchsten Aktivität behalten, während die übrigen Neuronen innerhalb eines Fensters verworfen werden. Demnach wird durch diese Lage die Dimension der Neuronenstruktur und somit die Laufzeit verringert, was in den meisten Fällen jedoch keine Verschlechterung des Lernerfolgs des CNN nach sich zieht. Im Gegenteil, es bietet sogar die Möglichkeit Übertraining zu vermindern. Übertraining bedeutet hierbei, dass sich das CNN zu stark an den Trainingsdatensatz anpasst und somit Fluktuationen innerhalb einer Klasse nicht mehr richtig zuordnen kann.

Eine weitere Möglichkeit Übertraining zu vermeiden ist das Einbinden der sogenannten Dropout Lagen. Hierbei wird ein festgelegter Bruchteil an Neuronen einer Lage zufällig

ausgewählt und in einem Trainingsschritt verworfen.

Zumeist bestehen CNNs zunächst aus einer abwechselnden Struktur aus Conv2D Lagen und Pooling Lagen bis die Dimension der Neuronenstruktur hinreichend stark reduziert ist. Als nächsten Schritt, werden die Ausgangswerte der letzten Lage dieser Struktur in eine 1 dimensionale Liste abgespeichert, welche als Flatten Lage bezeichnet wird. Diese Ausgangswerte werden wiederum in vollständig vernetzte dichte Lagen übergeben. Für diese werden die Anzahl an Neuronen definiert, die sich am Ausgang der Lagen befinden. Vollständig vernetzt bedeutet hierbei, dass jedes Neuron einer dichten Lage mit allen Neuronen der vorherigen und nachfolgenden Lage vernetzt ist.

Die letzte Lage oder Ausgangslage des CNN besteht aus einer vollständig vernetzten dichte Lage, bei der die Anzahl an Neuronen der Anzahl an Klassen des Datensatzes entspricht. Das hier verwendete CNN besitzt demnach vier Neuronen in der Ausgangslage.

Das CNN wird auf 70 % des Datensatzes trainiert, wobei der Lernerfolg durch die restlichen 30 % des Datensatzes validiert wird. Bei der Aufteilung wird darauf geachtet, dass die Zusammensetzung des Datensatzes in Tabelle 1 in beiden Teildatensätze beibehalten wird. Die sogenannte Batch Größe legt fest, wie viele Aufnahmen in einem Trainingsschritt durch das CNN propagiert werden, wobei eine Epoche des Trainings dann erfolgt ist, wenn der komplette Trainingsdatensatz durch das CNN propagiert wurde. Das Training erfolgt im vorliegenden Fall über 40 Epochen mit einer Batch Größe von 100.

Zur Klassifizierung des in diesem Projektbericht analysierten Datensatzes wird folgende Struktur als erste Referenzstruktur definiert:

- Eingangslage mit Eingangsdaten der Form  $(400 \times 400)$
- Conv2D Lage mit 64 Faltungsmatrizen der Dimension  $(4 \times 4)$  und Schrittweite  $(2, 2)$   
 $\Rightarrow (199 \times 199 \times 64)$  Ausgangsstruktur
- Pooling Lage mit  $(3 \times 3)$  Fenster  $\Rightarrow (66 \times 66 \times 64)$  Ausgangsstruktur
- Conv2D Lage mit 32 Faltungsmatrizen der Dimension  $(4 \times 4)$  und Schrittweite  $(2, 2)$   
 $\Rightarrow (32 \times 32 \times 32)$  Ausgangsstruktur
- Pooling Lage mit  $(3 \times 3)$  Fenster  $\Rightarrow (10 \times 10 \times 32)$  Ausgangsstruktur
- Dropout der Größe 0.25
- Dichte Lage mit 1000 Neuronen  $\Rightarrow (10 \times 10 \times 1000)$  Ausgangsstruktur
- Dichte Lage mit 250 Neuronen  $\Rightarrow (10 \times 10 \times 250)$  Ausgangsstruktur
- Flatten Lage mit 25000 Ausgangsneuronen
- Dichte Lage mit 100 Ausgangsneuronen
- Dropout der Größe 0.5
- Dichte Lage mit 32 Ausgangsneuronen
- Dichte Ausgangslage mit 4 Ausgangsneuronen

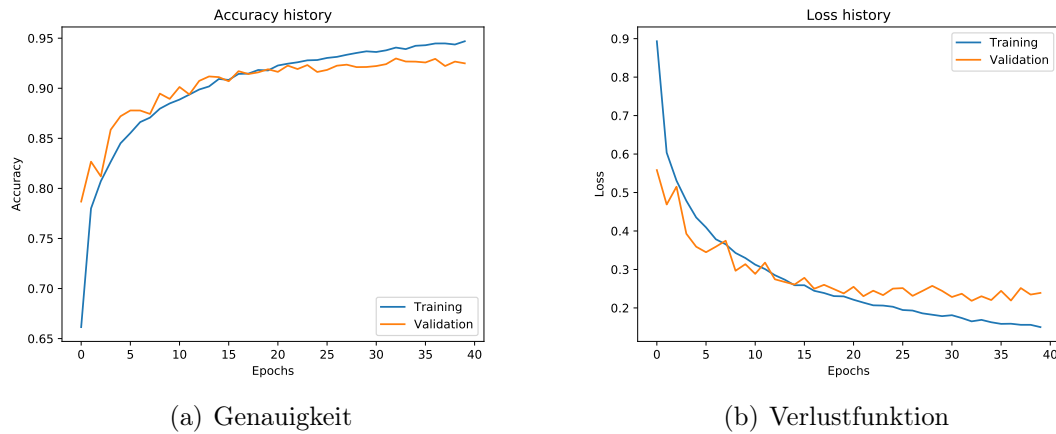


Abbildung 2: Genauigkeit 2(a) und Wert der Verlustfunktion 2(b) nach jeder Epoche des Trainings bei Propagation des Trainingsdatensatzes und Validierungsdatensatzes durch das CNN mit der Referenzstruktur.

Als Aktivierungsfunktion der versteckten Lagen wird die elu (exponential linear unit) Funktion verwendet und als Aktivierungsfunktion der Ausgangslage wird die softmax Funktion verwendet. Zudem wird die kategorische Kreuzentropie als Verlustfunktion und der Adam Optimierer [4] verwendet. Als Metrik wird die Genauigkeit verwendet, welcher sich durch das Verhältnis der Anzahl der richtig klassifizierten Aufnahmen und der Gesamtanzahl an Aufnahmen innerhalb eines Datensatzes berechnet. Der Lernerfolg wird durch das Aufzeichnen der Genauigkeit und dem Wert der Verlustfunktion nach jeder Epoche des Lernens auf dem Trainingsdatensatz ermittelt und durch die entsprechenden Werte, welche durch die Propagation des Validierungsdatensatzes durch die Netzstruktur ermittelt wird, validiert.

Es ergeben sich die in Abbildung 2 dargestellten Kurven für die Genauigkeiten und Verlustfunktion nach jeder Epoche. Es lässt sich erkennen, dass die Referenzstruktur bereits gute Ergebnisse liefert, da eine Genauigkeit auf dem Validierungsdatensatz von über 90 % erzielt wird. Da sich weder die Genauigkeiten noch die Werte der Verlustfunktion auf dem Validierungsdatensatz stark von den entsprechenden Werten auf dem Trainingsdatensatz unterscheiden, liegt nur geringes Übertraining vor, wodurch auf weitere Optimierungen des CNN in Bezug auf die Verminderung des Übertrainings beispielsweise durch das Einfügen von zusätzlichen Dropout Lagen oder Regularisierung, verzichtet wird.

Als Maß für den Lernerfolg dient zudem die Verwirrungsmatrix, welche die Prozentuale Verteilung der Aufnahmen einer Klasse auf die durch das vorhergesagten Klassenzugehörigkeit angibt. Das bedeutet demnach, dass sich die Zeilen der Matrix zu 100 % addieren und die Matrix im Idealfall nur diagonale Einträge hat. Für die Referenzstruktur ist die Verwirrungsmatrix in 3(a) dargestellt und es lässt sich eine diagonal dominante Struktur feststellen. Es lässt sich jedoch auch erkennen, dass die Aufnahmen der Klasse DRUSEN nur zu 64 % richtig und zu 27 % falsch als CNV klassifiziert werden. Wird anstelle der oben definierten Genauigkeit, die Genauigkeit des Netzes durch den Mittelwert der Prozentwerte der diagonalen Elemente angegeben so ergibt sich eine Genauigkeit von 86 %.

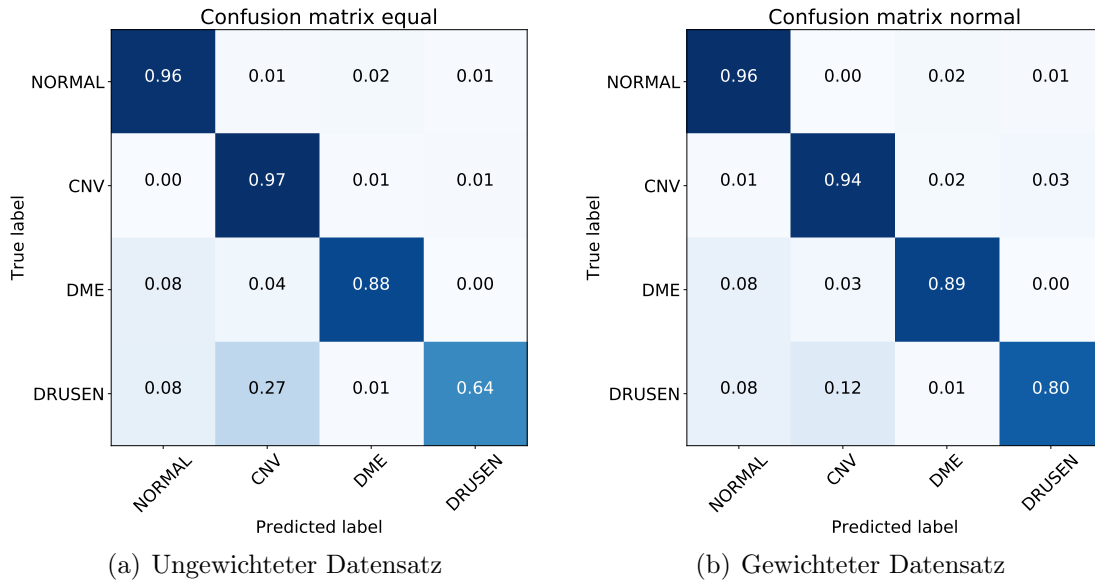


Abbildung 3: Verwirrungsmatrizen nach dem Training ohne Gewichtung des Datensatzes in Abbildung 3(a) und mit verwendeter Gewichtung in Abbildung 3(b).

Da es gewünscht ist, dass für jede Klasse die höchstmögliche Genauigkeit erreicht, wird diese Genauigkeit im Folgenden als Maß für den Lernerfolg verwendet und Gesamtgenauigkeit bezeichnet.

Wie sich in Tabelle 1 erkennen lässt, sind die Klassen im Datensatz unterschiedlich stark repräsentiert. Die Klassen, die durch die Referenzstruktur am schlechtesten klassifiziert werden sind die Klassen DME und DRUSEN, die am schwächsten im Datensatz vertreten sind. Um im Training dies zu berücksichtigen, wird jeder Aufnahme innerhalb des Datensatzes ein Gewicht zugeordnet, sodass sich die Gewichte der Aufnahmen innerhalb einer Klasse zu 1 addieren, und das Gewicht in der Verlustfunktion berücksichtigt. Für die Referenzstruktur ergibt dies die in Abbildung 3(b) dargestellte Verwirrungsmatrix. Die Gesamtgenauigkeit steigt hier auf 90 % und die Klasse DRUSEN wird deutlich besser klassifiziert, wobei die Klassen NORMAL und CNV weiterhin sehr gut klassifiziert werden. Daher werden stets im Folgenden diese Gewichte im Training berücksichtigt.

Die weiterführende Optimierung der Netzstruktur gestaltet sich als sehr zeitaufwendig, da das Training der Referenzstruktur bereits 17 Stunden dauert. Daher werden im Folgenden zwei Optimierungsansätze diskutiert, die im Rahmen der Projektarbeit durchgeführt werden.

Zunächst wird getestet welche Auswirkung die Änderung der Aktivierungsfunktion der versteckten Lagen auf den Lernerfolg des CNN haben. In Abbildung 4(a) ist die resultierende Verwirrungsmatrix dargestellt, wenn relu (rectangular linear unit) anstelle von elu verwendet wird. Die Gesamtgenauigkeit ist mit 89 % etwas geringer, sodass keine Verbesserung erzielt werden kann.

Zudem wird festgestellt, dass die dichte Lage nach der Flatten Lage enorm viele Parameter aufweist, sodass die Größe der dichten Lagen vor der Flatten Lage angepasst



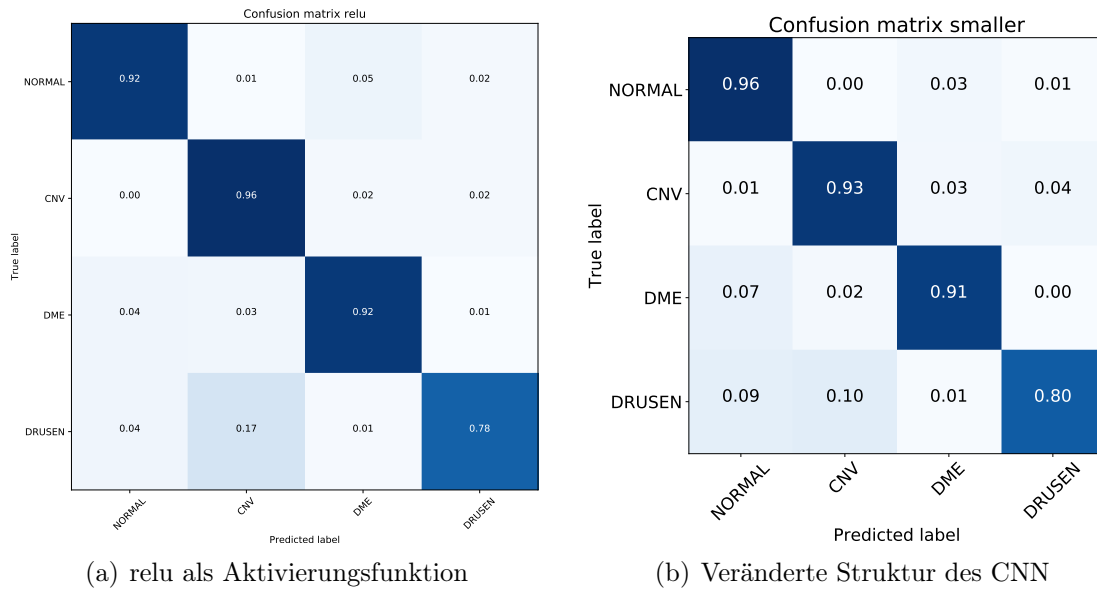


Abbildung 4: Verwirrungsmatrix bei Verwendung des gewichteten Datensatzes des CNN mit der Referenzstruktur mit relu als Aktivierungsfunktion 4(a) und veränderter Struktur der dichten Lagen 4(b).

werden. Die Lage mit 1000 Neuronen wird auf 256 Neuronen reduziert, die Lage mit 256 Neuronen auf 128 Neuronen reduziert. Somit reduziert sich die Trainingsdauer stark. Die resultierenden Genauigkeitswerte, Werte der Verlustfunktion und die Verwirrungsmatrix sind für diese Konfiguration in dargestellt. Es lässt sich erkennen, dass sich die Genauigkeit leicht verbessert, jedoch auch 90 % beträgt. Somit wird diese Struktur als finale

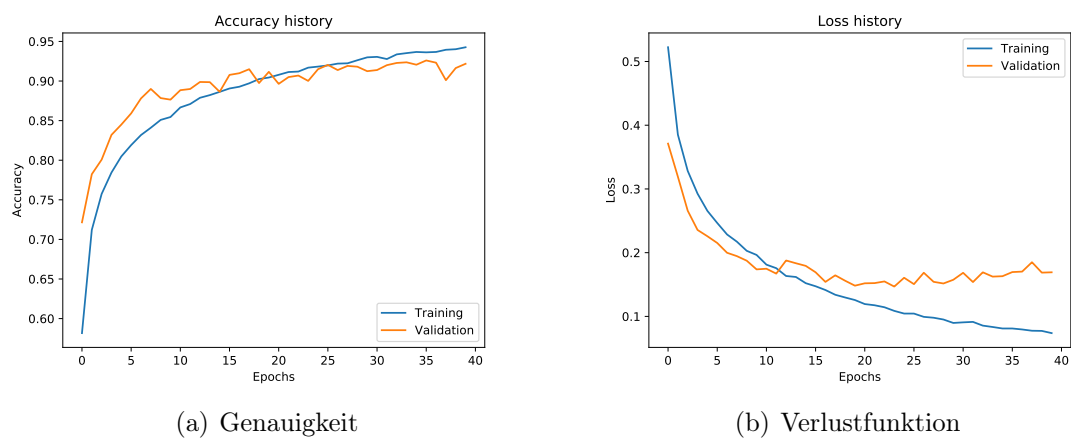


Abbildung 5: Genauigkeit 5(a) und Wert der Verlustfunktion 5(b) nach jeder Epoche des Trainings bei Propagation des Trainingsdatensatzes und Validierungsdatensatzes durch das CNN mit der angepassten Struktur der dichten Lagen.

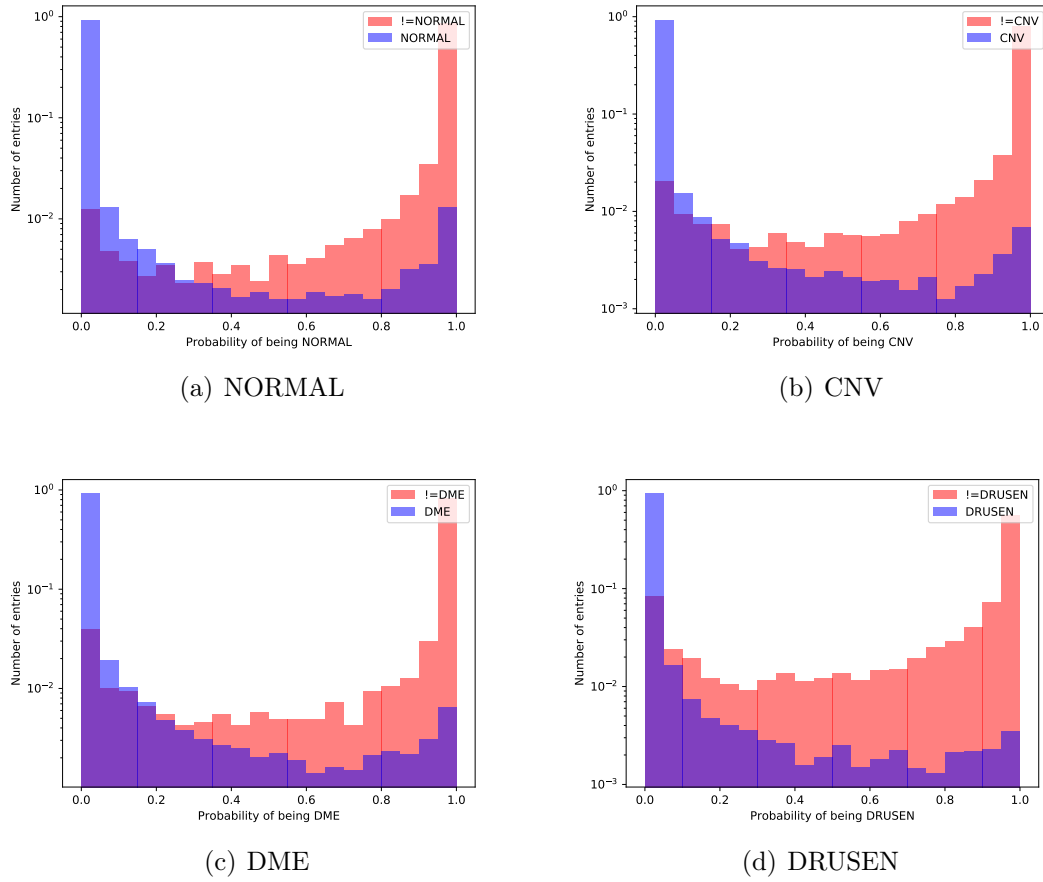


Abbildung 6: Verteilung der von dem CNN vorhergesagten Wahrscheinlichkeiten für die Aufnahmen innerhalb einer Klasse ein Mitglied der Klasse zu sein und für die Aufnahmen, die nicht dieser Klasse angehören. Die Abbildungen 6(a), 6(b), 6(c) und 6(d) zeigen dies für die Klassen NORMAL, CNV, DME respektive DRUSEN.

Struktur in dieser Projektarbeit verwendet.

Die Werte der Neuronen der Ausgangslage lassen sich als Wahrscheinlichkeiten für Zugehörigkeit einer Aufnahmen zu einer Klasse interpretieren. Abbildung 6 zeigt die Verteilungen der Wahrscheinlichkeitswerte für die Aufnahmen innerhalb einer Klasse X der Klasse X anzugehören und der Wahrscheinlichkeitswerte für die Aufnahmen außerhalb einer Klasse X der Klasse X anzugehören. Dies bedeutet anschaulich, dass im Idealfall die roten Verteilungen in Abbildung 6 nur bei 1 Aufnahmen aufweisen, wohingegen die blauen Verteilungen nur bei 0 Aufnahmen aufweisen. Demnach liegt für alle Verteilungen eine sehr gute Diskriminierung von nicht der Klasse zugehörigen Aufnahmen vor.

Jedoch zeigt Abbildung , dass für die Klasse DRUSEN die Wahrscheinlichkeitsverteilung der zu DRUSEN zugehörigen Aufnahmen im Vergleich zu den Abbildungen 6(a)-6(c) einen kleineren Ausschlag bei dem Wert 1 und eine deutlich höher Flanke zu kleinen Werten aufweist, wodurch hier eine deutlich schwächere Diskriminierung vorhanden ist.

## 4. Flaches tiefes neuronales Netz zur Klassifizierung von OCT Aufnahmen der Retina

Um die Wahl des CNN zu validieren wird diese mit einer alternativen Methode verglichen. Diese basiert auf einem flachen tiefen neuronalen Netz (DNN), welches aus vollständig vernetzten dichten Lagen besteht.

Hierzu müssen die analysierten Bilder zunächst in eine 1 dimensionale Liste aus Werten umgewandelt werden. Hierzu muss beachtet werden, dass diese Liste nicht zu groß werden darf, da ansonsten das Training und die Optimierung des neuronalen Netzes zu zeitaufwendig ist. Demnach wird die Dimension der Aufnahmen zunächst auf  $(50 \times 100)$  reduziert.

Darauf aufbauend werden drei Verfahren untersucht aus diesen Bilder eine 1 dimensionale Liste zu erstellen, welche als Eingangseigenschaften des DNN der Aufnahmen verwendet werden. Zum einen werden alle Pixelwerte einer  $(50 \times 100)$  Aufnahme in eine 1 dimensionale Liste umgewandelt und dem DNN übergeben. Erste Studien hierzu liefern eine Genauigkeit der Klassifikation von ungefähr 55 %. Die zweite Methode basiert darauf für jede  $x$ - und  $y$ -Linie des Aufnahme den Mittelwert aller Pixelwerte der jeweiligen Linie zu berechnen. Demnach generiert sich eine Liste aus 150 Werten. Die Werte werden dabei auf den maximalen Wert dieser List skaliert. Auch bei dieser Methodik wird eine Genauigkeit von ungefähr 55 % erzielt. Die letzte Methode basiert darauf das Bild weiter zu verkleinern, indem ein Fenster der Dimension  $(2 \times 4)$  definiert wird, welches die Aufnahmen abrastert. Dabei soll keine Überlappung der einzelnen Schritte entstehen. Der Mittelwert der im Fenster liegenden Pixel wird für jeden Schritt der Abrasterung berechnet. Anschließend werden die so erhaltenen Mittelwerte auf den größten Mittelwert einer Aufnahme skaliert und dem DNN übergeben. In den vereinfachten ersten Studien zeigte sich bei dieser Methode eine Genauigkeit von 70 %, wodurch diese vielversprechend erscheint und im Folgenden betrachtet wird.

Um eine geeignete Referenzstruktur des DNN zu finden mit der das CNN verglichen werden kann, werden verschiedene Netzstrukturen ausprobiert. Es werden 5 Grundstrukturen des DNN definiert, welche in Tabelle 2 dargestellt sind und als Modell  $i$  ( $i = 0, 1, 2, 3, 4$ ) bezeichnet werden. Dabei folgt auf jede versteckte Lage eine Dropout Lage, wobei die Anzahl an Ausgangsknoten und verwendeten Dropout-Rate der  $i$ -ten Lage durch das  $i$ -te Element in den entsprechenden Listen daragestellt ist.

Für jede der fünf Grundstrukturen werden die Aktivierungsfunktionen elu und relu für die versteckten Lagen und die Aktivierungsfunktionen softmax und sigmoid für die Aus-

Tabelle 2: Getestete Grundstrukturen des DNN.

	Struktur der versteckten dichten Lagen	Struktur der Dropout Lagen
Modell 0	(1024, 512, 128, 64, 32)	(0.5, 0.4, 0.4, 0.3, 0.2)
Modell 1	(1024, 512, 256, 128, 64, 32, 16)	(0.5, 0.4, 0.4, 0.4, 0.2, 0.2, 0.1)
Modell 2	(512, 256, 128, 64, 32, 16)	(0.4, 0.4, 0.3, 0.3, 0.2, 0.1)
Modell 3	(1024, 256, 64, 16)	(0.6, 0.4, 0.2, 0.1)
Modell 4	(512, 128, 32)	(0.5, 0.3, 0.1)

#### 4. Flaches tiefes neuronales Netz zur Klassifizierung von OCT Aufnahmen der Retina 10

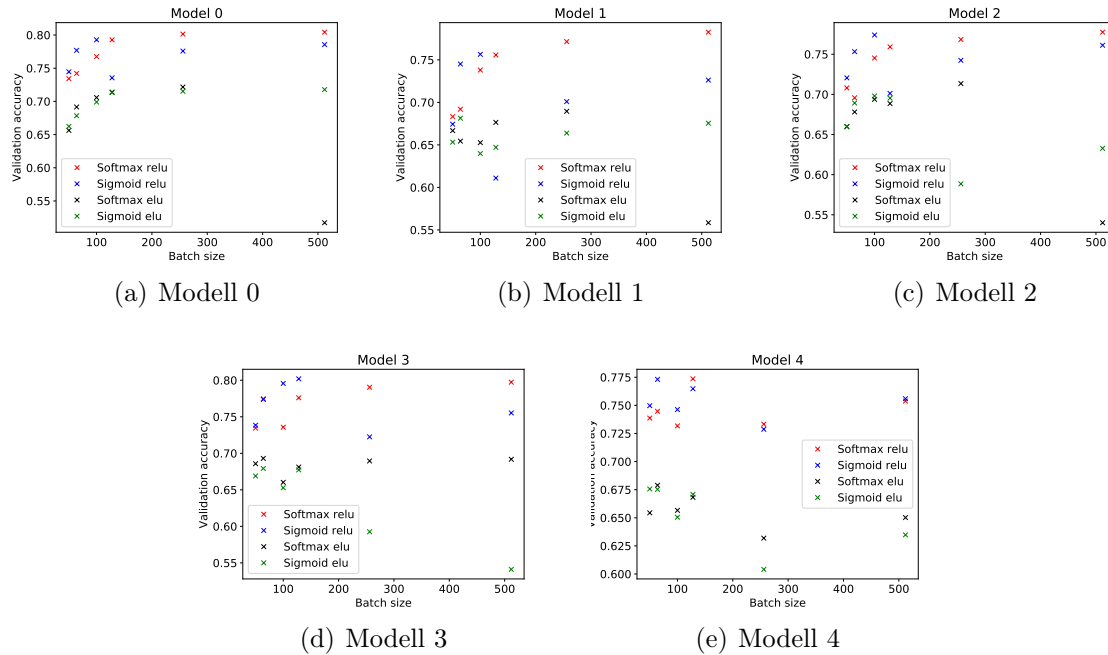


Abbildung 7: Erhaltene Genauigkeiten des DNN für die verschiedenen Konfigurationen der einzelnen Modelle als Funktion der Batch Größe.

gangslage variiert. Zudem werden die Batch Größen 50, 100, 128, 256 und 512 für jede der sich ergebenden DNN Konfigurationen getestet. Insgesamt werden somit 120 Konfigurationen für das Training des DNN getestet.

Als Verlustfunktion wird wie beim CNN die kategorische Kreuzentropie verwendet und als Metrik die Genauigkeit betrachtet. Der Adam Optimierer wird auch hierbei mit einer angepassten Lernrate von 0.0001 benutzt. Der Datensatz wird in den Trainings-, Validierungs- und Testdatensatz aufgeteilt, die aus 67,5 %, 25 % und 7,5 % des kompletten Datensatzes bestehen. Die verschiedenen Konfigurationen werden jeweils 150 Epochen lang trainiert. Die Werte der Genauigkeit bei Propagation des Validierungsdatensatzes als Funktion der Batch Größe ist für jede Konfiguration eines Modells in Abbildung 7 dargestellt.

Um ungeeignete DNN Konfigurationen herauszufiltern werden zwei verschiedene Kriterien definiert, die von den Konfigurationen erfüllt werden müssen. Zum einen muss die Genauigkeit auf dem Validierungsdatensatz größer als 73 % sein. Das zweite Kriterium filtert Konfigurationen heraus, die ein starkes Übertraining aufweisen, indem gefordert wird, dass der Wert der Verlustfunktion nach der ersten Epoche um mindestens 5 % nach der letzten Epoche gesunken ist. Nach diesen Anwendung dieser Selektionskriterien verbleiben 12 DNN Konfigurationen. Es stellt sich hierbei vor allem heraus, dass hohe Batch Größen ungeeignet sind, da keine der Konfigurationen mit einer Batch Größe von 512 und nur eine Konfiguration mit einer Batch Größe von 256 die Selektionsschritte passieren. Um nun die optimale herauszufiltern, wird die Gesamtgenauigkeit wie im Falle des CNN berechnet und das Maximum gesucht. Dieses ergibt sich für das Modell 0 unter Verwendung der relu Funktion als Aktivierungsfunktion der versteckten Lagen und der sigmoid

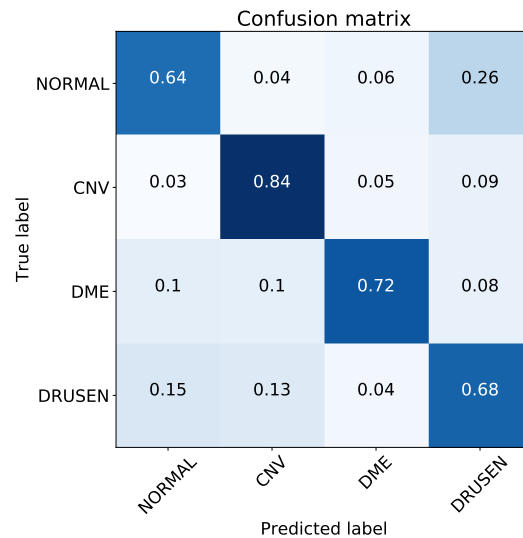


Abbildung 8: Verwirrungsmatrix der gewählten DNN Konfiguration ermittelt auf dem Trainingsdatensatz.

Funktion als Aktivierungsfunktion der Ausgangslage bei einer Batch Größe von 50 und beträgt 72 % auf dem Validierungsdatensatz. Der Verlauf der Genauigkeit und der Werte der Verlustfunktion sind als Funktion der Epochen in Abbildung 9(a) respektive 9(b) dargestellt.

Anhand 9 lässt sich zudem feststellen, dass kein signifikantes Übertraining vorhanden ist und die Genauigkeit bereits einen Sättigungswert erreicht hat. In Abbildung 8 ist die entsprechende Verwirrungsmatrix dargestellt, welche auf dem Trainingsdatensatz ermittelt wird. Auch hier errechnet sich eine Gesamtgenauigkeit von 72 %, wodurch bestätigt wird, dass das DNN eine geeignete Struktur aufweist, die kein starkes Übertraining besitzt.

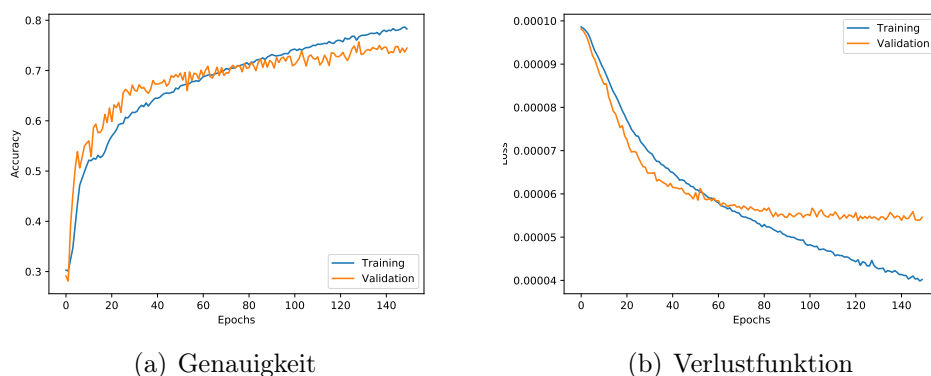


Abbildung 9: Genauigkeit und Wert der Verlustfunktion für die gewählte DNN Struktur als Funktion der Epoche bei Propagation des Trainings- und Validierungsdatensatzes.

## 5. Zusammenfassung und Schlussfolgerungen

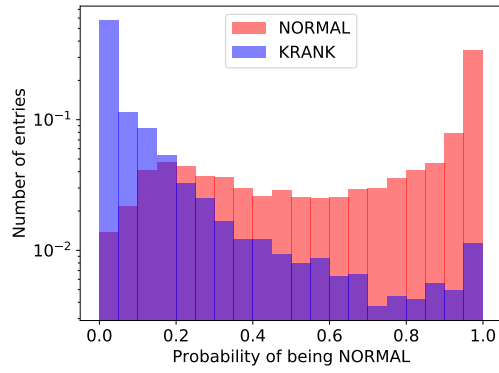
In dem vorliegenden Projektbericht wird eine Methode zur Klassifizierung von Retina Aufnahmen, welche durch die optische Kohärenztomographie (OCT) erstellt werden, basierend auf Methoden des überwachten maschinellen Lernen. Die Aufnahmen sind dabei in vier Klassen unterteilt. Drei dieser Klassen weisen Erkrankungen der Retina auf, während die vierte Klasse Aufnahmen zeigen bei denen keine der drei Erkrankungen vorliegen. Die vorgestellte Methode baut auf den sogenannten tiefen faltenden Netze (CNN) auf. Eine Struktur, die sich aus faltenden Lagen, Aggregationsschichten und dichten Lagen bei Verwendung des sogenannten Dropouts sehr gute Ergebnisse liefert. Es wird gezeigt, dass der Lernerfolg des CNN erheblich gesteigert wird, wenn die Zusammensetzung des Datensatzes in Form von Gewichten so berücksichtigt wird, dass jede der Klassen während des Trainings gleich behandelt wird. Zudem kann der Lernerfolg des Netzes durch die Reduzierung der Dimension der dichten Lagen gesteigert. Aufgrund der langen Laufzeit eines Trainingsprozesses kann die Konfiguration des CNN nicht durch eine größer angelegte Optimierungsstudie optimiert werden, jedoch liefert die in diesem Bericht vorgestellte finale Struktur bereits sehr gute Ergebnisse und erreicht eine Gesamtgenauigkeit von 90 %.

Die Wahl eines CNN wird dadurch validiert, dass der Lernerfolg eines vollständig vernetzten neuronalen Netzes (DNN) auf dem selben Datensatz ermittelt wird. Zu diesem Zweck wird die Struktur des Netzes, sowie die Aktivierungsfunktionen und die Batch Größe optimiert. Es stellt sich dabei heraus, dass kleinere Batch Größen bessere Ergebnisse erzählen. Die höchste Genauigkeit, die jedoch erzielt werden kann, liegt bei 72 % und ist somit deutlich geringer als die des CNN. Daher wird deutlich, dass die Wahl eines CNN für die gewählte Aufgabenstellung OCT Aufnahmen der Retina zu klassifizieren am besten geeignet scheint.

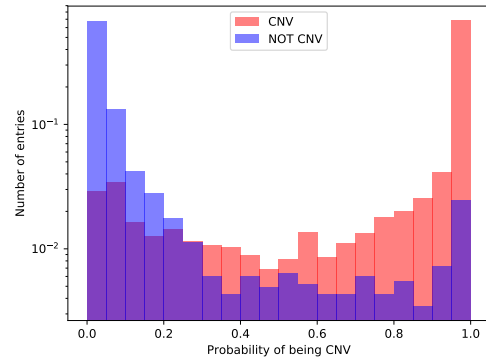
## Literatur

- [1] W. Drexler and J. Fujimoto, *Optical Coherence Tomography* (Springer, 2008).
- [2] Retinal OCT Images (optical coherence tomography), <http://www-app3.gfz-potsdam.de/obs/niemegk/monrep/wwhmean.html>, [Online; abgerufen am 31.07.2018].
- [3] D. C. Ciresan, U. Meier, and J. Schmidhuber, CoRR **abs/1202.2745** (2012), 1202.2745.
- [4] D. P. Kingma and J. Ba, CoRR **abs/1412.6980** (2014), 1412.6980.

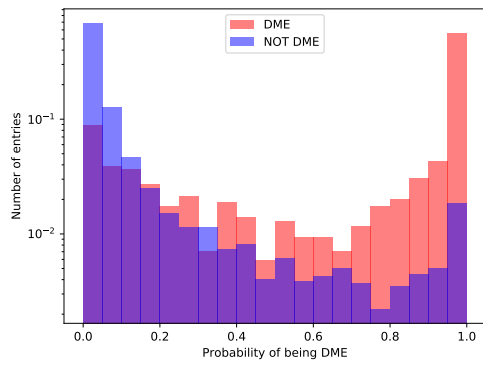
## A. Anhang



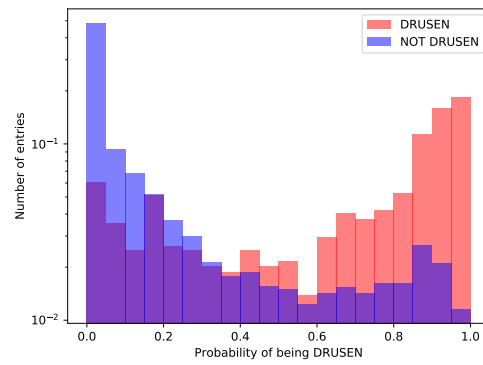
(a) Genauigkeit



(b) Verlustfunktion

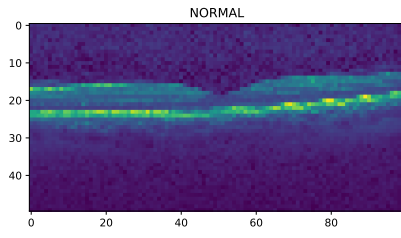


(c) Genauigkeit

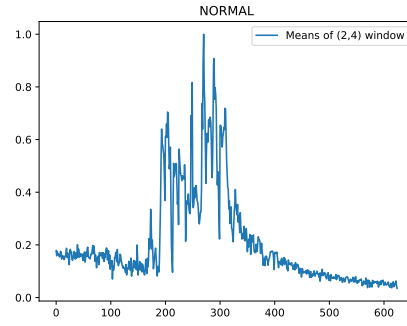


(d) Verlustfunktion

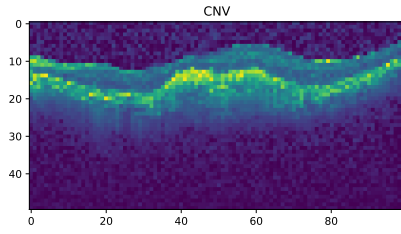




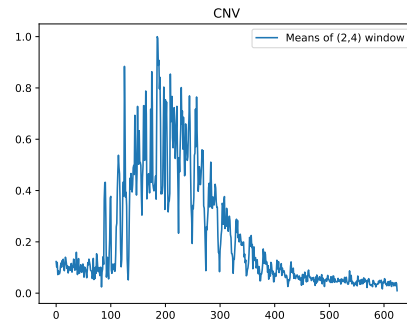
(e) Normal



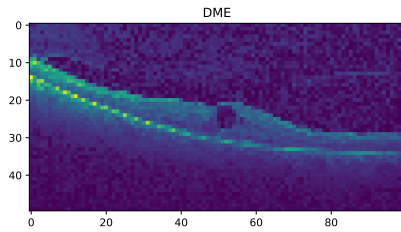
(f) Normal



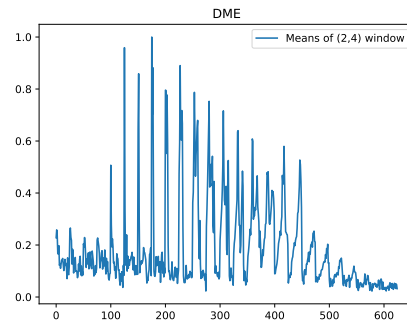
(g) CNV



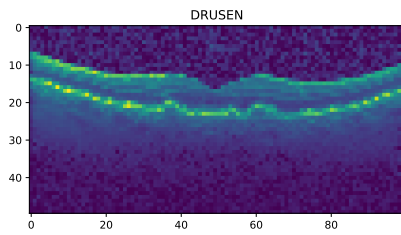
(h) CNV



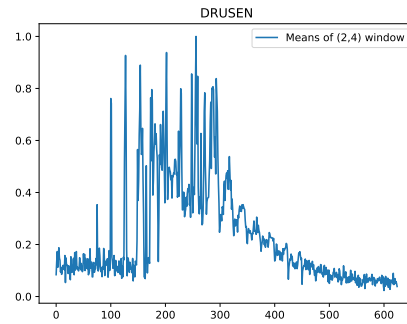
(i) DME



(j) DME



(k) DRUSEN



(l) DRUSEN

Abbildung 10: Beispielbilder für jede der Klassen im (50, 100) Format, sowie die aus der Abrasterung durch (2, 4) Fenster erhaltene Verteilung der Mittelwerte.