

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

**Klassifizierung durch Methoden des  
Maschinellen Lernens von  
Erkrankung der Retina mittels  
Aufnahmen der Optischen  
Kohärenztomographie**

Kevin Sedlaczek

kevin.sedlaczek@tu-dortmund.de

Björn Wendland

bjoern.wendland@tu-dortmund.de

Projektbericht vorgelegt von Björn Wendland

Datum der Abgabe: 31.07.2018

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Analysierter Datensatz</b>	<b>2</b>
<b>3. Methoden zur Klassifizierung von OCT Aufnahmen der Retina</b>	<b>3</b>
3.1. Tiefes faltendes neuronales Netz (CNN) . . . . .	3
3.2. Flaches tiefes neuronales Netz (DNN) . . . . .	9
<b>4. Zusammenfassung und Schlussfolgerungen</b>	<b>13</b>
<b>A. Anhang</b>	<b>15</b>

# 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 entstehenden Teilstrahlen trifft auf einen Spiegel, an dem er reflektiert wird, und wird als Referenzstrahl benutzt. Der andere Teilstrahl wird auf die zu untersuchende Probe gesendet. Dringt dieser Probenstrahl in die Probe ein, so kann er an den Probenschichten reflektiert werden.

Trifft der reflektierte Probenstrahl mit dem Referenzstrahl am semitransparenten Spiegel wieder zusammen, interferieren die beiden Teilstrahlen, wobei die Form des Interferenzmusters von der Probentiefe abhängt, bei der die Reflexion des Probenstrahls stattfindet, und von einem Detektor analysiert wird. Während der Analyse überlagern sich mehrere Signale, die durch Reflexion in unterschiedlichen Probentiefen entstehen. Durch Fourier-Transformation werden diese Signale getrennt, wodurch die komplette Information über das Tiefenprofil der Probe extrahiert werden kann.

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. Drei dieser Klassen beinhalten Aufnahmen von erkrankten Retinas. Eine der Krankheiten ist die choroidale Neovaskularisation (CNV), bei der sich irreguläre Blutgefäße in der Retina bilden, aus welchen Flüssigkeiten treten und somit zu Schwellungen der Retina führen können. Drusen (DRUSEN) bezeichnen Abfallprodukten aus der zentralen Retina, die sich unterhalb der Retina ansammeln. Das diabetische Makulaödem (DME) ist eine Schwellung der Makula, die im Zuge der Bildung von kleinen Gefäßausbuchtungen in der Retina entsteht. Die vierte Klasse (NORMAL) beinhaltet Retina Aufnahmen, die keine dieser Erkrankungen aufweisen.

Im Bereich der Bildklassifizierung gibt es bereits zahlreiche Beispiele, bei denen Methoden des maschinellen Lernens eingesetzt werden und gute Ergebnisse erzielen. In dem Projektbericht wird eine Methode basierend auf überwachtem, maschinellern 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 (DNN) zurückgreift, welches aus flachen dichten Lagen besteht.

## 2. Analysierter 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	NORMAL	CNV	DME	DRUSEN	Total
Anzahl	26315	37205	11348	8616	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 erkennbar ist. Auch DRUSEN sind in Abbildung 1(d) gut sichtbar, 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 vielen Aufnahmen ist dies jedoch nicht der Fall, sodass zur Klassifizierung eine andere Methodik oder Erfahrung in der Auswertung dieser Aufnahmen notwendig ist.

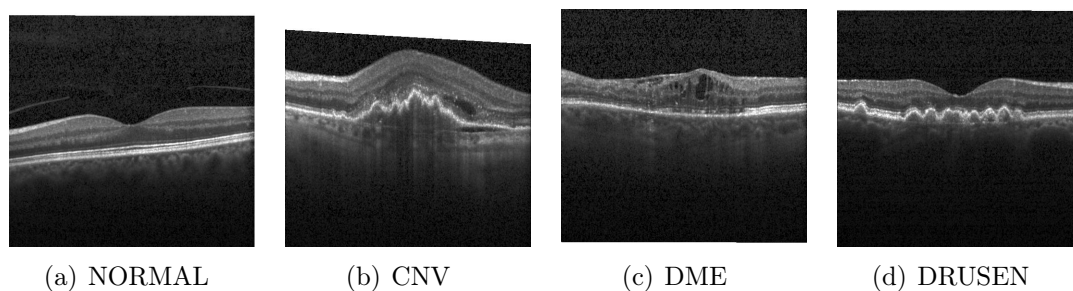


Abbildung 1: OCT Aufnahmen der Retina, die beispielhaft eine Aufnahme jeder untersuchten Klasse im Datensatz zeigen. Dabei zeigt Abbildung 1(a) eine OCT Aufnahme der Klasse NORMAL, die von keiner der Krankheiten, CNV, DME und DRUSEN, aufweist, welche in den Abbildungen 1(b), 1(c) respektive 1(d) dargestellt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

## 3. Methoden zur Klassifizierung von OCT Aufnahmen der Retina

### 3.1. Tiefes faltendes neuronales Netz (CNN)

Die Klassifizierung von OCT Aufnahmen der Retina stellt eine Aufgabe der Bildklassifizierung dar. Für die Bildklassifizierung durch maschinelles Lernen 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 des verwendeten Datensatzes 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 alle Pixelwerte des Datensatzes auf den Wertebereich  $[0, 1]$  linear transformiert.

Demzufolge erhält das CNN die Pixelwerte in Form einer zweidimensionalen Liste der Dimension  $(400 \times 400)$  als Eingangswerte. Ein 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 zweidimensionale Neuronenstruktur in festgelegten Schrittweiten abrastert. Bei jedem Schritt wird über eine diskrete Faltung die Aktivität jedes Neurons, das innerhalb des Kernels liegen, berechnet. Die Addition der Aktivitäten der einzelnen Neuronen ergibt den Ausgangswert für jeden Schritt. Anschaulich bedeutet dies, dass eine  $(400 \times 400)$  Neuronenstruktur, die beispielsweise mit einem  $(2 \times 2)$  Kernel in Schrittweiten der Form  $(2, 2)$  abgerastert wird, eine Ausgangsstruktur der Form  $(200 \times 200)$  generiert. Die Anzahl der Kernel, die zur Abrastierung 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 Fensters zu einem Neuron zusammenfasst. Wird beispielsweise ein  $(2 \times 2)$  Fenster definiert, welches eine  $(200 \times 200)$  Neuronenstruktur ohne Überlappung der einzelnen Schritte abfährt, werden in jedem Schritt vier Neuronen zu einem zusammengefasst, sodass sich eine  $(100 \times 100)$  Ausgangsstruktur ergibt. In dieser Arbeit wird bei jedem Schritt das Neuron mit der höchsten Aktivität behalten, während die übrigen Neuronen innerhalb eines Fensters verworfen werden. Demnach wird

durch Pooling Lagen die Dimension der Neuronenstruktur und somit die Laufzeit verringert, was in den meisten Fällen jedoch keine Verschlechterung des Lernerfolgs des CNN nach sich zieht. Durch das Entfernen von Neuronen geringerer Aktivität bietet es zudem 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 korrekt identifiziert, wodurch die Genauigkeit auf anderen Datensätzen desselben Typs geringer ist.

Eine weitere Möglichkeit Übertraining zu vermeiden ist das Einbinden der sogenannten Dropout Lagen. Hierbei wird ein festgelegter Bruchteil (Dropout Rate) an Neuronen einer Lage zufällig ausgewählt und in einem Trainingsschritt verworfen.

Zumeist bestehen CNNs aus einer abwechselnden Struktur aus Conv2D Lagen und Pooling Lagen bis die Dimension der Neuronenstruktur hinreichend stark reduziert ist. Daraufhin werden die Ausgangswerte der letzten Lage dieser Struktur in eine eindimensionale Liste abgespeichert, welche als Flatten Lage bezeichnet wird. Diese Ausgangswerte werden vollständig vernetzten dichten Lagen übergeben, für welche die Anzahl an Neuronen definiert werden, 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 dichten 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.

Zur Klassifizierung der in diesem Projektbericht analysierten Aufnahmen wird zunächst eine Referenzstruktur gesucht, welche sich für das Training auf dem kompletten Datensatz eignet, da dieses Training sich sehr zeitaufwendig gestaltet. Hierzu werden auf einem kleineren Datensatz Strukturen getestet und die vielversprechendste als Referenzstruktur verwendet,

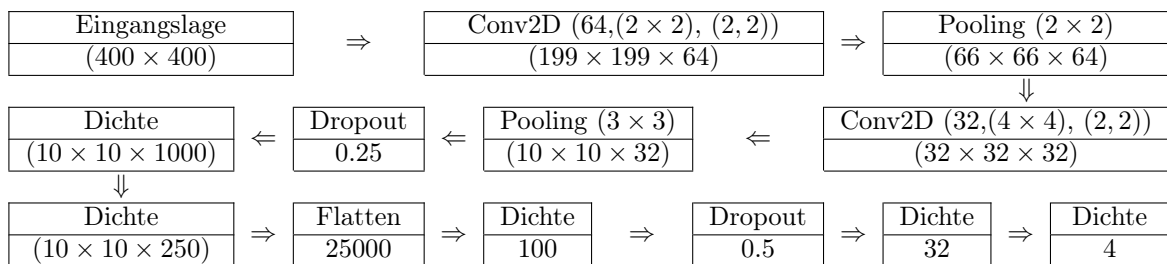


Abbildung 2: Schematische Darstellung der Referenzstruktur des CNN. Die Werte in den unteren Zeilen repräsentieren die Neuronenstruktur am Ausgang jeder Lage. Im Falle einer Dropout Lage stellt der Wert die festgelegte Rate dar. Für die Conv2D Lagen werden zudem die Anzahl an verwendeten Kernels sowie ihrer Dimension und die Schrittweite angegeben. Für die Pooling Lagen werden die definierten Dimensionen der verwendeten Fenster angegeben.

welche in Abbildung 2 schematisch dargestellt ist.

Das CNN wird auf 70 % des Datensatzes trainiert, wobei der Lernerfolg durch die restlichen 30 % des Datensatzes validiert wird. Bei der Aufteilung wird die relative Zusammensetzung des Datensatzes in Tabelle 1 in beiden Teildatensätze beibehalten. Die sogenannte Batch Größe legt fest, wie viele Aufnahmen in einem Trainingsschritt durch das CNN propagiert werden. Eine Epoche des Trainings ist dann absolviert, wenn der komplette Trainingsdatensatz durch das CNN propagiert wurde. Das Training erfolgt im vorliegenden Fall über 40 Epochen mit einer Batch Größe von 100. Als Aktivierungsfunktion der versteckten Layer und Ausgangslage wird die elu (exponential linear unit) Funktion respektive softmax Funktion verwendet. Zudem wird die kategorische Kreuzentropie als Verlustfunktion und der Adam Optimierer [4] mit Standardparametern verwendet. Als Metrik wird die Genauigkeit verwendet, welche 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 bei Propagation des Validierungsdatensatzes validiert. Es ergeben sich die in Abbildung 3 dargestellten Werte für die Genauigkeit und Verlustfunktion nach jeder Epoche.

Anhand Abbildung 3(a) lässt sich erkennen, dass die Referenzstruktur bereits gute Ergebnisse liefert, da eine Genauigkeit auf dem Validierungsdatensatz von über 90 % erzielt wird. Aufgrund der Tatsache, dass sich weder die Genauigkeiten noch die Werte der Verlustfunktion auf dem Validierungsdatensatz stark von den entsprechenden Werten auf dem

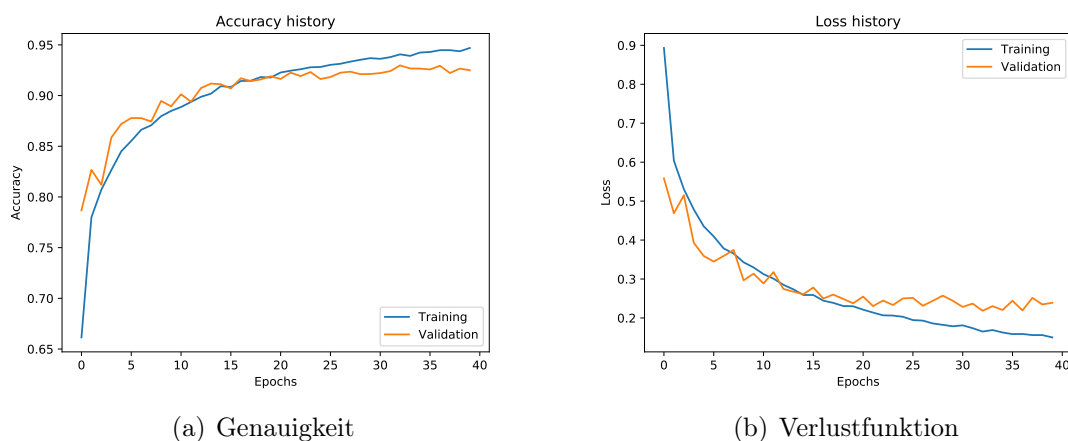


Abbildung 3: Genauigkeit 3(a) und Wert der Verlustfunktion 3(b) nach jeder Epoche des Trainings bei Propagation des Trainings- und Validierungsdatensatzes durch das CNN mit der Referenzstruktur.

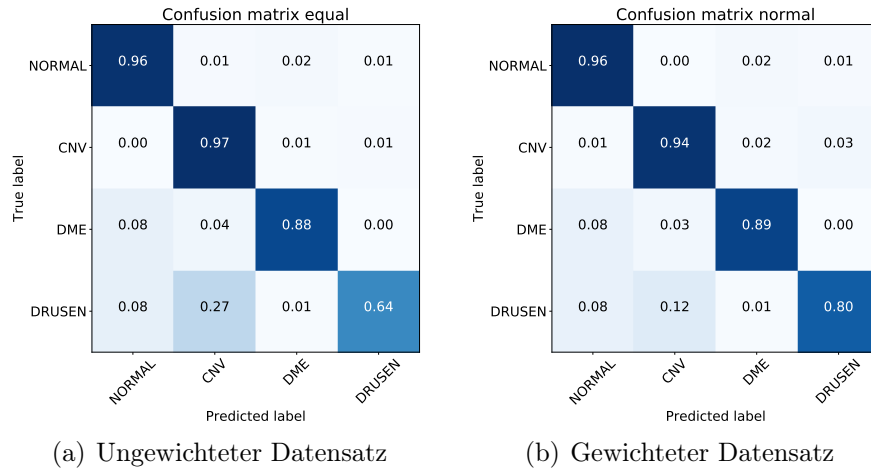


Abbildung 4: Verwirrungsmatrizen nach dem Training ohne Gewichtung des Datensatzes in Abbildung 4(a) und mit verwendeter Gewichtung in Abbildung 4(b).

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 des CNN dient zudem die Verwirrungsmatrix, welche die prozentuale Verteilung der Aufnahmen einer Klasse auf die durch das CNN 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 4(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 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 %. Da es gewünscht ist, dass das CNN für jede Klasse die höchstmögliche Genauigkeit erzielt, wird diese Genauigkeit im Folgenden als Maß für den Lernerfolg verwendet und als 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 dies im Training 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 4(b) dargestellte Verwirrungsmatrix. Die Gesamtgenauigkeit steigt hier auf 90 % und die Klasse DRUSEN wird deutlich besser klassifiziert und erreicht eine Gesamtgenauigkeit von 80 %, wobei die anderen Klassen weiterhin



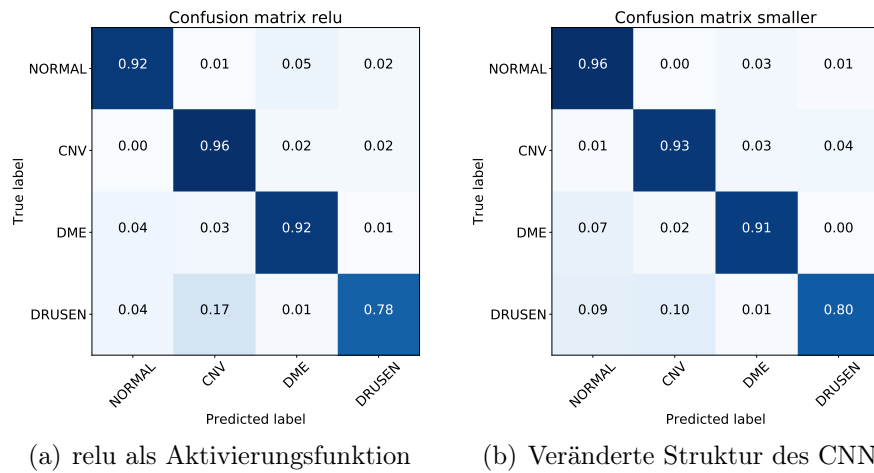


Abbildung 5: Verwirrungsmatrix bei Verwendung des gewichteten Datensatzes des CNN mit der Referenzstruktur mit relu als Aktivierungsfunktion der versteckten Lagen 5(a) und veränderter Struktur der dichten Lagen 5(b).

sehr gut klassifiziert werden. Daher werden im Folgenden stets 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 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 5(a) ist die resultierende Verwirrungsmatrix dargestellt, wenn relu (rectified linear unit) anstelle von elu verwendet wird.

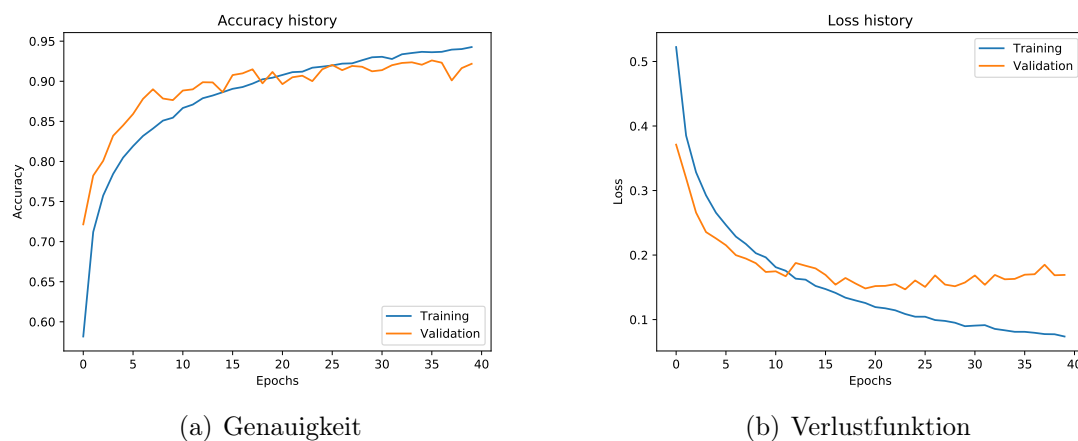


Abbildung 6: Genauigkeit 6(a) und Wert der Verlustfunktion 6(b) nach jeder Epoche des Trainings bei Propagation des Trainingsdatensatzes und Validierungsdatensatzes durch das CNN mit der angepassten Struktur der dichten Lagen.

Die Gesamtgenauigkeit ist mit 89 % etwas geringer, sodass keine Verbesserung erzielt werden kann.

Zudem wird festgestellt, dass die erste dichte Lage nach der Flatten Lage enorm viele Parameter aufweist, sodass die Größe der dichten Lagen vor der Flatten Lage angepasst werden. Die Lage mit 1000 Neuronen wird auf 256 Neuronen, die Lage mit 256 Neuronen auf 128 Neuronen reduziert, wodurch sich zudem die Trainingsdauer stark vermindert. Die resultierenden Genauigkeitswerte und Werte der Verlustfunktion nach jeder Epoche und die Verwirrungsmatrix sind in den Abbildungen 6(a), 6(b) respektive 5(b) dargestellt. Es lässt sich erkennen, dass sich die Genauigkeit leicht verbessert und 90 % beträgt. Somit wird diese Struktur als finale Struktur in dieser Projektarbeit verwendet.

Die Werte der Neuronen der Ausgangslage lassen sich als Wahrscheinlichkeiten für die

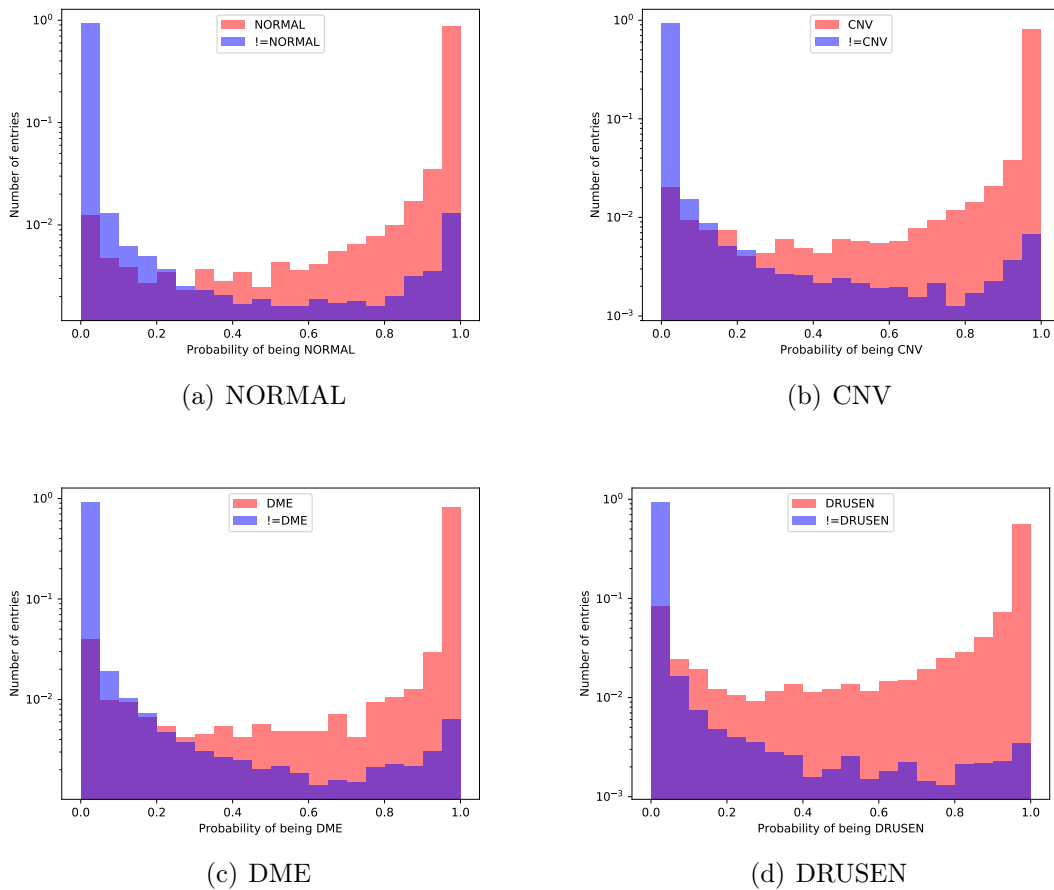


Abbildung 7: Verteilung der von dem CNN vorhergesagten Wahrscheinlichkeiten für die Aufnahmen innerhalb einer Klasse  $X$  und für die Aufnahmen, die nicht dieser Klasse angehören ( $\neq X$ ), der Klasse  $X$  anzugehören. Die Abbildungen 7(a), 7(b), 7(c) und 7(d) zeigen dies für die Klassen NORMAL, CNV, DME respektive DRUSEN.

Zugehörigkeit einer Aufnahmen zu einer Klasse interpretieren. Abbildung 7 zeigt die Verteilungen der Wahrscheinlichkeitswerte für die Aufnahmen innerhalb einer Klasse  $X$  und für die Aufnahmen außerhalb einer Klasse  $X$  ( $\neq X$ ) der Klasse  $X$  anzugehören. Dies bedeutet anschaulich, dass im Idealfall die roten Verteilungen in Abbildung 7 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 7(d), dass für die Klasse DRUSEN die Wahrscheinlichkeitsverteilung der zu DRUSEN zugehörigen Aufnahmen im Vergleich zu den Abbildungen 7(a)-7(c) einen kleineren Ausschlag bei dem Wert 1 und eine deutlich höhere Flanke zu kleinen Werten aufweist, wodurch hier eine deutlich schwächere Diskriminierung vorhanden ist.

### 3.2. Flaches tiefes neuronales Netz (DNN)

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 eindimensionale Liste aus Werten umgewandelt werden. Dabei 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.

In ersten Studien auf einem kleineren Datensatz erweist es sich als vielversprechend die Dimension der Aufnahmen weiter zu verringern. Hierzu wird ein Fenster der Dimension  $(2 \times 4)$  definiert, welches die Aufnahmen ohne Überlappung der einzelnen Schritte abstrahiert und den Mittelwert der im Fenster liegenden Pixelwerte in jedem Schritt berechnet. Anschließend werden die so erhaltenen 625 Mittelwerte auf den größten Mittelwert einer Aufnahme skaliert und dem DNN als eindimensionale Liste übergeben. In den ersten Studien zeigte sich bei dieser Methode eine Genauigkeit von 70 % und gröbere Körnungen der Aufnahmen sowie andere Fenstergrößen lieferten keine Verbesserungen, wodurch diese Methodik 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 getestet. Es werden 5 Grundstrukturen des DNN definiert, welche in Tabelle 3 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 in Tabelle 3 dargestellt sind.

Für jede der fünf Grundstrukturen werden die Aktivierungsfunktionen elu und relu sowie

softmax und sigmoid für die versteckten Lage respektive Ausgangslage getestet. Zudem werden die Batch Größen 50, 100, 128, 256 und 512 für jede der sich ergebenden DNN Konfigurationen eingestellt. Insgesamt werden somit 120 Konfigurationen für das Training des DNN überprüft.

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 % respektive 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 8 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 der 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 die optimale Konfiguration zu wählen, 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 Funktion als Aktivierungsfunktion der Ausgangslage bei einer Batch Größe von 50 und beträgt 72 % auf dem Validierungsdatensatz. Die Werte der Verlustfunktion nach jeder Epoche sind in Abbildung 9(a) dargestellt.

Anhand Abbildung 9 ist zudem ersichtlich, dass kein signifikantes Übertraining vorhanden

Tabelle 3: Getestete Grundstrukturen des DNN. Die Anzahl an Neuronen in der  $i$ -ten versteckten Lage und die Dropout Rate in der  $i$ -ten Dropout Lage sind durch das  $i$ -te Element der entsprechenden List dargestellt. Auf die  $i$ -te versteckte Lage folgt dabei die  $i$ -te Dropout Lage.

	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)

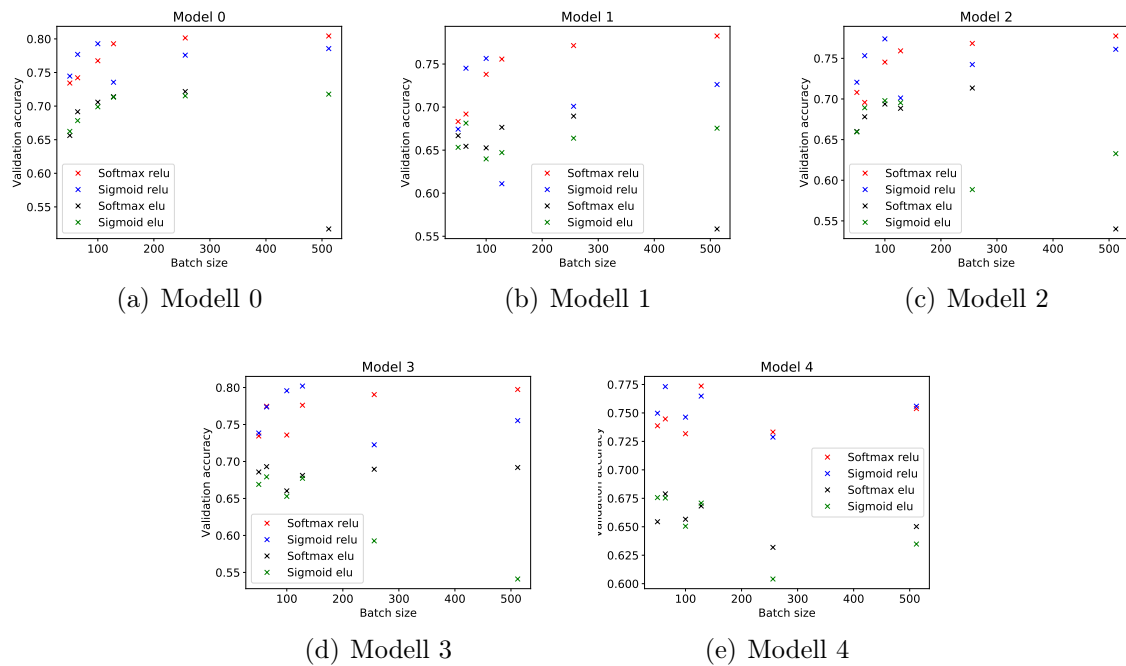


Abbildung 8: Erhaltene Genauigkeiten des DNN für die verschiedenen Konfigurationen der einzelnen Modelle als Funktion der Batch Größe.

ist und die Genauigkeit bereits einen Sättigungswert erreicht hat. In Abbildung 9(b) ist die entsprechende Verwirrungsmatrix dargestellt, welche auf dem Testdatensatz 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.

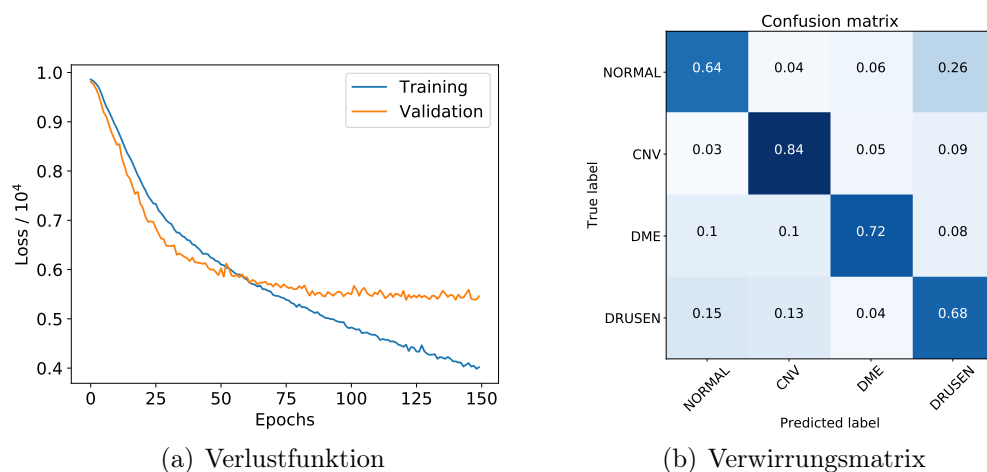


Abbildung 9: Wert der Verlustfunktion nach jeder Epoche 9(a) bei Propagation des Trainings- und Validierungsdatensatzes und Verwirrungsmatrix 9(b) bei Propagation des Testdatensatzes für die gewählte DNN Struktur .

## 4. Zusammenfassung und Schlussfolgerungen

In dem vorliegenden Projektbericht wird eine Methode des Maschinellen Lernens zur Klassifizierung von Erkrankungen der Retina vorgestellt, indem Aufnahmen der optischen Kohärenztomographie (OCT) analysiert werden. Die Aufnahmen beinhalten dabei drei unterschiedliche Erkrankungen der Retina und Aufnahmen, bei denen keine dieser Erkrankungen vorliegt.

Die vorgestellte Methode verwendet ein tiefes faltendes neuronales Netz (CNN), wobei eine Struktur gewählt wird, die sich aus faltenden Lagen, Aggregationsschichten und vollständig vernetzten dichten Lagen bei Verwendung des sogenannten Dropouts sehr gute Ergebnisse mit einer Genauigkeit von 90 % liefert. Es wird gezeigt, dass der Lernerfolg des CNN erheblich gesteigert wird, wenn die Zusammensetzung des Datensatzes in Form von Gewichten berücksichtigt wird, sodass jede der Klassen während des Trainings gleich behandelt wird. Das Training des CNN gestaltet sich jedoch als sehr zeitaufwendig, sodass eine vollständige Optimierung des Netzes nicht vollzogen wird. Es wird jedoch gezeigt, dass die Reduzierung der Dimension der dichten Lagen eine etwas verbesserte Genauigkeit liefert. Zudem wird dadurch die Trainingsdauer reduziert, sodass in weiterführenden Studien eine weitere Reduktion der Dimension dieser Lagen vielversprechend erscheint.

Darüber hinaus wird festgestellt, dass die elu (exponential linear unit) als Aktivierungsfunktion eine etwas bessere Genauigkeit liefert als die relu (rectified linear unit) Funktion. Beim Training über 40 Epochen wird kein Übertraining festgestellt, sodass eine Optimierung in Hinblick auf die Verminderung des Übertrainings nicht erfolgen muss. Außerdem kann die Dimension vor der ersten flachen vollständig vernetzten dichten Lage noch weiter reduziert und der Lernerfolg eventuell gesteigert werden, indem weitere Conv2D oder Pooling Lagen hinzugefügt werden.

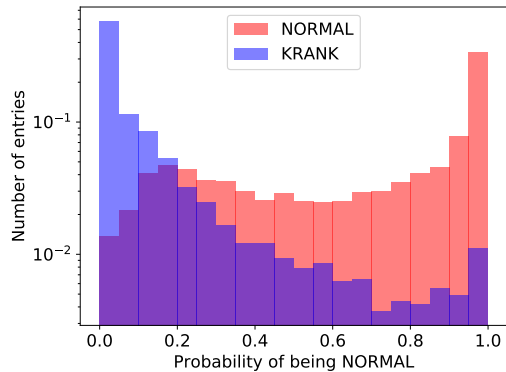
Die Wahl eines CNN wird dadurch validiert, dass der Lernerfolg eines optimierten flachen neuronalen Netzes (DNN) auf dem selben Datensatz ermittelt wird. Zu diesem Zweck wird die Struktur des Netzes, sowie die Aktivierungsfunktionen für die versteckten Lagen und die Ausgangslage sowie die Batch Größe optimiert. Es stellt sich dabei heraus, dass kleinere Batch Größen bessere Ergebnisse erzielen. Die höchste Genauigkeit, die erzielt werden kann, liegt hier 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 besser geeignet ist als ein DNN.

Abschließend lässt sich demnach schlussfolgern, dass die hier gewählte Struktur des CNN im Rahmen dieser Projektarbeit mit einer Genauigkeit von 90 % ein zufriedenstellendes Ergebnis liefert.

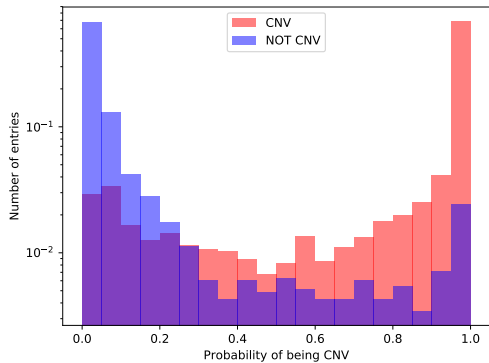
## Literatur

- [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/whhmean.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.

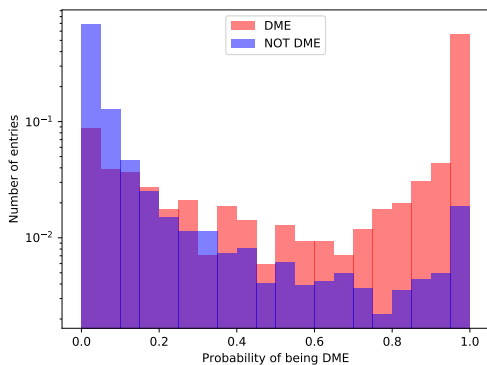
A. Anhang



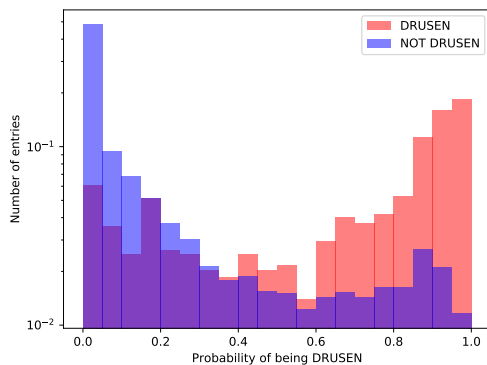
(a) Genauigkeit



(b) Verlustfunktion

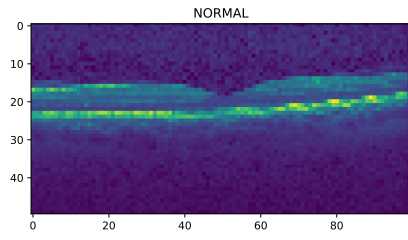


(c) Genauigkeit

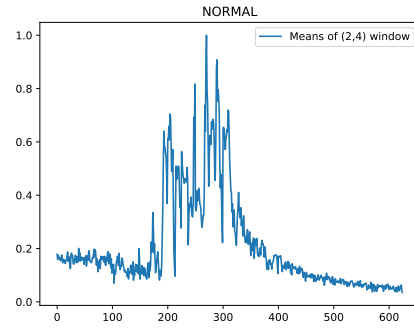


(d) Verlustfunktion

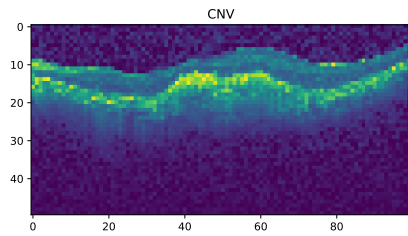




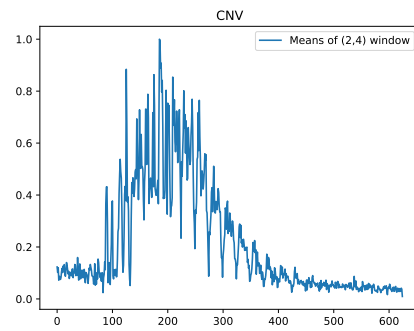
(e) Normal



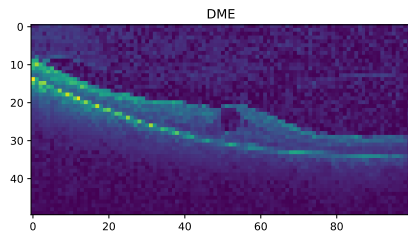
(f) Normal



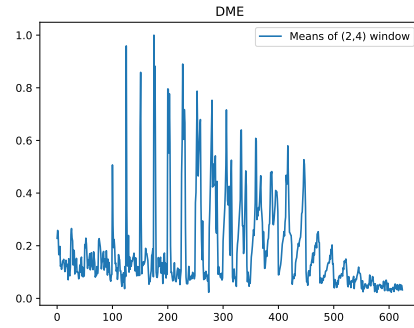
(g) CNV



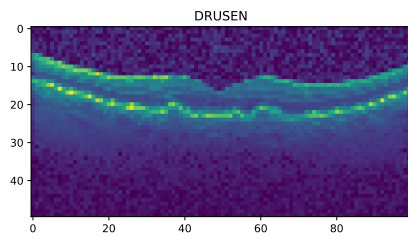
(h) CNV



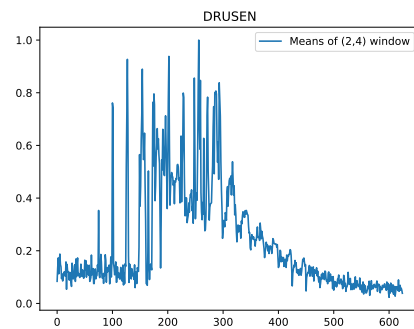
(i) DME



(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.