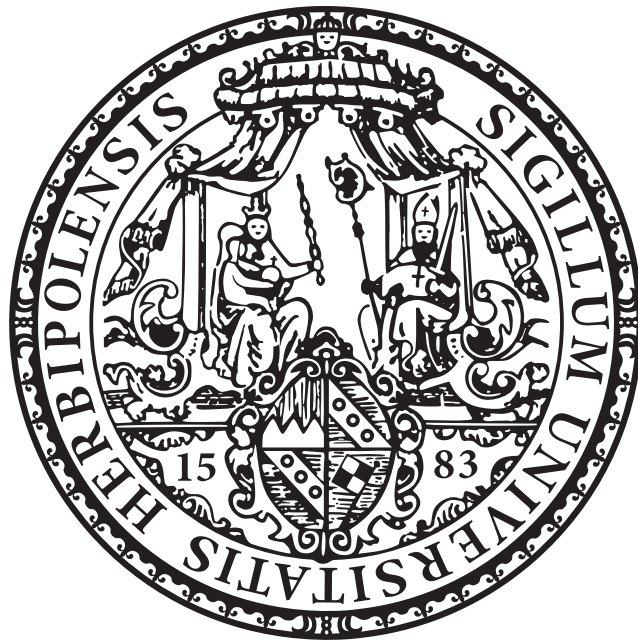


# Automatisierte Klassifizierung und Viabilitätsanalyse von Phytoplankton



Dissertation zur Erlangung des  
naturwissenschaftlichen Doktorgrades  
der Julius-Maximilians-Universität Würzburg

vorgelegt von  
Katja Schulze  
aus Berlin

Würzburg, 2014

Eingereicht am: .....

Mitglieder der Promotionskommission:

Vorsitzender:	Prof. Dr. Markus Engstler
1. Gutachter:	Prof. Dr. Thomas Dandekar
2. Gutachter:	Prof. Dr. Marcus Frohme

Tag des Promotionskolloquiums: .....

Doktorurkunde ausgehändigt am: .....

# Danksagung

Prof. Dr. Marcus Frohme danke ich für die Ermöglichung dieser Arbeit in einer sehr offenen und kollegialen Arbeitsatmosphäre. Seine Unterstützung und Ratschläge waren immer eine wertvolle Hilfe.

Prof. Dr. Thomas Dandekar danke ich für immer wieder bereichernde Hinweise und die Bereitschaft, die Promotion an der Universität Würzburg zu ermöglichen.

Der gesamten Arbeitsgruppe für molekulare Biotechnologie und funktionale Genomik an der TH Wildau danke ich für die vielen hilfreichen Diskussionen und das wunderbare Arbeitsklima.

Allen Kooperationspartnern im Projekt danke ich für die Möglichkeit der Zusammenarbeit.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
1.1	Plankton - Einordnung des Begriffs . . . . .	3
1.2	Rolle des Phytoplanktons bei der Einschätzung von Gewässern . . . . .	5
1.2.1	Konventionelle Methoden . . . . .	6
1.2.2	Automatisierung der Analyse . . . . .	7
1.3	Fluoreszenzcharakteristika des Phytoplanktons . . . . .	9
1.4	Bestehende Methoden zur Viabilitätsanalyse . . . . .	10
1.5	Bildverarbeitung und Bilderkennung in der Biologie . . . . .	11
<b>2</b>	<b>Zielstellung</b>	<b>14</b>
<b>3</b>	<b>Material und Methoden</b>	<b>15</b>
3.1	Materialien . . . . .	15
3.1.1	Verbrauchsmaterialien . . . . .	15
3.1.2	Biologisches Ausgangsmaterial . . . . .	16
3.1.3	Geräte . . . . .	17
3.1.4	Software . . . . .	18
3.2	Methoden . . . . .	19
3.2.1	Mikrobiologische Methoden . . . . .	19
3.2.2	Automatisierte Mikroskopie . . . . .	20
3.2.3	Methoden der Bildverarbeitung und Bilderkennung . . . . .	22
<b>4</b>	<b>Ergebnisse</b>	<b>47</b>
4.1	Automatisierte Identifikation von Süßwasser - Phytoplankton . . . . .	47
4.1.1	Automatisierte Mikroskopie . . . . .	49
	Hellfeld . . . . .	49

Fluoreszenz . . . . .	50
4.1.2 Bildverarbeitung und Bilderkennung . . . . .	52
Adaption von ImageJ . . . . .	52
Segmentierung . . . . .	53
Merkmalsauswahl und Klassifikation mit neuronalen Netzen .	53
Vergleich mit weiteren Klassifikationsverfahren . . . . .	57
4.1.3 Vergrößerung des Testsets und Integration weiterer Merkmale	61
4.2 Entwicklung eines Viabilitätsassays . . . . .	66
4.2.1 Einsatz der Fluoreszenzmikroskopie . . . . .	66
4.2.2 Automatisierung . . . . .	68
4.2.3 Reproduzierbarkeit . . . . .	70
4.2.4 Validierung der Methode . . . . .	71
4.2.5 Anwendungsbeispiel . . . . .	73
4.2.6 Kombination mit Absorptionsspektren zur Vitalitätsanalyse auf Zellebene . . . . .	75
<b>5 Diskussion</b>	<b>77</b>
5.1 Automatisierte Identifikation von Süßwasser - Phytoplankton . . . . .	77
5.1.1 Mikroskopie . . . . .	77
5.1.2 Bildverarbeitung und Bilderkennung . . . . .	79
5.1.3 Vergrößerung des Testsets und Integration weiterer Merkmale	85
5.2 Entwicklung eines Viabilitätsassays . . . . .	88
5.3 Fazit und Ausblick . . . . .	92
<b>Zusammenfassung</b>	<b>97</b>
<b>Abstract</b>	<b>99</b>
<b>Publikationen</b>	<b>101</b>
<b>Erklärung</b>	<b>102</b>

# Abkürzungsverzeichnis

°C Grad Celsius

**3NBBM+V** Bolds Basal Medium mit Vitaminen und 3fach Nitrat

**Abb.** Abbildung

**API** *Application Programming Interfaces*

**BG11** Medium für Blau Grün Algen

**CEDD** *Color and Edge Directivity Descriptor*

**CHIRP** *Composite Hypercubes on Iterated Random Projections*

**CHL** Chlorophyll

**dB** Dezibel

**DCT** *Discrete Cosine Transformation*

**EFD** Elliptische Fourier Deskriptoren

**ErbsMS** Erdwasser Medium aus Erbse und lehmiger Erde mit Sand

**ES** Erddekot und Salze

**FAD** Flavin-Adenin-Dinukleotid

**FCTH** *Fuzzy Color and Texture Histogram*

**FP** Falsch Positiv

**FURIA** *Fuzzy Unordered Rule Induction Algorithm*

**g** mittlere Erdbeschleunigung

**GUI** *Graphical User Interface*

**HSB** *Hue Saturation Brightness*

**IDL** *Interactive Data Language*

**KbE** Kolonie bildende Einheiten

**µl** Mikroliter

**LBP** *Local Binary Patterns*

**min** Minute

**ml** Milliliter

**NADP** Nicotinsäureamid-Adenin-Dinukleotid-Phosphat  
**NEAT** *NeuroEvolution of Augmenting Topologies*  
**nm** Nanometer  
**OD** Optische Dichte  
**PC** Phycocyanin  
**PE** Phycoerythrin  
**PLASA** *PLAnkton Structure Aalysis*  
**QFF** *Quick Full Focus*  
**RGB** Rot-Grün-Blau  
**RIPPER** *Repeated Incremental Pruning to Produce Error Reduction*  
**rpm** Umdrehungen pro Minute (*revolutions per minute*)  
**s** Sekunde  
**sp.** Spezies  
**SVM** Support Vektor Maschine  
**Weka** *Waikato Environment for Knowledge Analysis*

# 1 Einleitung

Biologische Phänomene lassen sich heute durch eine Vielzahl von mikroskopischen Verfahren und Methoden besser denn je abbilden. Mittlerweile ist es in fast jedem Labor möglich, mikroskopische Bilder auf Rechnern gut darzustellen und zu speichern. Dabei können Daten in vielfältiger Weise als Stillbilder, Videos, 3D Volumenbilder oder sogar 4D Volumenvideo aufgenommen werden. Zudem können über Fluoreszenzmethoden verschiedenste Probenbestandteile sichtbar gemacht und besser voneinander differenziert werden.

Die Auswertung dieser Daten gestaltet sich jedoch häufig als sehr monoton und nimmt viel Arbeitszeit in Anspruch. Großer Nachteil der Auswertung von Hand ist auch eine oft subjektive Einschätzung der Daten. Dabei ist die Qualität der Analyse stark von den Fähigkeiten und Vorlieben einzelner Personen abhängig. Zur Vereinfachung von mikroskopischen Analysen, Verbesserung derer Reproduzierbarkeit und Eingrenzung von Fehlerquellen kommen deshalb in der Biologie zunehmend Methoden der Bildverarbeitung und Bilderkennung zum Einsatz.

Diese Probleme gelten auch für Phytoplankter, die häufig über mikroskopische Methoden charakterisiert und analysiert werden. Im Rahmen dieser Arbeit wurden dabei Fragestellungen bearbeitet, die es mit Hilfe von automatischer Mikroskopie, Bildverarbeitung und Bilderkennung möglich machen sollen, deren Analyse zu vereinfachen und zu verbessern.

Im ersten Abschnitt der Arbeit wird hierfür der Aufbau eines Systems zur automatisierten Analyse von Phytoplanktongemeinschaften dargestellt. Diese liefern, aufgrund ihrer Diversität und der Veränderung von Populationen in Abhängigkeit von den vorherrschenden Lebensbedingungen, einen wichtigen Parameter für die Bewertung von Wasserqualitäten. Dabei ist die heute als Standardmethode durchgeführte Analyse von Hand sehr arbeitsintensiv, mit viel Spezialwissen verknüpft und fehleranfällig.



In der derzeitigen Analyse von Phytoplanktern ist eine Bestimmung des viablen Anteils an Organismen sehr komplex. Bestehende Methoden sind entweder sehr zeintensiv oder ermöglichen keinen Einblick auf Organismusebene, sondern nur einen Durchschnitt aller Organismen. Eine Analyse auf der Zellebene kann jedoch einen wertvollen Faktor zur detaillierten Analyse von Populationsveränderungen und deren Ursachen darstellen. Auf Grundlage dieser Problematik wird im zweiten Teil der Arbeit die Entwicklung und Validierung einer Methode zur automatisierten und schnellen Analyse der Viabilität auf Zellebene dargestellt. Die Methode wurde im Rahmen dieser Arbeit zunächst für das Cyanobakterium *Synechocystis* sp. 6803 etabliert, ist grundsätzlich aber auch für andere Phytoplankter anwendbar. Als Mitglieder des Phytoplanktons werden Cyanobakterien zum einen bei der Bewertung der Wasserqualität erfasst. Zum anderen kommt ihnen ein gesteigertes Interesse für die Nutzung zur Biobrennstoffproduktion zu und geeignete Methoden zur Kulturüberwachung werden immer bedeutender.

## 1.1 Plankton - Einordnung des Begriffs

Der Begriff Plankton wurde erstmals vom deutschen Zoologen Victor Hensen verwendet. Hiermit sollte die Beschreibung einer Gesellschaft an Lebewesen, die zuvor von Johannes Mueller mit Hilfe eines feinen durchlässigen Netzes gefischt wurde, besser gelingen. Dabei ist der Begriff vom griechischen Wort planktos (altgr. *πλαγκτόν*) abgeleitet und bedeutet soviel wie driften. Hensens erste Definition dieser Gemeinschaft von Organismen lautete dabei wie folgt: „*Ich verstehe darunter Alles was im Wasser treibt, einerlei ob hoch oder tief, ob todt oder lebendig*“ (Hensen, 1887). Dabei wird impliziert, dass Plankter nicht in der Lage sind ihren Aufenthaltsort entgegen Strömungen und Eigenbewegungen des Wassers zu verändern. Dies grenzt sie vom Nekton ab, das seine Position im Wasser trotz größerer Strömungen selbst bestimmen kann (z.B. Fische oder Säugetiere) (Reynolds, 1984).

Die heutige Definition des Planktons orientiert sich dabei noch immer an dieser ersten Definition. Unter dem Begriff Plankton wird heute eine Gesellschaft an Lebewesen zusammengefasst, die meist relativ klein sind (Mikrometer - Zentimeter), im Wasser leben und mit dem Strom des Wassers driften. Hier kann noch zwischen Lebewesen unterschieden werden, die in ihrem gesamten Lebenszyklus dem Plankton zugeord-

net werden (Holoplankton) und Lebewesen, die nur einige Zyklen ihres Lebens als Plankter verbringen (Meroplankton).

Eine weitere Einteilung erfolgt über die trophische Rolle der Organismen. Hier werden die im Folgenden näher erläuterten drei Gruppen unterschieden:

### **Phytoplankton**

Unter diesem Begriff werden alle photosynthetisch aktiven Organismen zusammengefasst, die die oberen, sonnendurchstrahlten Gewässerlagen besiedeln. Aufgrund ihrer photosynthetischen Aktivität bauen sie über die CO<sub>2</sub>-Fixierung mit Hilfe von Sonnenenergie Biomasse auf und sind die Primärproduzenten in Gewässern. Der Hauptanteil der Spezies im Phytoplankton wird von Kieselalgen, Grünalgen, Goldalgen, Dinoflagellaten, Cyanobakterien, Cryptomonaden und Haptophyta gebildet (Graham et al., 2009).

### **Zooplankton**

Diese Gruppe von Planktern fasst alle Organismen zusammen, die andere Planktonorganismen als Nahrungsquelle nutzen. Dabei kann zwischen Planktern, die sich direkt von Phytoplankton ernähren (Herbivore) und Planktern, die sich von anderem Zooplankton ernähren (Carnivore) unterschieden werden. Sie alle bilden die Nahrungsquellen von Konsumenten mit höherem trophischen Level (Lalli und Parsons, 1997).

### **Bakterioplankton**

Unter dieser Gruppe werden Bakterien zusammengefasst, die entweder frei im Wasser oder an Detritus bzw. andere Plankter angeheftet vorkommen. Vorrangig handelt es sich hierbei um heterotrophe Bakterien, welche ihre Energie über die Zersetzung von organischen Materialien beziehen und somit eine wichtige Rolle in der Remineralisierung spielen (Sommer, 2005).

Bei der Einteilung über die trophische Rolle kann nicht immer eine klare Trennlinie gezogen werden. Als Grenzfall existieren mixotrophe Plankter, die Photosyn-

these betreiben können und sich auch von anderen Organismen ernähren können. Beispiel sind einige Dinoflagellaten, die in Abhängigkeit von den vorherrschenden Bedingungen phototroph oder heterotroph leben können (Nabors, 2007).

Zudem findet eine Einteilung nach Größe statt, die in den Bereichen von mehreren Zentimetern und Bruchteilen von Mikrometern liegt (Tabelle 1.1).

Tabelle 1.1: Einteilung des Planktons nach Organismusgröße nach Sommer (2005).

<b>Bezeichnung</b>	<b>Größenordnung</b>	<b>Organismen</b>
Megaplankton	> 20 mm	größte Zooplankter bspw. Quallen
Macroplankton	2-20 mm	große Zooplankter bspw. Fischlarven
Mesoplankton	0,2-2 mm	“klassische“ Zooplankter bspw. Ruderfußkrebse, größte Einzeller, kolonienbildende Phytoplankter bspw. Volvox
Mikroplankton	20-200 µm	Großteil der Phytoplankter, Protozoen
Nanoplankton	2-20 µm	Phytoplankter, Protozoen
Picoplankton	0,2-2 µm	Bakterien, kleinste Phytoplankter
Femtoplankton	< 0,2 µm	Viren, Phagen

Da Plankton sowohl in Salz- als auch Süßwasser anzutreffen ist und hier deutlich unterschiedliche Lebensbedingungen vorliegen, wird darüber hinaus zwischen marinem Plankton und Süßwasserplankton differenziert.

## 1.2 Rolle des Phytoplanktons bei der Einschätzung von Gewässern

Der Schutz von Gewässern und die damit verbundene ökologische Bewertung wird über die Europäischen Wasserrahmenrichtlinie 2000/60/EG geregelt. Dabei werden Daten über die biologischen Komponenten Phytoplankton, Phytobenthos, Makrophyten, Makrozoobenthos sowie Fische neben hydrochemischen und gewässerstrukturellen Parametern zur Analyse genutzt (Europäische Union, 2000). Dem Phytoplankton kommt, neben dem Phytobenthos und den Makrophyten, eine wichtige Rolle zu, da hier die Zusammenhänge zwischen Artzusammensetzung eines Gewässers und dem Vorkommen von Indikatorarten, sowie dem ökologischen Zustand von Gewässern am intensivsten erforscht und am eindeutigsten hergestellt sind. Die Analyse des

Phytoplanktons hat zudem die Vorteile, dass die Taxonomie weit entwickelt ist und dass vergleichsweise geringe Kosten durch die Probennahme und Analytik entstehen. Dabei müssen die Phytoplanktonabundanz und -zusammensetzung sowie Blütenfrequenz und Blütenintensität erfasst werden (Euopäische Union, 2000; Solimini et al., 2006). Die erfassten Werte werden vorrangig zur Ermittlung der Eutrophierung der Gewässer eingesetzt (Carvalho et al., 2013). Zudem sind aber auch die Bestimmungen von organischen Verschmutzungen und Versauerungen möglich (Maileht et al., 2013).

Da die Planktongemeinschaften regionalen Veränderungen unterliegen, müssen nationale Anpassungen für die Analyse erfolgen. In der Überwachung von deutschen Gewässern wird der Parameter Phytoplankton zur Bewertung von Seen sowie planktondominierten Fließgewässern eingesetzt (Mischke et al., 2005; Mischke, 2006; Mischke und Nixdorf, 2008; Nixdorf et al., 2006; Mischke et al., 2011). Für eine Standardisierung der Phytoplanktonanalyse wurden dabei in nationalen Studien Taxa identifiziert, die als Indikator für unterschiedliche Gewässerqualitäten genutzt werden können. Diese wurden mit einer jeweils festgelegten Mindestbestimmungstiefe als Indikator taxa für stehende Gewässer und Fließgewässer in der harmonisierten Taxaliste für Phytoplankton zusammengefasst (Mischke und Kusber, 2006).

### **1.2.1 Konventionelle Methoden**

Bei der klassischen Analyse von Gewässern werden zur Probengewinnung Wasserschöpfer verwendet. Anschließend wird die Probe mit Lugol'scher Lösung fixiert, um eine Veränderung der Probenzusammensetzung zu verhindern. Nach dem Utermöhl-Verfahren werden darin befindliche Phytoplankter qualitativ und quantitativ erfasst (Hoehn, 1998; Utermöhl, 1958). Dabei werden die Phytoplankter in einer Sedimentationskammer abgesetzt und anschließend im Umkehr-Mikroskop mikroskopiert. Für die Bestimmung der Wasserqualität müssen hierbei pro Probe mindestens 400 Organismen erfasst werden (Mischke et al., 2005).

Des Weiteren ist die Ermittlung der Biomasse des Phytoplanktons notwendig, welche durch das Ausmessen der Organismen mit einem Meßokular erfolgt (Hoehn, 1998). Nach Auszählung, Ermittlung der Biomasse und Berücksichtigung der jeweiligen biologischen Gewichtung des ermittelten Phytoplanktons kann der Trophieindex

angegeben werden, welcher Aufschluss über die ökologische „Qualität“ gibt. Zur Beurteilung eines Gewässers werden jährlich je nach Gewässertyp mindestens fünf bis sieben Phytoplanktonproben entnommen (Mischke et al., 2005; Nixdorf et al., 2006). In alternativen Ansätzen werden Fluoreszenz-Mikroskope und Auflicht-Mikroskope verwendet. Fluoreszenz-Mikroskope finden z.B. bei speziellen Fragestellungen wie der Unterscheidung zwischen kleinen Blaualgen und kleinen Grünalgen bzw. Bakterien Anwendung. Zudem werden Auflicht-Mikroskope verwendet, um Phytoplanktonorganismen aus Membranfilterproben zu identifizieren. Dies wird vor allem dann angewandt, wenn Organismen mit Gasvakuolen bzw. schlechten Sedimentationseigenschaften identifiziert werden sollen (Hoehn, 1998).

Durch die hohe morphologische Komplexität des Phytoplanktons wird die mikroskopische Analyse derzeit noch von Hand durchgeführt. Neben dem hohen Zeit- und Arbeitsaufwand, der durch die notwendige Erfassung der einzelnen Organismen entsteht, können hier auch Probleme in der Reproduzierbarkeit der Ergebnisse bestehen. Diese können dabei zwischen verschiedenen Personen aber auch einer einzelnen Person auftreten (Culverhouse et al., 2003).

### **1.2.2 Automatisierung der Analyse**

Durch die Automatisierung von Analysen soll die Reproduzierbarkeit verbessert und der Arbeitsaufwand für den Menschen verringert werden. Für die Planktonidentifikation wurden dabei bereits verschiedene Ansätze verfolgt.

Der Einsatz eines Flowcytometers ermöglicht eine sehr schnelle Analyse verschiedener Phytoplanktongruppen und hat den Vorteil, dass die Fixierung der Proben unnötig wird (Cellamare et al., 2010; Dubelaar und Jonker, 2000). Nachteilig ist hier, dass die Organismen nur mit Hilfe von Lichtstreuungs- und Fluoreszenzsignalen unterschieden werden können, was vor allem bei Nano- und Mikroplankton zu einer schlechten Speziesauflösung führt (Dubelaar und Jonker, 2000; Erickson et al., 2012).

Ein weitere Möglichkeit stellt die parallele Analyse der gesamten Gemeinschaft über metagenomische Ansätze dar. Da diese Methode auf genomischen Daten basiert, hat sie großes Potential für eine hohe taxonomische Auflösung (Yan und Yu, 2011). Durch noch hohe anfallende Kosten ist dieser Ansatz für den Routineeinsatz derzeit jedoch nicht geeignet. Analog können auch moderne barcoding Methoden, die auf der

Sequenz und Struktur von genetischen Markern wie ribosomaler RNA basieren, für eine parallele Analyse genutzt werden (Keller et al., 2010).

Für den Einsatz zur Quantifizierung besteht dabei bei beiden Ansätze die Schwierigkeit, dass einzelne DNA oder RNA Moleküle nicht immer einem Organismus entsprechen und z.B. einzelne Organismen zum Teil auch mehrere Genome besitzen können. Zudem können auch Variationen in Abhängigkeit von verschiedenen Wachstumsphasen auftreten. Ferner müssen für eine Identifikation über die DNA sequenzierte Referenzgenome zur Verfügung stehen, was zur Folge hat, dass nur bereits erfasste Spezies in die Analyse integriert werden können.

Da die Taxonomie von Plankton auf morphologischen Unterschieden basiert, ist es nahe liegend die digitale Bildanalyse für eine Unterscheidung zu nutzen. Hier existiert eine Vielzahl von Systemen welche Scanner, Durchflusssysteme und Videosysteme umfassen (Benfield et al., 2007; Culverhouse, 2007). Diese wurden überwiegend für die Analyse von Zooplankton entwickelt und besitzen eine zu geringe Bildauflösung, um eine genaue Unterscheidung von Phytoplankton zu ermöglichen. Aufgrund ihrer hohen Auflösung sind Mikroskope besser für eine Analyse dieser Unterklasse des Planktons geeignet. Da sie auch vorrangig in der Routineuntersuchung eingesetzt werden, kann zudem auch eine gute Überprüfung von automatisiert gewonnenen Ergebnissen erfolgen. Ein publiziertes System ist das Plankton Struktur Analyse (PLASA) System, welches für die Unterscheidung von Phytoplankton in Rahmen einer ökotoxikologischen Studie entwickelt wurde. Es integriert automatisierte Mikroskopie der Proben und eine anschließende automatisierte Bildanalyse (Rodenacker et al., 2006). Für eine bessere Unterscheidung von verschiedenen Phytoplanktonarten und von Phytoplankton und anderen Partikeln in der Probe (wie Zooplankton, Detritus und anorganische Partikel) wurden Fluoreszenzaufnahmen von Chlorophyll und Phycoerythrin in das System integriert. Dabei wurden die Proben zur Erhaltung ihrer Fluoreszenzeigenschaften nicht wie üblich mit Lugol'scher Lösung, sondern mit Formaldehyd fixiert. Die Software wurde in IDL, einer proprietären Programmiersprache für die eine kommerzielle Lizenz benötigt wird, entwickelt. Dies hat möglicherweise die Weiterentwicklung des Systems verhindert, da bis heute nur eine Unterscheidung von fünf verschiedenen Taxa beschrieben wurde (Rodenacker et al., 2006).

### 1.3 Fluoreszenzcharakteristika des Phytoplanktons

Verschiedene Objekte emittieren bei der Anregung mit meist kurzwelligem Licht ohne weitere Präparation Fluoreszenzsignale. Ausgehend von verschiedenen Photopigmenten des Photosystems können solche Autofluoreszenzsignale auch bei phototrophen Organismen beobachtet werden. Photopigmente stehen dabei am Beginn der photosynthetischen Umwandlung von Lichtenergie in chemische Energie. Sie absorbieren Licht in definierten Wellenlängen und ermöglichen somit die Einkopplung der Energie in das Photosystem. Dabei kommt es immer auch zu Verlusten, die neben Wärme und Phosphoreszenz auch als Fluoreszenzsignal abgegeben werden. Um das Spektrum der Sonne möglichst gut ausnutzen zu können, sind verschiedene Photopigmente mit unterschiedlichen Anregungswellenlängen in das Photosystem integriert. Dabei können sie über entsprechend ausgewählte Anregungswellenlängen mit Hilfe der Fluoreszenzmikroskopie separat voneinander identifiziert werden (Tabelle 1.2) (X. Hu et al., 2010; Graham et al., 2009; Taiz und Zeiger, 2010).

Tabelle 1.2: Fluoreszenzeigenschaften der wichtigsten Photopigmente des Phytoplanktons nach X. Hu et al. (2010) und Graham et al. (2009)

<b>Pigment</b>	<b>Hauptanregungsbande (nm)</b>	<b>Hauptemissionsbande (nm)</b>
Chlorophyll a	ca. 430	680
Chlorophyll b	ca. 460	680*
Chlorophyll c	ca. 440	680*
Karotenoide	ca. 450 und 510	680*
Phycocyanin	ca. 620	630-680
Phycoerythrin	ca. 540	550-580

\* Akzessorische Chlorophyllmoleküle und Karotenoide transferieren die absorbierte Energie zum Reaktionszentrum von Chlorophyll a und besitzen keine Autofluoreszenz.

Die unterschiedlichen Klassen des Phytoplanktons weisen weiterhin unterschiedliche Zusammenstellungen in ihren Pigmentierungen auf (X. Hu et al., 2010; Beutler et al., 2002). Die Hauptpigmente der wichtigsten Vertreter des Süßwasser-Phytoplanktons sind für eine Übersicht in Tabelle 1.3 zusammengestellt. Die Karotenoide wurden

dabei aufgrund der großen Vielfalt nicht berücksichtigt. Da Chlorophyll a in allen Phytoplanktonorganismen vorhanden ist, kann es gut für die Abtrennung von restlichen Bestandteilen genutzt werden. Alle weiteren Pigmente ermöglichen eine Differenzierung zwischen einzelnen Gruppen.

Tabelle 1.3: Hauptphotopigmente der dominanten Vertreter des Süßwasser-Phytoplanktons (X. Hu et al., 2010; Beutler et al., 2002; Streble und Krauter, 2006)

<b>Algenklasse</b>	<b>Photopigmente</b>
Dinophyta (Dinoflagellaten)	Chlorophyll a & c
Bacillariophyta (Kieselalgen)	Chlorophyll a & c
Chrysophyceae (Goldalgen)	Chlorophyll a & c
Cyanobacteria (Blaualgen)	Chlorophyll a, Phyco cyanin, Phycoerythrin
Cryptophyta (Kryptomonaden)	Chlorophyll a & c, Phyco cyanin, Phycoerythrin
Chlorophyceae (Grünalgen)	Chlorophyll a & b,
Euglenophyta (Augenflagellaten)	Chlorophyll a & b
Xanthophyceae (Gelbgrünalgen)	Chlorophyll a

Neben den Autofluoreszenzsignalen der Photosysteme wurde von Tang und Dobbs (2007) eine unspezifische grüne Autofluoreszenz in verschiedenen Phytoplanktern beschrieben. Diese wird bei ähnlicher Wellenlänge wie die Autofluoreszenz des Chlorophylls angeregt. Allerdings ist der Effekt im Phytoplankton nur schlecht untersucht und die hierfür verantwortlichen Moleküle sind weitestgehend unbekannt. Ein ähnliches Phänomen wurde bereits in höheren Pflanzen beschrieben und wird dort von einer Vielzahl von unterschiedlichen Molekülen wie Flavonoiden (z.B. FAD), Zimtsäuren (z.B. Ferulasäure), Betaxanthinen und Pyridinnukleotiden (z.B. NAD(P)H) hervorgerufen (Hense et al., 2008). Inwiefern diese Moleküle auch im Phytoplankton eine Rolle spielen, ist noch nicht weiter beschrieben.

## 1.4 Bestehende Methoden zur Viabilitätsanalyse

Ein wichtiger Parameter zur Analyse von Phytoplanktonpopulationen oder aber auch von Monokulturen im Labor ist deren Anteil an viablen Zellen. Hiermit können Rück-



schlüsse auf optimale Kulturbedingungen und die vorhandenen Wachstumsbedingungen getroffen werden. Zudem können wertvolle Informationen über Stress, Zellalter und respiratorische Rate durch sich verändernde Bedingungen gewonnen werden.

Eine sehr einfache bestehende Methoden zur Fitnessanalyse sind Absorptionsspektren, die eine Übersicht über den Gesamtgehalt an Pigmenten und deren Verteilung in einer Kultur geben. Hierfür kann auch die Konzentration an Chlorophyll oder die Chlorophyllkonzentration/OD<sub>750</sub> herangezogen werden (Tandeau de Marsac, 1977). Nachteil beider Methoden ist, dass die Analyse einzelner Zellen nicht möglich ist, sondern nur ein Durchschnitt über die gesamte Probe gebildet werden kann. Somit kann keine Aussage über den Anteil an viablen und nicht viablen Zellen in der Kultur getroffen werden.

Die Standardmethode zur Bestimmung des Anteils an viablen Zellen in Monokulturen ist das Ausplattieren und eine anschließende Koloniezahlbestimmung. Allerdings ermöglicht sie keine zeitnahe Analyse, da Kolonien meist erst nach einer Woche Wachstum gezählt werden können.

Zudem kann die Viabilität von Zellen über Fluoreszenzfärbungen ermittelt werden. Dabei kann zum Beispiel die Autofluoreszenz von Chlorophyll für die Identifikation von viablen Zellen und der Farbstoff SYTOX Green für die Identifikation von nicht viablen Zellen genutzt werden (Sato et al., 2004). Alternativ wurde auch der Einsatz des Fluoreszenzfarbstoffes Neutral Red beschrieben (Zetsche und Meysman, 2012). Auch wenn die Fluoreszenzfärbung schneller Ergebnisse liefert, erfordert sie Zeit und ist mit zusätzlichen Kosten verbunden. Für eine Automatisierung wurde der Einsatz von Flowcytometern beschrieben, bei dem der Anteil an viablen Zellen einer *Microcystis* Kultur nach vorhergehender Fluoreszenzfärbung bestimmt wurde (Lee et al., 2000). Dieser Ansatz hat den Vorteil schneller und reproduzierbarer zu sein. Allerdings ist auch hier eine Färbung erforderlich.

## 1.5 Bildverarbeitung und Bilderkennung in der Biologie

Um den Arbeitsaufwand bei der Bildauswertung zu reduzieren und die Analysen vergleichbar zu machen, werden in der Biologie zunehmend Methoden der Informatik eingesetzt. Nach Peng (2008) wurde die Verarbeitung und Auswertung von Bildern biologischer Proben erstmals als *bioimage informatics* neu definiert. Diese

Bereich kann grob in die beiden Themengebiete Bildverarbeitung und Bilderkennung aufgliedert werden. Die Bildverarbeitung hat dabei das Ziel, die im Bild vorhandenen Informationen durch verschiedenste Manipulationen bestmöglich abzubilden und/oder einzelne Bestandteile aus dem Gesamtbild zu extrahieren. Dabei ist die Interpretation und Analyse der Daten weiterhin vom Menschen abhängig. Dagegen wird in der Bilderkennung versucht, das Eingangsbild in entsprechende vordefinierte Klassen einzuordnen (Uchida, 2013). Der Übergang zwischen beiden Bereichen gestaltet sich allerdings sehr fließend. Um optimale Ergebnisse zu erzielen, werden zum Lösen konkreter Fragestellungen daher meist Methoden aus beiden Bereichen miteinander kombiniert.

Für die Analyse von biologischen Bildern sollten dabei einige besondere Faktoren berücksichtigt werden. Zum einen sind die Objekte selbst oft erschwert abzubilden. Häufige Charakteristika, die dies verursachen sind eine sehr geringe Größe, Transparenz, ungenaue Begrenzungen und Texturen, Überlappungen, Veränderungen innerhalb des Objektes (hier vor allem durch Sterben von Zellen oder Photobleaching-Effekte) oder auch Deformationen während des Sterbens. Zum anderen können durch die in der Mikroskopie vorhandenen Aufnahmebedingungen Probleme entstehen. Beispielsweise enthalten Fluoreszenzaufnahmen aufgrund der geringen Signalintensität oft ein hohes Hintergrundrauschen, welches die weitere Auswertung erschwert (Uchida, 2013). All dies lässt die Verarbeitung und Erkennung von biologischen Bildern deutlich komplexer werden, als sie bei bekannten Fragestellungen, wie der Erkennung von Schrift oder besser definierten Gegenständen wie Autos ist.

Die Anwendungsgebiete der Bildverarbeitung und Bilderkennung sind in der Biologie weit gefächert, da viele Prozesse die auf Bildern festgehalten werden können automatisch analysiert werden können. Einsatzgebiete sind z.B. die Quantifizierung und Identifizierung unterschiedlicher Zell-Phänotypen und subzellulärer Merkmale (Carpenter et al., 2006), die dreidimensionale Kartierung von Modellorganismen oder einzelner Zellen über Expressionsbilder (Long et al., 2009; Carson et al., 2005), die Analyse von biologischen Prozessen innerhalb von Zellen über z.B. das Verfolgen von Mikrotubuli (Altınok et al., 2007) oder aber auch die Rekonstruktion und Darstellung von dreidimensionalen Strukturen (Schmid et al., 2010; Longair et al., 2011).

Dabei ermöglicht die Kombination mit automatisierten Mikroskopieroutinen, dass ei-

ne Vielzahl von bildbasierten Anwendungen im Hochdurchsatz durchgeführt werden können (Wollman und Stuurman, 2007). Zudem kommen Bildverarbeitungstechniken auch in Hochdurchsatzprozessen wie der Microarrayauswertung (White et al., 2005) oder dem Quantifizieren von Fluoreszenzsignalen von Next Gen Sequenzierungs-Methoden (Shendure und Ji, 2008; Metzker, 2010; Loman et al., 2012) zum Einsatz.

## 2 Zielstellung

Zielstellung der Arbeit war es Methoden der Mikroskopie, Bildverarbeitung und Bilderkennung für die Charakterisierungen verschiedener Phytoplankter einzusetzen, um deren Analyse zu verbessern und zu automatisieren.

Im ersten Teil der Arbeit war es Ziel, ein System zur automatisierten Erkennung verschiedener Phytoplankter zu etablieren, welches in der Wasserqualitätsanalyse eingesetzt werden kann. Hierzu wurden für eine Automatisierung geeignete mikroskopische Verfahren für Hellfeld und Fluoreszenz etabliert. Es wurden für die Fragestellung geeignete Methoden zur Bildauswertung als Plugin für die Bildverarbeitungssoftware ImageJ erstellt. Um eine Unterscheidung der Plankter zu ermöglichen, wurden zudem geeignete Klassifikationsverfahren ausgewählt und in ImageJ integriert. Das aufgebaute System wurde mit definierten Mischproben überprüft. Hierbei wurde auf Indikatortaxa, die besonders gut für die Analyse der Wasserqualität geeignet sind, zurückgegriffen.

Im zweiten Teil der Arbeit wurde eine neue Methode für die vereinfachte Viabilitätsanalyse von Phytoplanktern entwickelt. Diese kann zum einen bei der Analyse von Phytoplanktongemeinschaften neue Erkenntnisse ermöglichen. Zum anderen kann hierdurch die Überwachung von Wachstumsbedingungen in Algenmonokulturen verbessert werden. Hierbei sollte der Effekt, dass neben der roten Chlorophyllfluoreszenz bei gleicher Anregungswellenlänge eine grüne unspezifische Fluoreszenz in den Organismen beobachtet werden konnte, genutzt werden. Im Rahmen der Arbeit wurde anhand des Modellorganismus *Synechoscystis* sp. 6803 untersucht, inwiefern der Effekt zur Unterscheidung von viablen und nicht viablen Organismen genutzt werden kann und welche Einsatzmöglichkeiten er bietet. Die Bildaufnahme und die Bildauswertung wurden über entsprechende Methoden automatisiert.

# 3 Material und Methoden

## 3.1 Materialien

Von den für die Laborarbeiten verwendeten Substanzen werden hier nur Nicht-Standard-Substanzen aufgeführt und genauer spezifiziert. Alle Substanzen hatten üblicherweise p. A. Qualität und wurden, wenn nicht anders spezifiziert, von Roth, Sigma-Aldrich oder Merck bezogen. Es werden nur Laborgerätschaften aufgeführt, die nicht zur Standardeinrichtung eines Labors gehören oder bei denen die genauere Spezifikation aus anderen Gründen nötig ist.

### 3.1.1 Verbrauchsmaterialien

#### Medien und Lösungen (alphabetisch)

3NBBM+V Medium	8,82 mM NaNO <sub>3</sub> , 0,17 mM CaCl <sub>2</sub> * 2 H <sub>2</sub> O, 3,04 mM MgSO <sub>4</sub> * 7 H <sub>2</sub> O, 0,33 mM K <sub>2</sub> HPO <sub>4</sub> * 3 H <sub>2</sub> O, 1,29 mM KH <sub>2</sub> PO <sub>4</sub> , 0,43 mM NaCl, 11 nM Vitamin B <sub>12</sub> , 0.003 mM Vitamin B <sub>1</sub> , 1 nm Biotin, 0.02 mM EDTA und Spurenelemente
BG11 Medium	17,65 mM NaNO <sub>3</sub> , 0,18 mM K <sub>2</sub> HPO <sub>4</sub> * 3 H <sub>2</sub> O, 0,03 mM Citronensäure, 0,003 mM EDTA (Dinatrium Magnesium), 0,19 mM Na <sub>2</sub> CO <sub>3</sub> , 0,03 mM Ammoniumeisencitrat und Spurenelemente
Diat Medium	0,41 mM Na <sub>2</sub> SiO <sub>3</sub> , 4 nM Vitamin B <sub>12</sub> , 3 %Vol. Erdextrakt, 0.003 mM Ammoniumeisencitrat, 0,03 mM Citronensäure, 0,18 mM K <sub>2</sub> HPO <sub>4</sub> * 3 H <sub>2</sub> O, 0,02 mM FeSO <sub>4</sub> , 0,01 mM EDTA und Spurenelemente

ErbsMS	1 Erbse und 200 ml eines Sand-Gartenerde Gemischs (Verhältnis 1:1) wurde in einem 1 l Erlmayerkolben mit 1 l dest. Wasser aufgefüllt und an zwei aufeinander folgenden Tagen für 1 Stunde gekocht.
Erdextrakt	100 g Gartenerde wurden mit 1 l dest. H <sub>2</sub> O aufgefüllt und an zwei aufeinander folgenden Tagen für jeweils 1 Stunde gekocht.
ES Medium	2 mM KNO <sub>3</sub> , 0,11 mM K <sub>2</sub> HPO <sub>4</sub> , 0,08 mM MgSO <sub>4</sub> * 7 H <sub>2</sub> O, 3 %Vol. Erdextrakt, 0,02 mM FeSO <sub>4</sub> , 0,01 mM EDTA, Spurenelemente
ES-B <sub>12</sub> Medium	ES Medium, 4 nM Vitamin B <sub>12</sub>
Euglena Medium	12,19 mM Natriumacetate, 1 g/l Rindextrakt, 2 g/l Bacto-Tryptone, 2 g/l Hefeextrakt, 3 %Vol. Erdextrakt
Formaldehyd Lösung	90 %Vol. PBS Puffer, 10 %Vol. Paraformaldehyd
+V Medium	90 %Vol. ES Medium, 10 %Vol. Euglena Medium, 0,002 mM Vitamin B <sub>1</sub> und 4 nM Vitamin B <sub>12</sub>

### 3.1.2 Biologisches Ausgangsmaterial

#### Viabilitätsanalyse

Für die Entwicklung und Etablierung des Viabilitätsassays wurde der Stamm *Synechocystis* sp. PCC6803 von Cyano Biofuels GmbH zur Verfügung gestellt. Dieser wurde in T75 Kulturflaschen in BG11 Medium kultiviert. Der Stamm wurde im Minitron-Schüttler mit Photosyntheselicht bei 35°C, 55 rpm und einer kontinuierlichen Belichtung von 20 µE/m<sup>2</sup> gehalten. Die Kultur wurde durch Verdünnen in der logarithmischen Wachstumsphase gehalten.

#### Automatisierte Phytoplanktonerkennung

Die für diese Arbeiten genutzten Stämme sind im Folgenden inklusive ihrer Herkunft aufgelistet. Sie wurden in T75 Kulturflaschen in den jeweils vermerkten Kulturmedien kultiviert. Die Kulturen wurden bei Raumtemperatur und Tageslicht gehalten. Sie wurden in wöchentlichen Abständen mikroskopiert und verdünnt.

*Cyclotella meneghiniana* (Göttingen, SAG 2136, ES-B<sub>12</sub> Medium)

<i>Anabaena</i> sp.	(Cyano Biotech GmbH, CBT 149, BG11 Medium)
<i>Chlorogonium elongatum</i>	(Göttingen, SAG 31.98, 3NBBM+V Medium)
<i>Cryptomonas ovata</i>	(Göttingen, SAG 979-3, ES-B <sub>12</sub> Medium)
<i>Desmodesmus perforatus</i>	(U. Mischke, Müggelsee, Berlin)
<i>Staurastrum tetracerum</i>	(Göttingen, SAG 7.94, ErbsMS Medium)
<i>Botryococcus braunii</i>	(Göttingen, SAG 807-1, BG11 Medium)
<i>Pediastrum duplex</i>	(Göttingen, SAG 28.83, ErbsMS Medium)
<i>Trachelomonas volvocina</i>	(Göttingen, SAG 1283-4, +V Medium)
<i>Crucigenia tetrapedia</i>	(Göttingen, SAG 9.81, BG11 Medium)
<i>Fragilaria crotonensis</i>	(Göttingen, SAG 28.96, Diat Medium)
<i>Nitzschia palea</i>	(Göttingen, SAG 1052-3a, Diat Medium)
<i>Monoraphidium minutum</i>	(Göttingen, SAG 243-1, ES Medium)
<i>Ankistrodesmus falcatus</i>	(Göttingen, SAG 202-2, BG11 Medium)
<i>Monoraphidium contortum</i>	(Göttingen, SAG 47.80, ES Medium)
<i>Pandorina morum</i>	(Göttingen, SAG 60-1a, ES Medium)
<i>Tetraedron caudatum</i>	(Göttingen, SAG 2091, ErbsMS Medium)
<i>Actinastrum hantzschii</i>	(Göttingen, SAG 52.81, ES Medium)
<i>Aphanizomenon gracile</i>	(Göttingen, SAG 31.79, ES Medium)
<i>Limnothrix redekei</i>	(Göttingen, SAG 3.89, BG11 Medium)
<i>Merismopedia glauca</i>	(Göttingen, SAG 48.79, ES Medium)
<i>Coelastrum microporum</i>	(Göttingen, SAG 2292, ErbsMS Medium)

### 3.1.3 Geräte

Automatisches inverses Fluoreszenzmikroskop Bio-revo BZ 9000 mit Steuerungssoftware BZ II Viewer	(Keyence, Japan)
Fluoreszenzfilterset Chlorophyll a, b & c	
Anregung: BrightLine HC 435/40 nm	(Semrock, USA)
Strahlenteiler: HC BS 510 nm	(Semrock, USA)
Emission: E 515 nm Langpass	(Chroma, USA)

Fluoreszenzfilterset Phycoerythrin	
Anregung: BrightLine HC 543/22 nm	(Semrock, USA)
Strahlenteiler: HC BS 562 nm	(Semrock, USA)
Emission: BrightLine 593/40 nm	(Semrock, USA)
Fluoreszenzfilterset Phycocyanin	
Anregung: BrightLine HC 600/37 nm	(Semrock, USA)
Strahlenteiler: Q 625 nm LP	(Chroma, USA)
Emission: BrightLine HC 655/40 nm	(Semrock, USA)
Neubauer Zählkammer (Tiefe 0,02 mm)	(Marienfeld, Deutschland)
Minitron Schüttler mit Photosyntheselampe	(Infors HT, Schweiz)
Spektrophotometer UV 1800	(Shimadzu, Japan)
Utermöhl-Kammer	(Hydro-Bios, Kiel)

### 3.1.4 Software

AutoHotkey - <i>Programm zur Automatisierung von Maus- und Tastaturbefehlen</i>	(Chris Mallet, 2004-2013)
Eclipse 3.7.2 - <i>Integrierte Entwicklungsumgebung zur Entwicklung von Software</i>	(Eclipse Foundation, Inc., Kanada)
Encog 3.1 - <i>Umgebung für verschiedene Algorithmen des maschinellen Lernens</i>	(Heaton, 2010)
ImageJ 1.43u - <i>Bildverarbeitungsprogramm</i>	(Rasband, 1997)
ImageJ Directionality Histogram Plugin	(Tinevez, 2010; Rezakhaniha et al., 2012)
ImageJ Elliptic Fourier Descriptor (EFD) Plugin	(Boudier und Tupper, 2009)
ImageJ Image Moments Plugin	(Prodanov, 2008)
ImageJ Texture Analyzer Plugin	(Cabrera, 2006)
Lire Java Bibliothek 0.9.3 - <i>Umgebung für Algorithmen der Merkmalsberechnung</i>	(Lux und Chatzichristofis, 2008)
R base package 2.14.1-1 - <i>Umgebung für statistische Berechnungen und Erstellung von Grafiken</i>	(R Foundation)
RStudio 0.97.551 - <i>Integrierte Entwicklungsumgebung für R</i>	(RStudio, Inc, 2009-2012)



Waikato Environment for Knowledge Analysis (Hall et al., 2009)  
(Weka) 3.7.9 - *Sammlung von Algorithmen des maschinellen Lernens*

## 3.2 Methoden

Im Folgenden sind die in der Zielstellung beschriebenen Methoden zur automatisierten Analyse unterschiedlicher Phytoplankter beschrieben.

### 3.2.1 Mikrobiologische Methoden

#### Kulturvorbereitung zur Validierung der Viabilitätsmessung

Eine viable *Synechocystis*-Kultur wurde in zwei Proben aufgeteilt (jeweils 20 ml und  $OD_{750} = 2$ ). Um nicht viable Zellen zu erhalten, wurde eine der beiden Proben in einer Schottflasche für 5 Tage bei 60 °C auf einem Magnetprüher inkubiert. Die zweite Probe wurde währenddessen im Minitron-Schüttler mit Photosyntheselicht bei 35°C, 55 rpm und einer kontinuierlichen Belichtung von 20  $\mu\text{E}/\text{m}^2$  weiter kultiviert. Von beiden Kulturen wurde anschließend die Zellzahl mit Hilfe einer Helber Zählkammer bestimmt und die Konzentration auf  $5 \cdot 10^7$  Zellen/ml eingestellt. Die Kulturen wurden in den Verhältnissen von 1:0, 4:1, 2:1, 1:4 und 0:1 gemischt.

#### $OD_{750}$ Messung und Absorptionsspektroskopie

Die optische Dichte der *Synechocystis*-Kultur wurde am Spektrophotometer bei 750 nm bestimmt. Zur Aufnahme der Absorptionsspektren wurden die Proben auf eine optische Dichte von 0,65 eingestellt. Die Spektren wurden im Bereich von 300 - 800 nm aufgenommen.

#### Konzentrationsbestimmung von Chlorophyll

Nach Tandeau de Marsac (1977) wurden 2 ml der *Synechocystis*-Kultur bei 15000 g für 10 min zentrifugiert. 1900  $\mu\text{l}$  des Überstandes wurden verworfen und die Zellen mit Hilfe eines Ultraschallbades resuspendiert. Anschließend wurden 900  $\mu\text{l}$  Methanol (100 %) zur Probe gegeben und diese für 1 Stunde bei +4 °C und Dunkelheit inkubiert. Nachdem die Probe für 10 min bei 15000 g zentrifugiert wurde, wurde

die Absorption bei 650 nm bestimmt. Die Chlorophyllkonzentration wurde mit Hilfe folgender Formel bestimmt:

$$\text{Chlorophyll}(\mu\text{g}/\text{ml}) = (\text{Abs}_{650} * 13,9)/2\text{ml} \quad (3.1)$$

Die Bestimmung der Chlorophyllkonzentration erfolgte für alle Proben in Duplikaten.

### **Ausplattieren der *Synechocystis*-Kultur**

Um die Anzahl an viablen Zellen pro ml zu bestimmen, wurden  $5 * 10^4$  Zellen einer Probe auf BG11 - Agarplatten (BG11, 1% Agar) ausplattiert (innerer Durchmesser: 87 mm ; Wachstumsfläche:  $60,1 \text{ cm}^2$ ). Anschließend wurden die Platten für eine Woche bei Raumtemperatur und  $10 \mu\text{E}/\text{m}^2\text{s}$  inkubiert. Jede Probe wurde in Triplikaten plattiert.

### **Sedimentation von Phytoplanktonproben**

Die Planktonproben wurden mit Paraformaldehyd (1%) fixiert und für die anschließende Mikroskopie nach Utermöhl (1958) in einer Utermöhl-Kammer sedimentiert. Das Probenvolumen wurde jeweils so angepasst, dass die sedimentierten Organismen sich möglichst nicht auf dem Boden der Kammer überlagerten. Zum Erstellen der Trainingsdaten wurden vorwiegend Monokulturen und zum Erstellen der Testdaten wurden Mischkulturen sedimentiert.

## **3.2.2 Automatisierte Mikroskopie**

### **Viabilitätsanalyse**

Die Probe wurde in einer Helber Zählkammer mit Hilfe des BZ 9000 Mikroskopes automatisiert mikroskopiert. Zur Minimierung von Zellbewegungen wurde die Kultur zuvor mit Glycerin (99.5%, p.a.) im Verhältnis 1:2 gemischt. Um eine fehlerfreie Bildanalyse zu ermöglichen, fand die Aufnahme der Bilder außerhalb des Zählgitters statt. Zur Automatisierung der Aufnahme wurde die Softwarefunktion AutoRangePhoto in der Keyence Software genutzt. Hierzu müssen vier vordefinierte Ecken eines Vierecks von Hand fokussiert werden. Anschließend werden die Schärferebenen innerhalb des Vierecks extrapoliert und die Bilder können automatisch aufgenommen

werden. Es wurde ein Volumen von 0,084  $\mu\text{l}$  mikroskopiert. Für jede Position wurde jeweils ein Fluoreszenzbild mit dem Filter zur Anregung von Chlorophyll und ein Bild im Phasenkontrast aufgenommen.

### **Koloniezahlbestimmung**

Für die Koloniezahlbestimmung der ausplattierten *Synechocystis*-Kulturen wurden, analog zur Methode unter dem Abschnitt **Viabilitätsanalyse**, Fluoreszenzsignale für Phycocyanin automatisch aufgenommen. Die Bildaufnahme wurde mit geschlossener Agarplatte durchgeführt und es wurde durch den Agar hindurch mikroskopiert. Für jede Platte wurde eine Fläche von 10.5  $\text{cm}^2$  automatisch erfasst.

### **Planktonerkennung**

Die automatisierte Mikroskopie wurde am BZ 9000 Mikroskop durchgeführt. Da einzelne benötigte Funktionen nicht in der Steuerungssoftware miteinander kombiniert werden konnten, wurde auf die Software AutoHotKey zurückgegriffen, mit deren Hilfe alle vorhandenen Funktionen automatisch angesteuert werden konnten. Es wurde eine Routine definiert, in der Bilder für 80 unterschiedliche Positionen innerhalb eines vordefinierten rechteckigen Bereiches aufgenommen werden können. Vor Beginn der Routine wurden Belichtung, Aperturstop und Belichtungszeit der Kamera auf vordefinierte Werte eingestellt (Hellfeld: Belichtungszeit 1/28s, Apertur 65% offen, Lichtintensität 84% ; Fluoreszenz: Belichtungszeit 1/4,5s und Verstärkung 18dB). Weiterhin wurde ein Weißabgleich durchgeführt und ein Bild des Hintergrundes für eine anschließende Ausleuchtungskorrektur gespeichert.

Da der Boden der verwendeten Utermöhl-Kammer nicht vollständig eben ist, wurde ein Autofokus zum Fokussieren der einzelnen Positionen eingesetzt. Es wurden fünf unterschiedliche Bilder mit einem 60x Objektiv aufgenommen: Ein Hellfeldbild, ein Quick Full Focus Bild (QFF) und Fluoreszenzbilder der Autofluoreszenz von Chlorophyll (ab 515 nm), Phycoerythrin (573 - 613 nm) und Phycocyanin (625 - 665 nm). Im QFF Bild wurden 58 verschiedenen Tiefenebenen (35 über und 23 unter der mit dem Autofokus ermittelten Position) in ein Bild integriert.

### 3.2.3 Methoden der Bildverarbeitung und Bilderkennung

#### ImageJ

Alle dargestellten Methoden der Bildverarbeitung und Bilderkennung wurden mit ImageJ durchgeführt. Der Grundaufbau der Software ist sehr einfach gehalten und alle vorhandenen Funktionen sind über einen Toolbar zugänglich (Abb. 3.1). Die Funktionalität der Software wurde für die Arbeiten über selbst erstellte und angepasste Plugins erweitert. Über die zur Verfügung stehende API können hier bereits bestehende Methoden gut mit neuen Methoden kombiniert werden. Hierdurch ist auch ein automatisierter Ablauf der Analysen möglich.

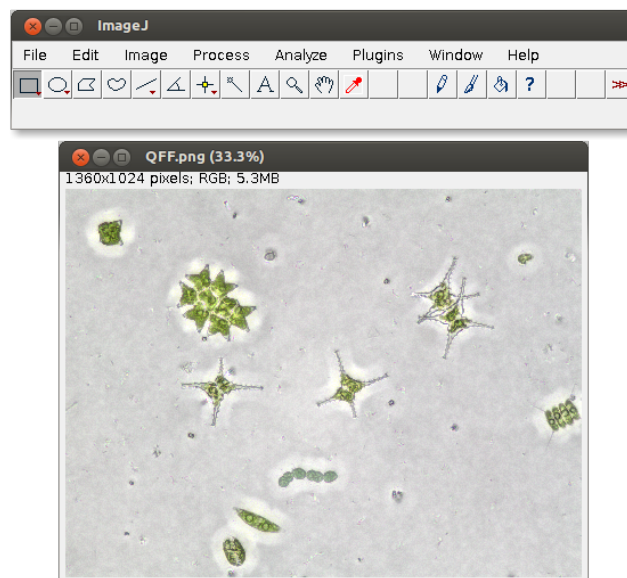


Abbildung 3.1: **Aufbau der ImageJ GUI.** Alle Funktionen von ImageJ können über die im oberen Teil der Abbildung dargestellten GUI aufgerufen werden. Zudem werden Erweiterungen der Software unter dem Bereich Plugins integriert. Die Darstellung und Verarbeitung von Bilder erfolgt in eigenständigen Fenstern.

#### Viabilitäts- und Zellzahlbestimmung

Für die automatisierte Viabilitäts- und Zellzahlbestimmung wurden Fluoreszenzbilder verwendet, die nach der beschriebenen Methode unter Abschnitt 3.2.2 - Viabilitätsanalyse aufgenommen wurden. Bilder in denen keine Zellen erfasst wurden, wurden

zur Vermeidung von Artefakten ausgeschlossen. Hierzu musste die globale maximale Helligkeitsintensität des Bildes kleiner als 15 sein (Intensitätsbereich 0 - 255). Für die Erkennung der Zellen wurden die Bilder zunächst in RGB-Farbkanäle aufgetrennt. Anschließend wurde für den roten und grünen Kanal separat ein Schwellwertbild erstellt. Der Schwellwert wurde automatisch über die Maximum Entropy Methode (Kapur et al., 1985) ermittelt. Beide Schwellwertbilder wurden anschließend kombiniert und die gefundenen Partikel mit der Particle Analyzer Funktion registriert (Igathinathane et al., 2008). Partikel, die kleiner als  $35 \mu\text{m}^2$  waren wurden von der weiteren Analyse ausgeschlossen. Für eine Unterscheidung von roten und grünen Zellen wurde der Mittelwert der Partikelintensität (Intensitätsbereich 0 - 255) im grünen Kanal bestimmt. Dabei wurden Partikel mit Werten kleiner als 50 als rote Zellen und Partikel mit Werten größer oder gleich 50 als grüne Zellen registriert. Nach Analyse aller Bilder wurde der Prozentsatz an roten und grünen Zellen und die Konzentration pro ml berechnet. Der Code des Plugins ist unter <https://github.com/KatjaSchulze/Viability> zugänglich.

### **Koloniezahlbestimmung**

Für diese Methode wurde auf Fluoreszenzbilder, die mit der Methode unter Abschnitt 3.2.2 - Koloniezahlbestimmung aufgenommen wurden, zurückgegriffen. Hierbei wurde über den Helligkeitskanal des Bildes ein Schwellwertbild mit Hilfe der Minimum Methode (Prewitt und Mendelsohn, 1966) bestimmt. Bilder, die keine Kolonien enthielten, wurden von der Analyse ausgeschlossen, indem ein maximaler Helligkeitswert im Gesamtbild größer 25 vorausgesetzt wurde. Anschließend wurden die Partikel mit der Particle Analyzer Funktion registriert und gezählt. Zur Berechnung der Viabilität der Kulturen, wurde die Zellzahl aller Proben mit der Zellzahl der Kontrolle, bei der nur vitale Zellen plattiert wurden, normalisiert. Der Code des Plugins ist unter <https://github.com/KatjaSchulze/CyanoColonyCounter> zugänglich.

### **Planktonerkennung**

Im Rahmen der Planktonerkennung wurde auf Bilder zurückgegriffen, die mit Hilfe der unter Abschnitt 3.2.2 - Planktonerkennung beschriebenen Methode erstellt wurden. Im Folgenden werden die einzelnen Methoden, die für diese Arbeiten genutzt

wurden, dargestellt. Alle verwendeten Methoden sind unter <https://github.com/KatjaSchulze/PlanktoVision> zugänglich.

### **Bildkorrektur**

Da die Helligkeitsverteilung in den erstellten Hellfeldbildern aufgrund einer ungleichmäßigen Ausleuchtung im Mikroskop unregelmäßig war, wurde vor jedem Durchlauf ein Bild ohne Probe zur Bildkorrektur aufgenommen. Mit Hilfe der CalculatorPlus Funktion von ImageJ wurde das Bild nach folgender Formel korrigiert:

$$I_{kor} = (I_p/I_{op}) * mean_{op} \quad (3.2)$$

Dabei ist  $I$  die Helligkeit eines Pixels des jeweiligen Bildes ( $kor$  - korrigiertes Bild;  $p$  - Bild der Probe,  $op$  - Bild ohne Probe) und  $mean_{op}$  der globale Mittelwert der Helligkeit des Bildes ohne Probe.

### **Segmentierung**

Um die Organismen vom Hintergrund der Mikroskopiebilder zu segmentieren, wurde ein Region Growing Ansatz gewählt (Adams und Bischof, 1994). Grundprinzip der Methode ist es, Pixel des Bildhintergrundes als Startpixel einer Region zu definieren und angrenzende Pixel hinzuzufügen, falls diese ein bestimmtes Kriterium erfüllen. Das Wachsen der Region ermöglicht somit eine Unterscheidung in Hintergrund und Organismen. Damit auch transparente Organismen gut detektiert werden können, wurde eine Kantenerkennung mit dem Sobel Operator und eine anschließende Kontrastanpassung (ImageJ - Enhance Contrast Funktion: Histogramm Normalisierung bei der 0.4% der voll gesättigten Pixel genutzt werden) als Vorverarbeitung durchgeführt. Um korrekte Startpixel im Hintergrund zu finden, wurde die Watershed Segmentierung der MaximumFinder Funktion von ImageJ genutzt, womit eine grobe Segmentierung des Bildes in verschiedene Bereiche möglich ist. Anschließend wurde für jeden Bereich der Modalwert bestimmt. Da bei dieser Art Bilder der Hintergrund die größte Fläche ausmacht, wurde die größte Fläche mit gleichem Modalwert als Hintergrund definiert. Zum Start des Region Growing wurden alle Pixel der Hintergrundfläche, die dem Modalwert dieser Fläche entsprachen, als Startpixel definiert. Anliegende Pixel wurden anschließend hinzugefügt, sofern sie das Hintergrundkrite-

rium erfüllen. Dabei ergaben sich die besten Ergebnisse, wenn die Standardabweichung der Helligkeit aller bereits zur Fläche hinzugefügter Pixel multipliziert mit 10 als Kriterium verwendet wurde. Die segmentierten Partikel wurden anschließend mit der Particle Analyzer Funktion registriert. Dabei wurden Partikel die kleiner als 100 Pixel (ca.  $3 \mu\text{m}^2$ ) waren und Partikel, die an den Kanten des Bildes lagen nicht weiter betrachtet.

### **Merkmalsberechnung**

Zur Unterscheidung der Organismen müssen verschiedene Merkmale berechnet werden, um diese weniger komplex und besser definiert beschreiben zu können.

Da einige Merkmale nicht unabhängig von der Rotation des Partikels sind, wurde diese in einem vorverarbeitenden Schritt korrigiert. Hierzu wurde der Partikel über seine Primärachse auf die x-Achse des Bildes gedreht. Die Primärachse wurde dabei über eine an den Partikel gefittete Ellipse bestimmt.

Nach Berechnung der im Folgenden erläuterten Merkmale, wurden diese für die jeweiligen Partikel in einer xls-Datei gespeichert. Der dazugehörige Bildausschnitt wurde als eigenständiges Bild gesichert, um eine spätere Einteilung in unterschiedliche Klassen zu ermöglichen. Die Merkmale wurden, wenn nicht anders angegeben, mit der in ImageJ integrierten Measure Routine und/oder den Funktionen `getStatistics()` und `getHistogramm()` ermittelt.

### ***Fläche***

Die Fläche eines Partikels in Quadratpixeln.

### ***Perimeter***

Die Länge der äußeren Grenzlinie eines Partikels in Pixeln.

### ***Breite / Höhe***

Die Breite / Höhe des kleinsten Rechtecks, das einen Partikel umschließt in Pixeln.

### ***Feret-Durchmesser***

Der längste Abstand zwischen zwei beliebigen Punkten entlang der Außenlinie eines Partikels in Pixeln.

### ***Minimaler Feret-Durchmesser***

Der kürzeste Abstand zwischen zwei beliebigen Punkten entlang der Außenlinie eines Partikels in Pixeln.

### ***Feret Winkel***

Der Winkel zwischen dem Feret-Durchmesser und einer Linie, die parallel zur X-Achse des Hauptbildes verläuft.

### ***Integrierte Dichte***

Das Produkt von Fläche und mittlerer Helligkeit des Partikels.

### ***Länge der Haupt- und Nebenachse***

Anhand einer bestmöglich an den Umriss des Partikels gefitteten Ellipse wird die Länge der längsten Achse (Hauptachse) und der 90° dazu gedrehten Achse (Nebenachse) in Pixeln ermittelt.

### ***Zirkularität***

Gibt an, wie sehr die Form des Partikels einem Kreis entspricht und wird wie folgt berechnet:

$$\text{Zirkularität} = \frac{4\pi * \text{Fläche}}{\text{Perimeter}} \quad (3.3)$$

Dabei entspricht 1 einem perfekten Kreis. Je näher der Wert entgegen 0 geht, desto länglicher ist die Form.

### ***Aspektverhältnis***

Ist das Verhältnis zwischen der Länge der Haupt- und der Nebenachse eines Partikels.

### ***Rundheit***

Ist sehr ähnlich zur Zirkularität und ist ein Maß dafür, wie rund der Partikel ist. Die Rundheit wird wie folgt berechnet:

$$\text{Rundheit} = \frac{4 * \text{Fläche}}{\pi * \text{Hauptachse}^2} \quad (3.4)$$

Sie entspricht 1 für einen Kreis und geht gegen 0 für sehr lang gestreckte Objekte.



### **Solidität**

Gibt an, wie stark die Kontur eines Partikels fragmentiert ist. Sie wird über das Verhältnis zwischen Partikelfläche und Fläche der Konvexhülle gebildet. Dabei ist die Konvexhülle die kleinste konvexe Fläche, die alle Pixel des Partikels enthält.

### **Schiefe**

Gibt an, wie asymmetrisch die Pixelintensitäten um den Mittelwert herum streuen. Dabei entspricht 0 einer vollkommen symmetrischen Verteilung. Werte  $< 0$  zeigen eine Verschiebung nach links, Werte  $> 0$  eine Verschiebung nach rechts an. Die Schiefe wird wie folgt berechnet:

$$\text{Schiefe} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^3 \quad (3.5)$$

mit  $n$  der Anzahl an Pixeln,  $x_i$  dem Pixelwert  $i$ ,  $\bar{x}$  dem Mittelwert und  $s$  der Standardabweichung.

### **Kurtosis**

Ist eine Maßzahl, die beschreibt wie steil die Pixelintensitäten um den Mittelwert herum verteilt sind beziehungsweise, wie stark die vorhandene Häufigkeitsverteilung um den Mittelwert herum gewölbt ist. Sie wird wie folgt berechnet:

$$\text{Kurtosis} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^4 \quad (3.6)$$

mit  $n$  der Anzahl an Pixeln,  $x_i$  dem Pixelwert  $i$ ,  $\bar{x}$  dem Mittelwert und  $s$  der Standardabweichung.

### **Hue**

Der Hue-Wert eines Pixels definiert den dargestellten Farbton. Innerhalb des Partikels wurden Mittelwert, Standardabweichung, Modalwert, minimaler Pixelwert, maximaler Pixelwert und das Histogramm für die vorhandenen Hue-Werte bestimmt.

### **Sättigung**

Die Sättigung eines Bildpixels definiert die Buntheit des dargestellten Farbtons. Innerhalb des Partikels wurden Mittelwert, Standardabweichung, Modalwert, minima-

ler Pixelwert, maximaler Pixelwert und das Histogramm für die vorhandenen Sättigungswerte bestimmt.

### ***Helligkeit***

Die Helligkeit eines Bildpixels definiert wie hell bzw. dunkel dieser erscheint. Innerhalb des Partikels wurden Mittelwert, Standardabweichung, Modalwert, minimaler Pixelwert, maximaler Pixelwert und das Histogramm für die vorhandenen Helligkeitswerte bestimmt.

### ***Statistik der Fluoreszenz***

Alle Merkmale wurden jeweils für die Fluoreszenzsignale von Phycoerythrin (PE), Phycocyanin (PC) und Chlorophyll (CHL) bestimmt. Für jedes Signal wurde im Kanal der Helligkeit des HSB - Farbraum innerhalb der jeweiligen Partikel Mittelwert, Modalwert, Standardabweichung sowie Minimum und Maximum der Pixelwerte ermittelt. Da in den Chlorophyllaufnahmen rote und grüne Signale auftreten, wurde hier das Bild zusätzlich in die Kanäle rot und grün aufgetrennt und jeweils Mittelwert, Modalwert, Standardabweichung sowie Minimum und Maximum innerhalb des Partikels bestimmt. Die gleichen Werte wurden zusätzlich für den Hue-Farbkanal des Bildes ermittelt.

### ***Fluoreszierender Anteil***

Alle Merkmale wurden jeweils für die Fluoreszenzsignale von Phycoerythrin (PE), Phycocyanin (PC) und Chlorophyll (CHL) bestimmt. Zur Ermittlung des fluoreszierenden Anteils innerhalb des Partikels wurde ein Schwellwertbild des Helligkeitskanals erstellt und anschließend der Verhältnis an Pixeln, die über dem Schwellwert lagen und der Gesamtfläche des Partikels bestimmt.

### ***Elliptische Fourier Deskriptoren der Konturlinie***

Mit Hilfe der Fourier Deskriptoren ist es möglich die Form der Konturlinie unabhängig von Rotation, Position und Größe des Partikels im Bild darzustellen. Mit einem in die Analyse integrierten und hierfür angepassten ImageJ Plugin (Boudier und Tupper, 2009) wurden zunächst die Fourier Koeffizienten der Konturlinie nach

Kuhl und Giardina (1982) ermittelt:

$$\begin{aligned}
xa_k &= \frac{2}{n} \sum_{i=1}^n x_i \cos \frac{2\pi ki}{n} \\
xb_k &= \frac{2}{n} \sum_{i=1}^n x_i \sin \frac{2\pi ki}{n} \\
ya_k &= \frac{2}{n} \sum_{i=1}^n y_i \cos \frac{2\pi ki}{n} \\
yb_k &= \frac{2}{n} \sum_{i=1}^n y_i \sin \frac{2\pi ki}{n}
\end{aligned} \tag{3.7}$$

Mit  $k$  der Nummer des zu ermittelnden Deskriptors,  $n$  der Anzahl an vorhandenen Konturpixeln,  $x_i$  der X-Koordinate des  $i$ -ten Pixels in der Konturlinie und  $y_i$  der Y-Koordinate des  $i$ -ten Pixels in der Konturlinie.

Die einzelnen Deskriptoren ergeben sich nun nach:

$$EFD_k = \sqrt{\frac{xa_k^2 + ya_k^2}{xa_1^2 + ya_1^2}} + \sqrt{\frac{xb_k^2 + yb_k^2}{xb_1^2 + yb_1^2}} \tag{3.8}$$

Für die Arbeiten wurden die ersten 30 Deskriptoren bestimmt.

### **Statistik der Grauwertmatrix**

Mit Hilfe einer Grauwertmatrix kann die Textur eines Bildes erfasst werden. Hierfür wird in der Matrix festgehalten, welche Pixelintensitäten sich in einem definiertem Abstand im Bild gegenüberliegen und wie häufig die jeweiligen Kombinationen auftreten.

Die Grauwertmatrix  $C$  eines Bildes  $I$  der Dimension  $n * m$  und dem Abstand  $\Delta x, \Delta y$  wird wie folgt ermittelt:

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{wenn } I(p, q) = i \text{ und } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{andernfalls} \end{cases} \tag{3.9}$$

Dabei sind  $i$  und  $j$  die Pixelintensitäten im Graustufenbild und  $p$  und  $q$  die aktuelle Position im Bild. Über den Abstand  $\Delta x, \Delta y$  kann dabei der Winkel in dem die Matrix

erfasst wird und die Distanz zwischen den Pixel, die miteinander in Verbindung gebracht werden verändert werden.

Anhand dieser Matrix können nach Haralick et al. (1973) und Walker et al. (1995) einzelne Werte berechnet werden, die die Textur beschreiben. Nachfolgende Werte wurden über die normalisierte Grauwertmatrix  $Cn$  (normalisiert über die Anzahl an Pixeln, die untersucht wurden) eines jeden Partikels für Rotationen von  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  und  $135^\circ$  und Distanzen von 1 - 5 Pixeln erfasst.

$$\text{Energie} = \sum_{i,j=0}^{N-1} Cn_{(i,j)}^2 \quad (3.10)$$

$$\text{Kontrast} = \sum_{i,j=0}^{N-1} Cn_{(i,j)}(i-j)^2 \quad (3.11)$$

$$\text{Homogenität} = \sum_{i,j=0}^{N-1} \frac{Cn_{(i,j)}}{1+(i-j)^2} \quad (3.12)$$

$$\text{Entropy} = \sum_{i,j=0}^{N-1} Cn_{(i,j)}(-\ln Cn_{(i,j)}) \quad (3.13)$$

$$\text{Korrelation} = \sum_{i,j=0}^{N-1} Cn_{(i,j)} \left( \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right)$$

mit  $\mu_i = \sum_{i,j=0}^{N-1} i * Cn_{(i,j)}$   $\mu_j = \sum_{i,j=0}^{N-1} j * Cn_{(i,j)}$  (3.14)

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} Cn_{(i,j)}(i-\mu_i)^2 \quad \sigma_j^2 = \sum_{i,j=0}^{N-1} Cn_{(i,j)}(j-\mu_j)^2$$

Zur Bestimmung der aufgelisteten Werte wurde das ImageJ-Plugin Texture Analyzer (Cabrera, 2006) mit Abänderungen in die Analyse integrierten.

### ***Richtungshistogramm***

In einem Richtungshistogramm wird die lokale Orientierung von Strukturen im Bild ermittelt und in Richtungsklassen zusammengefasst dargestellt. Hiermit können Haupt-

orientierungen innerhalb von Strukturen ermittelt werden. Handelt es sich z.B. um eine isotrope Struktur, d. h. der Anteil aller Richtungen ist gleich groß, ergibt sich ein flaches Histogramm. Bei Strukturen mit bevorzugten Richtungen ergeben sich Peaks im Histogramm. Zur Ermittlung des Histogramms wurde das Plugin Directionality (Tinevez, 2010; Rezakhaniha et al., 2012) abgeändert und in die Analyse integriert. Hierbei wird zunächst der Gradient  $x$  ( $G_x$ ) und  $y$  ( $G_y$ ) in Rotation von  $0^\circ$  und  $90^\circ$  im Bild  $I$  mit Hilfe eines 5x5 Sobel Filters ermittelt:

$$G_x = \begin{bmatrix} 2 & 3 & 4 & 3 & 2 \\ 1 & 2 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -2 & -3 & -2 & -1 \\ -2 & -3 & -4 & -3 & -2 \end{bmatrix} * I \quad G_y = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -3 & -2 & 0 & 2 & 3 \\ -4 & -3 & 0 & 3 & 4 \\ -3 & -2 & 0 & 2 & 3 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} * I \quad (3.15)$$

Die lokale Orientierung  $\Theta$  und die Stärke  $G$  der Kanten wird anschließend wie folgt bestimmt:

$$G = G_x^2 + G_y^2 \quad (3.16)$$

$$\Theta = \text{atan}(G_y, G_x) \quad (3.17)$$

Anschließend wird die Kantenstärke  $G$  mit dem bereits bestehenden Wert in der passenden Klasse des Histogramms aufaddiert. Die Zuordnung zu den Klassen erfolgt dabei über die lokale Orientierung  $\Theta$ . Es wurden 20 Histogrammklassen mit einer Klassenbreite von jeweils  $9^\circ$  verwendet.

### **Momente**

Mit Hilfe von Momenten kann die Verteilung der Helligkeitswerte innerhalb des segmentierten Partikels beschrieben werden. Hierbei handelt es sich um verschiedenen gewichtete Mittelwerte der Helligkeitswerte, die so gewählt sind, dass eine geometrische Interpretation möglich wird. Für die Berechnung der Momente wurde das ImageJ Plugin Image Moments angepasst und in die Analyse integriert (Prodanov, 2008). Nach M.-K. Hu (1962), Flusser (2000) und Flusser und Suk (2006) wurden hiermit 8 verschiedene Momente berechnet, die invariant bezüglich der Rotation, Skalierung und Translation sind.

### **Local Binary Patterns (LBP)**

LBP's sind Merkmale, die die Textur innerhalb eines Partikels abbilden und basieren auf relativen Unterschieden zwischen benachbarten Pixeln (Ojala et al., 2002). Die Methoden zur Berechnungen wurden neu in ImageJ implementiert.

Grundprinzip ist es, jeden Pixel einer definierten kreisförmigen Nachbarschaft mit dem zentralen Pixel zu vergleichen. Abhängig vom Unterschied wird der Schwellwert 1 oder 0 zugewiesen und alle Werte dieser Nachbarschaft werden aufaddiert und quadriert. Die kreisförmige Nachbarschaft wird hierbei durch  $(P, R)$  definiert.  $P$  steht für die Anzahl an genutzten Pixeln und  $R$  definiert den Radius der Nachbarschaft. Die verwendeten Pixel  $P$  um den zentralen Pixel  $(x, y)$  liegen auf den Koordinaten  $(x_p, y_p) = (x + R\cos(2\pi p/P), y - R\sin(2\pi p/P))$ . Wenn diese Koordinaten nicht existieren wird der Pixelwert interpoliert. Das LBP eines zentralen Pixel wird nun wie folgt ermittelt:

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(f(x, y) - f(x_p, y_p)) 2^p \quad (3.18)$$

mit  $s(z)$  als Schwellwertfunktion

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \quad (3.19)$$

Durch das Multiplizieren der einzelnen Schwellwerte mit dem binomialen Faktor  $2^p$  wird eine spezifische LBP-Bitsequenz erzeugt, die die lokale Textur beschreibt. Alle LBP-Sequenzen einer Region werden in einem Histogramm zusammengefasst, um die Textur der analysierten Region zu beschreiben.

Da LBP's abhängig von der Rotation des Partikels sind, wurde anschließend eine diskrete Fourier Transformation des Histogramms nach Ahonen et al. (2009) durchgeführt. Dabei ergibt sich ein rotationsunabhängiges Fourierspektrum, welches für die Klassifikation genutzt wurde. Die Spektren wurden für die LBP Histogramme  $LBP_{8,1}$ ,  $LBP_{8,2}$ ,  $LBP_{8,3}$ ,  $LBP_{16,2}$  und  $LBP_{16,3}$  erstellt.

### **Symmetrie**

Nach Fischer et al. (2000) wurde eine Methode zur Bestimmung der Punkt- und Achsensymmetrie der Partikelkontur in ImageJ neu implementiert.

Zunächst wird die zweidimensionale Kontur als eindimensionale Funktion  $d$  dargestellt. Hierzu wird zwischen dem Massenmittelpunkt des Partikels und definierten Punkten auf der Kontur der Abstand bestimmt. Die Konturpunkte werden im Uhrzeigersinn über einem konstanten Winkel  $\Delta\alpha$  als Abstand festgelegt. Dabei repräsentiert jeder Eintrag  $i$  in der Abstandsliste eine gerade Linie mit dem Winkel  $i \cdot \Delta\alpha$ , die durch den Mittelpunkt des Partikels läuft. Wenn an dieser Linie eine Achsensymmetrie vorliegt, müssen alle sich gegenüberliegenden Abstände gleich sein. Die Asymmetrie einer Form im Bezug zur geraden Linie  $i$  und der Anzahl an Abständen  $n$  in der Abstandsliste kann somit wie folgt bestimmt werden:

$$a = \sum_{j=0}^{(n/2)-1} |d_{(i+j) \bmod n} - d_{(i-j) \bmod n}| \quad (3.20)$$

Da bei einer Achsensymmetrie die Asymmetrie klein sein muss, wurde über einen festgelegten Schwellwert ermittelt, bei welchem Winkel diese vorliegt. Es wurden bis zu vier Symmetrieachsen ermittelt.

Die Punktsymmetrie zweiten Grades wurde entsprechend wie folgt bestimmt:

$$a = \sum_{i=0}^{(n-1)} |d_i - d_{(i+n/2) \bmod n}| \quad (3.21)$$

Hierbei werden nicht die sich jeweils auf der Konturlinie gegenüberliegenden Punkte, sondern Punkte die jeweils in einer Drehung von  $180^\circ$  zueinander liegen miteinander verglichen. Auch hier wurde über einen zuvor bestimmten Schwellwert anschließend festgestellt, ob im Partikel eine Punktsymmetrie vorliegt.

Nach einer ersten Überprüfung des Systems und einer Erweiterung der Daten wurde die Merkmalsberechnung erweitert. Im Folgenden werden die Merkmale aufgelistet, die nur für das vergrößerte Datenset verwendet wurden. Hierzu wurden Methoden der Lire Bibliothek (Lux und Chatzichristofis, 2008) in die ImageJ Analyse integriert und entsprechend angepasst.

### ***Einfache Merkmale***

Mit der BasicFeature()-Methode von Lire wurden Gesamthelligkeit, Kontrast, Anzahl an verschiedenen Hue Werten, Gesamtsättigung, Clipping, Komplexität, Schiefe und Energie des Partikels ermittelt.

### ***Kanten-Histogramm***

In einem Kanten Histogramm wird die räumliche Verteilung von Kanten erfasst und es stellt somit eine weitere Möglichkeit der Texturbeschreibung dar. Hierzu wurde das Bild in 16 gleich große Bereiche aufgeteilt und jeweils der Anteil an Kanten mit bestimmter Ausrichtung ermittelt. Dabei wurden 5 Kategorien unterschieden: vertikal, horizontal,  $45^\circ$ ,  $135^\circ$  und Kanten ohne Richtung (Sikora, 2001).

### ***DCT Koeffizienten Histogramm***

DCT Koeffizienten werden dazu genutzt, um Veränderungen und Verläufe von Farben im Bild zu erfassen. Hierfür wurde das Bild in  $8 \times 8$  Blöcke eingeteilt in die YUV Farbdarstellung umgewandelt. Für jeden Block und jeden Farbkanal erfolgt eine diskrete Kosinustransformation. Für jeden Block wird anschließend durch zick-zack scannen der Koeffizienten und eine nicht lineare Quantifizierung ein Deskriptor ermittelt. Hieraus ergab sich ein Histogramm mit 192 Werten.

### ***Tamura Merkmale***

Mit Hilfe von Tamura Merkmalen werden Eigenschaften eines Bildes beschrieben, die auf der menschlichen Wahrnehmung basierende wichtige Texturmerkmale erfassen. Nach Tamura et al. (1978) wurden die 6 verschiedenen Eigenschaften Granularität, Kontrast, Maß der Richtung, lineare Ähnlichkeit, Gleichmäßigkeit und Rauheit erfasst .

### ***Farb-Autokorrelogramm***

Mit Hilfe eines Farb-Autokorrelogramms kann die örtliche Korrelation von Farben innerhalb eines Bildes statistisch beschrieben werden. Dabei wird erfasst, wie häufig Pixel gleicher Farbe in zuvor definierten Abständen voneinander im Bild auftreten. Nach Huang et al. (1997) ist das Autokorrelogramm im Bild  $I$  mit den Pixeln  $p_1 = (x_1, y_1)$ ,  $p_2 = (x_2, y_2)$ , der Pixelfarbe  $I_c$ , dem definierten Abstand  $k$  und dem Abstand



$|p_1 - p_2|$  zwischen den beiden Pixeln wie folgt definiert:

$$\alpha_c^{(k)}(I) = \Pr_{p_1 \in I_c, p_2 \in I} (p_2 \in I_c | |p_1 - p_2| = k) \quad (3.22)$$

Das Farb-Autokorrelogramm wurde für 256 Farben  $I_c$  und für die Abstände  $k = 1, 2, 3, 4$  erstellt.

### ***Farbhistogramm***

Mit Hilfe eines Farbhistogramms kann die Verteilung von Farben im Bild erfasst werden. Es wurde jeweils ein Histogramm für den RGB - und den HSB - Farbraum aufgenommen. Zur Erstellung wird dabei die Häufigkeit der RGB - bzw. HSB - Kombination im jeweiligen Pixel ermittelt. Da die vorhandenen Intensitätsabstufungen von 0-255 in drei Kanälen  $256^3$  Kombinationen ergeben würden, wurden die Intensitäten der einzelnen Farbkanäle auf 8 Stufen reduziert.

### ***Joint Histogramm***

Im Joint Histogramm wird ein einfaches RGB Histogramm über Texturmerkmale weiter aufgeteilt. Dabei wird ein Pixel Rank definiert, bei dem die Pixelintensität eines Pixels mit den in der Nachbarschaft liegenden Pixeln verglichen wird. Hieraus ergibt sich ein zweidimensionales Histogramm bei dem Pixel anhand ihrer Farbe und dem Intensitätsrank eingeordnet werden.

### ***Color and Edge Directivity Deskriptor (CEDD)***

Im CED Deskriptor werden Farb- und Texturinformationen in ein Histogramm integriert (Chatzichristofis und Boutalis, 2008a). Hierbei wird ein Fuzzy Farbhistogramm, in dem das Bild in 24 verschiedene Farben gruppiert wird, mit 6 verschiedenen Kantentypen (keine, ohne Richtung, vertikal, horizontal,  $45^\circ$  diagonal und  $135^\circ$  diagonal) kombiniert. Dabei erfolgt im Fuzzy Farbhistogramm keine absolute Zuordnung in bestimmte Farbklassen, sondern es wird gewichtet, wie hoch die Angehörigkeit zu einer Klasse ist.

### ***Fuzzy Color and Texture Histogram (FCTH)***

Im FCT Histogramm wird das gleiche Fuzzy Farbhistogramm wie beim CEDD verwendet. Dieses wird jedoch mit einer ausführlicheren Kantenbeschreibung, die die

Kanten in 8 Kategorien unterteilt, kombiniert (Chatzichristofis und Boutalis, 2008b).

### ***Zernike Momente***

Ausgehend von Zernike-Polynomen können nach Teague (1980) Momente ermittelt werden, die die äußere Form beschreiben. Zur Berechnung wurden Routinen der ImgLib2 Bibliothek (Pietzsch et al., 2012) in ImageJ integriert.

### ***Opponent, Fuzzy Opponent und Joint Opponent Histogram***

Ein Opponent Histogramm kombiniert drei eindimensionale Histogramme, die auf dem entgegengesetzten Farbraum basieren und ergibt sich nach Van De Sande et al. (2010) für die drei Farbkanäle rot  $R$ , grün  $G$  und blau  $B$  wie folgt:

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \quad (3.23)$$

$O_1$  und  $O_2$  repräsentieren Farbinformationen und  $O_3$  Intensitätsinformationen des Bildes.

Im Fuzzy Opponent Histogramm erfolgt keine absolute Eingliederung der Pixelinformation, sondern es wird gewichtet, wie hoch die Angehörigkeit zu einer Klasse ist. Im Joint Opponent Histogramm wird analog zum Joint Histogramm eine weitere Unterteilung über den Pixelrang durchgeführt.

### ***Gabor Textur Merkmale***

Zur Bestimmung dieser Merkmale wird das Bild über eine Gabor Filter vorverarbeitet, um charakteristische Strukturen des Bildes für verschiedene Orientierungen und Frequenzen herauszufiltern. Extrahierte Merkmale ermöglichen anschließend eine vereinfachte Strukturbeschreibung.

### **Erstellung von Trainings- und Testdaten**

Für das Trainingsset wurden alle Monokulturen nach den unter 3.2.1 - Sedimentation von Phytoplanktonproben und 3.2.2 - Planktonerkennung beschriebenen Methode aufgenommen. Die Bilder wurden segmentiert und es wurden alle Merkmale für jeden segmentierten Partikel berechnet. Anschließend wurden die Partikel von Hand in die folgenden Kategorien unterteilt: Detritus, falsch segmentiert oder unbekannte Partikel oder eine der Gruppen des jeweiligen Taxons. Es wurden ca. 600 Partikel je Klasse verwendet.

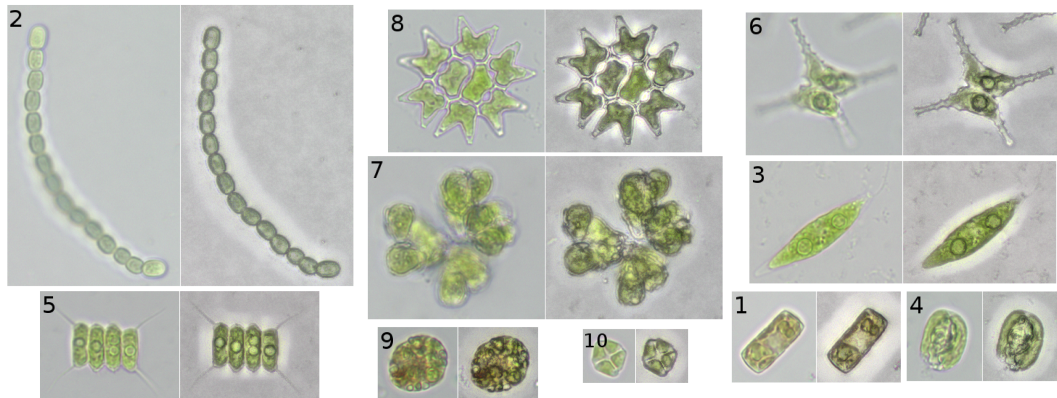
Für das Testset wurden Mischproben aller Taxa verwendet und analog zum Trainingsset verarbeitet. Es wurde Daten von verschiedenen Mischproben verwendet, um die Auswahl eines übertrainierten Klassifikators zu verhindern. Dabei wurden die einzelnen Proben unabhängig voneinander vorbereitet, mikroskopiert und bearbeitet.

#### **Daten für die Etablierung des Systems**

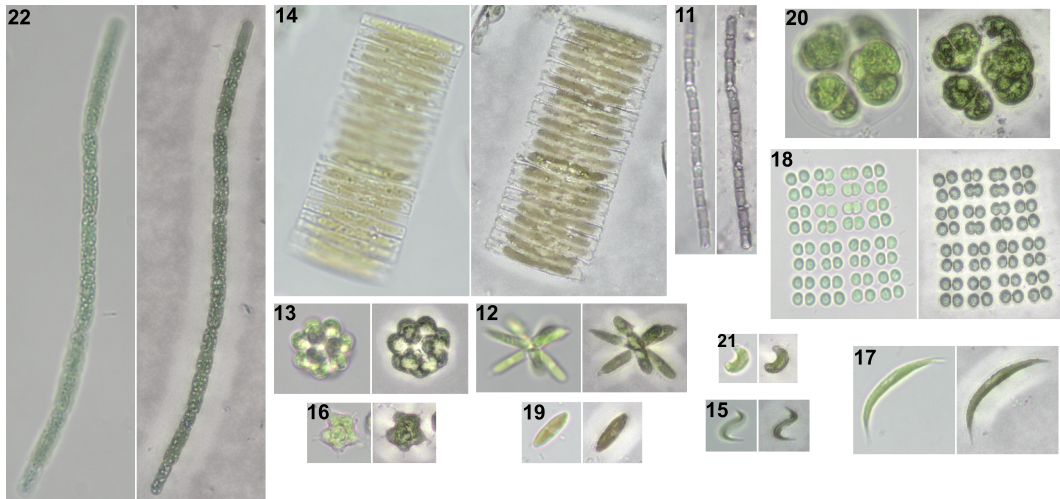
Für eine erste Überprüfung des Systems wurden zunächst Daten von zehn verschiedenen Taxa verwendet (Abb. 3.2(a)). Das Testset bestand aus Daten von vier verschiedenen Mischproben.

#### **Erweiterung der Daten**

Nach der Etablierung des Systems mit zehn verschiedenen Taxa wurde das Trainingsset auf insgesamt 22 Indikatortaxa erweitert (Abb. 3.2(b)). Das Testset wurde um Daten von vier weiteren Mischproben erweitert, in denen die bereits verwendeten Arten mit neuen Arten gemischt wurden.



(a) Aufnahmen der zehn Taxa, die für die Etablierung des Systems verwendet wurden: 1 - *Cyclotella*, 2 - *Anabeana*, 3 - *Chlorogonium*, 4 - *Cryptomonas*, 5 - *Desmodesmus*, 6 - *Staurastrum*, 7 - *Botryococcus*, 8 - *Pediastrum dup.*, 9 - *Trachelomonas* und 10 - *Crucigenia* (Schulze et al., 2013).



(b) Aufnahmen der Taxa, die bei der Erweiterung der Daten in die Analyse integriert wurden: 11 - *Limnothrix*, 12 - *Actinastrum*, 13 - *Coelastrum*, 14 - *Fragilaria*, 15 - *Monoraphidium cont.*, 16 - *Tetradron*, 17 - *Ankistrodesmus*, 18 - *Merismopedia*, 19 - *Nitzschia*, 20 - *Pandorina*, 21 - *Monoraphidium min.* und 22 - *Aphanizomenon*.

Abbildung 3.2: Mikroskopische Hellfeld und Quick Full Focus Bilder aller Taxa, die als Testdaten für das System verwendet wurden. Für jedes Taxon ist jeweils eine Hellfeldaufnahme (linke Seite) und eine QFF-Aufnahme (rechte Seite) abgebildet.

## Training und Klassifikation mit Testdaten zur Etablierung

Für die Klassifikation des ersten Testsets wurde ein künstliches neuronales Netz verwendet. Hierfür wurde Encog, ein Java Framework für neuronale Netze, neu in ImageJ integriert (Heaton, 2010).

Künstliche neuronale Netze bilden modellhaft die Funktionsweise von biologischen neuronalen Netzen nach. An das biologische System angelehnt, sind die kleinsten Funktionseinheiten Neuronen. Sie sind in der Lage Signale aufzunehmen, diese über Aktivierungsfunktionen zu verarbeiten und anschließend das verarbeitete Signal weiterzuleiten. Durch eine Vernetzung werden komplexe Klassifikationsprozesse möglich. In den durchgeführten Arbeiten kamen einfache Feedforward Netze zum Einsatz, in denen die verknüpften Neuronen in unterschiedlichen Schichten angeordnet sind (Abb. 3.3).

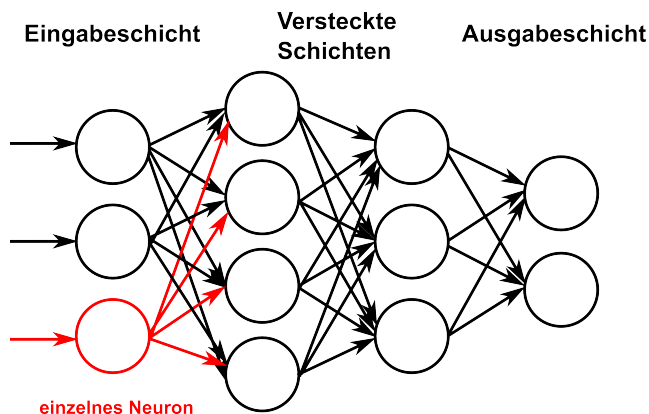


Abbildung 3.3: **Aufbau eines einfachen neuronalen Feedforward Netzes.** Die Grundeinheit bildet ein einzelnes Neuron (rot), welches ein Eingangssignal aufnimmt, verarbeitet und als Ausgangssignal weiterleitet. Durch die Verknüpfung von einzelnen Neuronen und die Anordnung in verschiedene Schichten werden komplexe Verarbeitungen möglich. In der Eingabeschicht werden Daten in das Netz eingespeist, die versteckte Schicht dient der Verarbeitung der Daten und die Ausgabeschicht ermöglicht eine Einteilung der Daten in verschiedene Klassen. In den durchgeführten Arbeiten wurde jedes berechnete Merkmal über ein einzelnes Neuron der Eingabeschicht in das Netzwerk eingespeist. Die einzelnen Neuronen in der Ausgabeschicht repräsentierten die zu unterscheidenden Phytoplankter.

Über die Eingabeschicht werden Daten, die für die Klassifikation genutzt werden sollen, in das Netz eingespeist. In verschiedenen versteckten Schichten, die in unter-

schiedlicher Anzahl auftreten können, findet die eigentliche Verarbeitung der Daten statt. In welche Klasse die Eingangssignale anschließend eingeordnet werden ist davon abhängig, welches Neuron in der Ausgabeschicht aktiviert wird. In den Arbeiten wurden dabei die zuvor berechneten Merkmale als Eingangssignale genutzt. Jedes Merkmale wurde hier einem Neuron in der Eingabeschicht zugeordnet. Die Anzahl der Neuronen in der Ausgabeschicht entsprach der Anzahl an zu unterscheidenden Klassen und ermöglichte so die Unterscheidung der Phytoplankter.

Die genaue Arbeitsweise eines Neurons ist in Abbildung 3.4 verdeutlicht. Die einge-

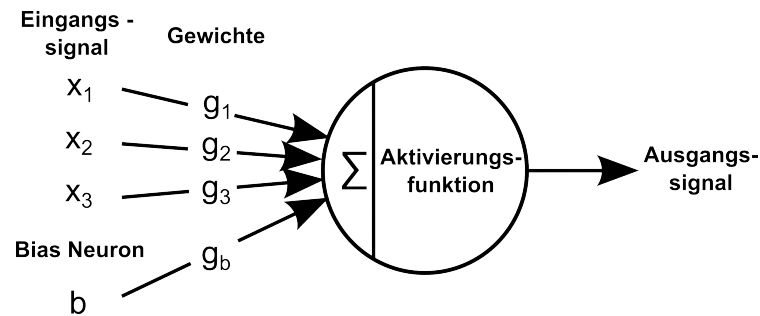


Abbildung 3.4: **Funktionsweise eines einzelnen Neurons.** Eingehende Signale werden durch Multiplikation mit einem Gewicht in ihrer Bedeutsamkeit eingestuft und anschließend aufaddiert. Anschließend wird über eine Aktivierungsfunktion bestimmt, welches Ausgangssignal an die nächsten Neuronen weitergeleitet wird. Neben den normalen Eingabewerten kann zusätzlich ein Bias-Wert verwendet werden. Dieser ermöglicht eine bessere Anpassung der Aktivierungsfunktion an die Daten. Durch Veränderung der Gewichte kann das Verhalten der Neuronen verändert werden.

hende Signale werden durch Multiplikation mit einem Gewicht in ihrer Bedeutsamkeit eingestuft und anschließend aufaddiert. Dabei kann jedes Eingangssignal negativ (hemmend) bzw. positiv (erregend) reguliert werden. Über die Aktivierungsfunktion wird festgelegt, welches Ausgangssignal weitergeleitet wird. Hier kann beispielsweise ein einfacher Schwellwert eingesetzt werden. In diesen Arbeiten wurde eine Elliot Aktivierungsfunktion gewählt (Elliott, 1993). Dabei wird das Ausgangssignal über eine sigmoidähnliche Funktion auf den Wertebereich -1 bis 1 skaliert und ermöglicht somit eine zusätzliche Regulierung der Signalintensität. Zudem wurden alle Neuronen mit einem Bias Neuron kombiniert. Bias Neuronen ermöglichen es, über eine eigene Gewichtung, die Aktivierungsfunktion entlang der x-Achse zu verschieben und somit besser an den zu trennenden Wertebereich anzupassen. Der Bias Wert wurde dabei

auf 1 gesetzt.

Um ein Netz für eine Klassifikationsaufgabe zu trainieren, können mehrere Parameter verändert werden. Der bedeutendste Schritt ist die Anpassung der Gewichte, die mit den Eingangssignalen eines jeden Neurons multipliziert werden. Hierzu wurde der Resilient Backpropagation Algorithmus (Riedmiller und Braun, 1993; Igel und Hüsken, 2000) genutzt. Dieser wurde über die HybridStrategy()-Funktion mit einem genetischen Algorithmus als weitere Lernstrategie kombiniert. Dabei wurden aus 50 Netzen mit gleicher Struktur aber randomisiert veränderten Gewichten, 10% der Netze ausgesucht, die die beste Klassifikationsrate aufweisen. Zusätzlich wurden, nach dem Vorbild des genetischen splicings, unterschiedliche Teile der Netze randomisiert miteinander kombiniert, um verbesserte Netze (Nachfahren) zu generieren. In Anlehnung an natürliche Mutationen, wurden in 20 % der neu ausgewählten Netze zufällige Veränderungen an den Gewichten vorgenommen. Der genetische Algorithmus kam immer dann zum Einsatz, sobald mit der Resilient Backpropagation die Lernrate innerhalb von 10 Iterationen geringer als  $1 * 10^{-5}$  % war.

Das Training wurde entweder nach einer erreichten Fehlerrate von 0% oder nach 3000 Iterationen gestoppt. Anschließend wurde das neuronale Netz und die dazugehörigen Regeln der Normalisierung gespeichert, um sie für die Klassifikation nutzen zu können.

Einen weiteren Einfluss auf die Effizienz eines Netzes hat die Anzahl an versteckten Schichten und die Anzahl an Neuronen in diesen Schichten. Eine automatische Vorhersage der besten Einstellung ist nicht möglich und diese muss empirisch bestimmt werden. Hierzu wurde ein in Encog integrierte Pruning-Ansatz eingesetzt, bei dem in einem zuvor festgelegten Bereich eine maximale Anzahl an Schichten und Neuronen pro Schicht schrittweise zum Netz hinzugefügt wurden. Über ein kurzes Training mit festgelegter Anzahl an Iterationen wurde ermittelt, wie gut die jeweilige Anordnung geeignet ist und die beste Anordnung ausgewählt. Das Pruning wurde immer dann angewendet, wenn die Netzstruktur verändert wurde wie z.B. bei einer veränderten Anzahl an genutzten Merkmalen oder einer veränderte Anzahl an Ausgangsneuronen. Für die Auswahl der geeignetsten Merkmale wurden alle verfügbaren Merkmale für ein erstes Training genutzt. Basierend auf diesen Ergebnissen wurde die Signifikanz eines jeden Neurons mit Hilfe der in Encog integrierten Funktion PruneSelecti-

ve.determineNeuronSignificance() ermittelt und die am besten geeigneten Merkmale bestimmt. Zusätzlich wurde empirisch getestet, welche Merkmale die besten Ergebnisse liefern, indem diese einzeln zum Netz hinzugefügt wurden und der Einfluss auf eine bessere Klassifikation mit Hilfe des Testsets ermittelt wurde.

Um das trainierte Netz mit dem Testset überprüfen zu können, wurden die Daten nach den gespeicherten Regeln normalisiert und mit dem Netz klassifiziert. Zur Überprüfung der Ergebnisse wurden die einzelnen Bilder der Partikel in Ordnern entsprechen ihrer Klassifikationsergebnisse sortiert und mit den Ergebnissen einer Klassifikation von Hand verglichen.

### **Vergleich verschiedener Klassifikationsmethoden**

Neben dem neuronalen Netz wurde die Klassifikation der Daten mit weiteren Klassifikationsmethoden durchgeführt. Dies sollte eine bessere Einordnung der Ergebnisse, die mit dem neuronalen Netz erzielt wurden, ermöglichen. Es wurde die Software Weka (Hall et al., 2009) verwendet, die eine Vielzahl von verschiedenen Verfahren zur Verfügung stellt. Als alternative Klassifikatoren wurden folgende in Weka verfügbare Methoden überprüft. Die entsprechende Klassenbezeichnung in Weka ist jeweils in Klammern angegeben:

- Klassifikatoren, die auf Entscheidungsbäumen basieren:
  - Funktionelle Bäume (FT) (Gama, 2004)
  - Logistische Modell Bäume (LMT) (Landwehr et al., 2005; Sumner et al., 2005)
  - Alternierender Multi-Klassen Entscheidungsbaum (LADTree) (Holmes et al., 2002)
  - C4.5 Entscheidungsbaum (J48) (Quinlan, 1993)
  - Forest of Random Trees (RandomForest) (Breiman, 2001)
- Klassifikatoren, die die Daten über Funktionen abgrenzen
  - Support Vector Machine trainiert über Sequential minimal optimization (SMO) (Platt, 1998; Keerthi et al., 2001)



- Lineare Logistische Regression (SimpleLogistic) (Landwehr et al., 2005; Sumner et al., 2005)
- Multinomiale Logistische Regression (Logistic) (Cessie und Houwelingen, 1992)
- Large Linear Klassifikation (LibLinear) (Fan et al., 2008)
- Bayes'sche Netze, bei denen auf Grundlage des Bayestheorem eine Einteilung der Daten über bedingte Wahrscheinlichkeiten erfolgt:
  - Bayes Netzwerk (BayesNet) (Bouckaert, 2004)
  - Naiver Bayes Klassifikator (NaiveBayes) (John und Langley, 1995)
- Klassifikatoren, bei denen formelle Regeln erstellt werden:
  - FURIA - Fuzzy unordered Rule Induction Algorithm (FURIA) (Hühn und Hüllermeier, 2009)
  - RIPPER - Repeated Incremental Pruning to Produce Error Reduction (JRip) (Cohen, 1995)
- die K-nearest Neighbor Methode (IBk) (Aha et al., 1991)
- CHIRP - Composite Hypercubes on Iterated Random Projections (CHIRP) (Wilkinson et al., 2011)

Als Trainings- und Testdaten wurden die Daten zur Etablierung des Systems, bestehend aus zehn Taxa, verwendet. Für das Training wurde auf die Merkmalsauswahl zurückgegriffen, welche für das neuronale Netz die beste Klassifikationsrate ergab. Ergänzend wurde Weka zur automatisierten Auswahl von Merkmalen eingesetzt und die ReliefF - Methode verwendet (Robnik-Sikonja und Kononenko, 1997). Diese ist als gut geeignete Auswahlmethode beschrieben, wenn Wechselwirkungen innerhalb des Merkmalsets auftreten, wie es bei biologischen Daten oft der Fall ist (Hall und Holmes, 2003). Prinzip der Methode ist es, die Qualität einzelner Merkmale durch den Unterschied zwischen den jeweils ähnlichsten Mitgliedern der verschiedenen Klassen zu ermitteln. Hierfür wird aus einer Klasse eine Instanz I zufällig ausgewählt. ReliefF sucht nun nach den k ähnlichsten Nachbarn in der selben Klasse (nähester Treffer N) und den k ähnlichsten Nachbarn von jeder anderen Klasse (nähester Fehltreffer

F). Basierend auf den Werten für I, N und F wird anschließend der Qualitätswert  $W_i$  für jedes  $i$ -te Merkmal angepasst. Wenn die Instanz I und die Instanzen in N verschiedene Werte für das jeweilige Merkmal aufweisen, wird die Qualität gesenkt. Wenn jedoch die Instanzen I und F unterschiedlich sind, dann wird der Qualitätswert erhöht.

In diesen Arbeiten wurden für die ReliefF Methode die in Weka gesetzten Einstellungen mit  $k = 10$  für die Qualitätsanalyse verwendet. Anschließend wurden alle Merkmale, die einen Qualitätswert größer als 0.001 zeigten für die Klassifikation genutzt.

### **Training und Klassifikation mit dem erweiterten Datenset**

Für die Unterscheidung von zusätzlichen Planktern wurden nach dem Test verschiedener Klassifikationsmethoden eine Support Vektor Maschine (SVM) als zweites Klassifikationsverfahren eingesetzt. Hierfür wurden in Encog integrierte Methoden, die auf der LibSVM von Chang und Lin (2011) basieren, genutzt. Somit konnten diese in die bereits bestehende Trainingsroutine in ImageJ integriert werden. Zur Nutzung der SVM wurden die entsprechenden Merkmale auf einen Wertebereich zwischen 0 und 1 normalisiert, da hier keine Eingabe von negativen Zahlen möglich ist.

Wie von Noble (2006) beschrieben, ermöglicht eine SVM, dass zwei Klassen von Daten über eine lineare Funktion voneinander abgetrennt werden können. So ist dies im eindimensionalen Raum ein Punkt, im zweidimensionalen Raum eine Linie und folglich im dreidimensionalen Raum eine Fläche (Abb. 3.5 a-c). Dabei wird die Dimension von der Anzahl der verwendeten Merkmale bestimmt und ist meist deutlich höher. Auch in höheren Dimensionen wird eine lineare Funktion verwendet, die mathematisch extrapoliert werden kann und als Hyperebene bezeichnet wird.

Eine Besonderheit der Support Vektor Maschine ist, dass bei der Auswahl der Hyperebene ein maximaler Abstand (maximum-margin) zwischen den zu separierenden Klassen eingehalten werden muss. Dabei dienen die zur Hyperebene am nächsten liegenden Daten als Support Vektor (Abb. 3.5 c und d). Hierdurch wird die Fähigkeit, die richtige Klasse für unbekannte Daten zu ermitteln, maximiert. Um auch die Unterscheidung von Klassen mit sich widersprechenden Daten zu ermöglichen, kann als freier Parameter der soft margin  $c$  definiert werden. Dieser gibt an, wie viele

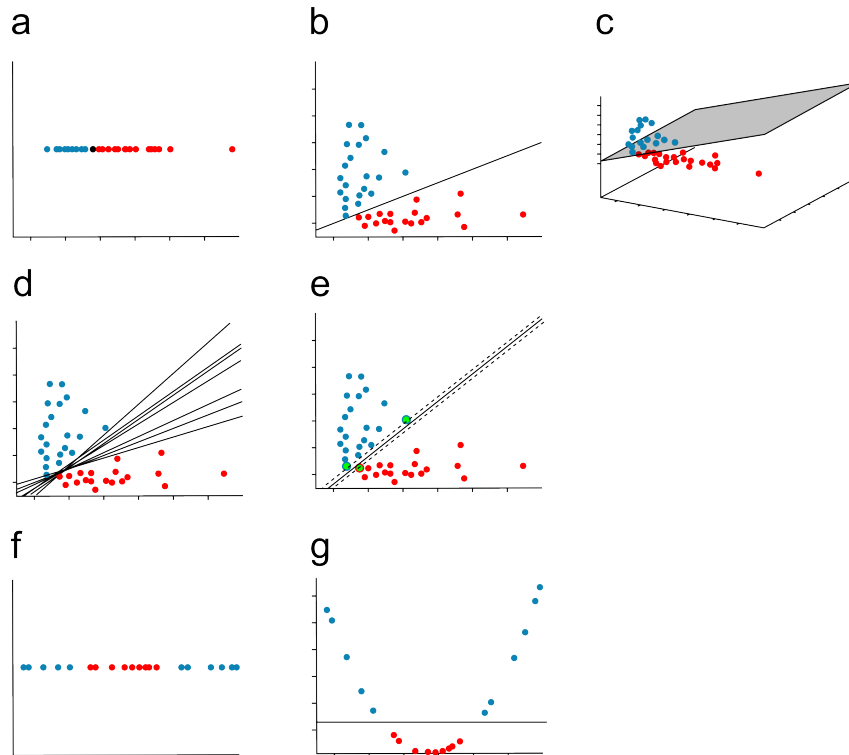


Abbildung 3.5: **Prinzipielle Funktionsweise einer Support Vektor Maschine.** Die Daten werden über die Hyperebene abgetrennt. Diese ist im (a) eindimensionalen Raum ein Punkt, (b) im zweidimensionalen Raum eine Linie und (c) im dreidimensionalen Raum eine Fläche. (d) Mehrere mögliche Hyperebenen, die eine Unterscheidung der Daten ermöglichen. (e) Die maximal margin Hyperebene mit drei Support Vektoren (grün). (f) Daten die im eindimensionalen Raum nicht linear voneinander getrennt werden können. (g) Über eine Kernelfunktion transformierte Daten werden im höher dimensionalen Raum trennbar (nach Noble (2006)).

Daten und mit welchem Abstand diese außerhalb der Klassengrenze liegen können, ohne das die endgültige Auswahl der Hyperebene beeinflusst wird.

Zur Separation von zwei Klassen, die in ihrer entsprechenden Dimension nicht über die Hyperplane unterschieden werden können, kommen Kernel-Funktionen zum Einsatz. Bei richtiger Auswahl transformieren sie die Daten so in einen höher dimensionalen Raum, dass sie dort wieder linear separiert werden können (Abb. 3.5 e und f). Bei Rücktransformation der linearen Funktion in die Ausgangsdimension ergeben sich somit komplexe nicht lineare Funktionen, die für die Trennung der Daten genutzt werden können.

Für die Versuche wurde als Kernelfunktion eine Gaußsche Radiale Basisfunktion genutzt, bei der die Daten über die Gaußsche Normalverteilung transformiert werden. Dabei wird die Glockenweite vom Parameter  $\gamma = -\frac{1}{2\sigma^2}$  reguliert. Der soft margin wurde eingesetzt, um das Auftreten von sich evtl. widersprechende Daten besser ausgleichen zu können. Für eine Optimierung der beiden Parameter  $c$  und  $\gamma$  wurde das Training über die Encog Funktion SVMSearchTrain() mit einer Gridsuche kombiniert. Als Trainingsmethode wurde die Sequential minimal optimization (Platt, 1998) genutzt.

Da eine einzelne SVM nur zwei Klassen voneinander unterscheiden kann, mussten mehrere SVM's erstellt werden. Diese wurden nach einem eins-gegen-eins Prinzip, dass in die LibSVM integriert ist, trainiert. Hierfür werden für jedes mögliche Pärchen Klassifikatoren erstellt und jeder wird mit den entsprechenden Daten trainiert. In der Klassifikation werden die Daten nun von jeder SVM in eine der beiden Klassen eingeteilt, für die die SVM trainiert wurde. Anschließend kann ermittelt werden, wie viele Klassifikatoren für welche Klasse gewählt haben. Die Daten werden dann in die Klasse eingeteilt, für die die meisten SVM's gewählt haben (Hsu und Lin, 2002).

Für das neuronale Netz wurde Training und Klassifikation wie zuvor beschrieben durchgeführt. Dabei wurde zusätzlich eine equilaterale Normalisierung der Klassen vorgenommen. Da bei der Fehlerberechnung immer der durch das Netzwerk ermittelte Ausgabewert mit dem idealen Ausgabewert verglichen wird, hat die Position der Klassen innerhalb der Ausgabeschicht einen Einfluss auf die Fehlerrate und kann diese verfälschen. Über eine equilateralen Normalisierung können die idealen Werte der Ausgabeschicht so angepasst werden, dass alle Klassen bei der Fehlerberechnung den gleichen Abstand voneinander aufweisen. Als Abstandsmaß wurde der Euklidische Abstand verwendet (Heaton, 2010). Für die Merkmalsauswahl wurde auf die zuvor beschriebene automatische Merkmalsauswahl in Weka mit der ReliefF - Methode zurückgegriffen. Wenn Merkmale nur aus einzelnen Ziffern bestanden, wurde hier ein empirisch ermittelter Qualitätsschwellwert von 0.001 gesetzt. Für Merkmale, die mehrere Ziffern beinhalten, wurden die Qualitätswerte gemittelt und bestimmt, wie hoch der Anteil an negativen Qualitätswerten ist. Hier wurde für die Auswahl ein Schwellwert von 0.001 für die mittlere Qualität und ein Anteil an negativen Werten von weniger als 15 % gesetzt.

## 4 Ergebnisse

### 4.1 Automatisierte Identifikation von Süßwasser - Phytoplankton

In diesem Abschnitt werden Ergebnisse für die automatisierte Erkennung verschiedener Phytoplankter dargestellt. Diese stellen einen wichtigen Parameter für die Wasserqualitätsanalyse dar. Eine Automatisierung von Mikroskopie und Bildauswertung soll helfen, die sonst sehr aufwendige Analyse zu vereinfachen.

Zunächst soll ein Überblick über den Ablauf der Analyse gegeben werden. Wie in Abbildung 4.1 zusammengefasst dargestellt, können die zwei Hauptschritte Mikroskopie und Bildanalyse unterschieden werden.

Zunächst wird die zu untersuchende Probe fixiert und in einer Utermöhl-Kammer sedimentiert. Für die Automatisierung wurde eine Mikroskopieroutine definiert, die es ermöglicht, dass auf verschiedenen Positionen entlang eines festgelegten Pfades verschiedene Bilder aufgenommen werden können. Hierbei werden Bilder im Hellfeld und für die Fluoreszenz aufgenommen. Anschließend erfolgt im Schritt der Bildauswertung die Erkennung der in den Mikroskopiebildern abgebildeten Organismen. Die Organismen werden dabei vom Hintergrund segmentiert, Merkmale berechnet und diese mit entsprechendem Einzelbild abgespeichert. Diese können nun entweder sortiert und für ein Training des Klassifikators genutzt werden, oder mit Hilfe eines trainierten Klassifikators analysiert werden. Für die Klassifikation wurden neuronale Netze oder Support Vektor Maschinen genutzt.

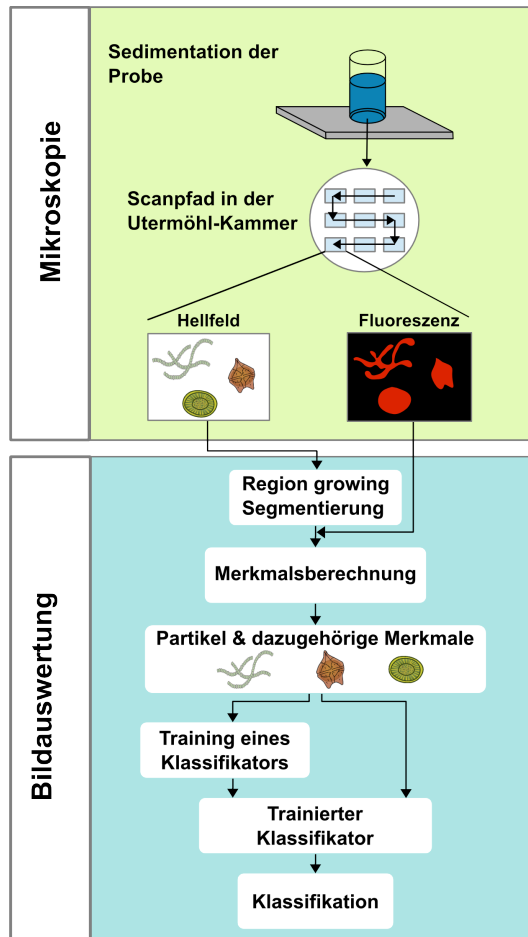


Abbildung 4.1: **Gesamtablauf der Erkennung von Phytoplanktern in Wasserproben.** Nach der Sedimentation der Probe werden mit Hilfe von automatisierter Mikroskopie verschiedene Bilder im Hellfeld und für die Fluoreszenz aufgenommen. In der Bildauswertung werden die Organismen vom Hintergrund getrennt und Merkmale berechnet. Anschließend kann ein Klassifikator anhand von erstellten Trainingsdaten trainiert werden und für eine Unterscheidung von unbekanntem Bildern genutzt werden (Schulze et al., 2013).

### 4.1.1 Automatisierte Mikroskopie

#### Hellfeld

Für die verwendeten Proben wurden normale Hellfeldbilder und Quick Full Focus Bilder (QFF), bei denen mehrere Fokusebenen in ein Bild integriert werden, aufgenommen. Mit den Hellfeldbildern war es vor allem in Mischproben oft nicht möglich alle Organismen gleichzeitig scharf abzubilden, da nur eine Fokusebene in das Bild integriert wird. In den neu verwendeten Quick Full Focus Bildern konnten sowohl in den verwendeten Monokulturen als auch in den Mischproben alle Organismen scharf in einem Bild aufgenommen werden (Abb. 4.2). Hierbei ist es allerdings notwendig Vibrationen während der Bildaufnahme zu unterbinden, um Bewegungen und daraus resultierende Artefakte im Bild zu vermeiden.

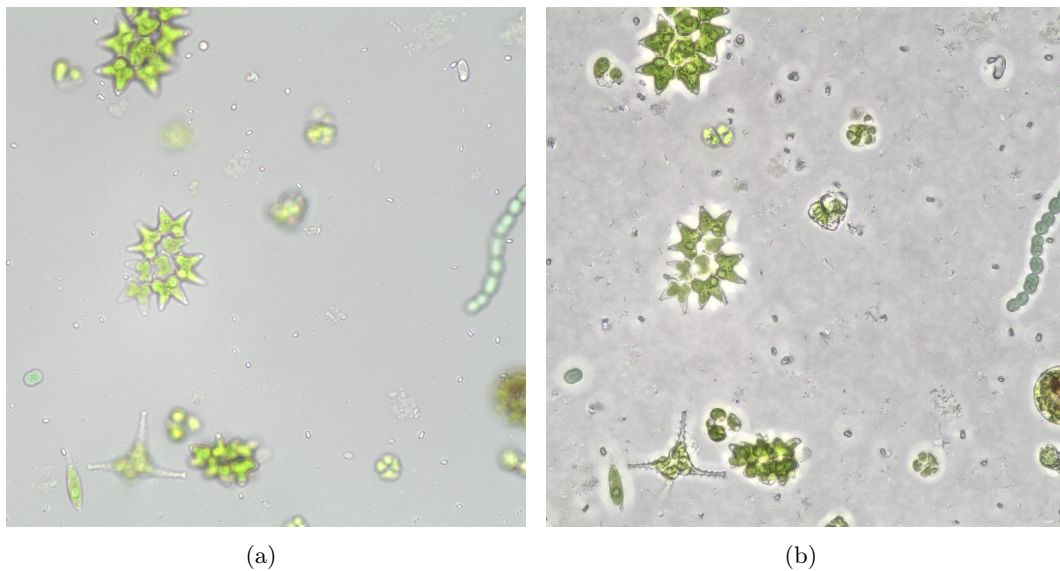


Abbildung 4.2: **Beispielhafter Vergleich zwischen einem normalen Hellfeldbild (a) und einem QFF Bild (b) einer Mischprobe.** In den Hellfeldbildern wurden Organismen nur teilweise scharf abgebildet oder konnten gar nicht in ihrer Fokusebene erfasst werden. In den QFF Bildern konnte alle Organismen vollständig abgebildet werden.

Bilder die innerhalb des definierten Aufnahmebereiches aufgenommen wurden zeigten eine konsistente Qualität in der Bildausleuchtung und Hintergrundfarbe und es konnten keine signifikanten Unterschiede in den Charakteristika der Proben während der Mikroskopie beobachtet werden (Abb. 4.3). Aufgrund eines unebenen Bodens der Sedimentationskammer musste ein Auto Fokus für jede mikroskopierte Position verwendet werden.

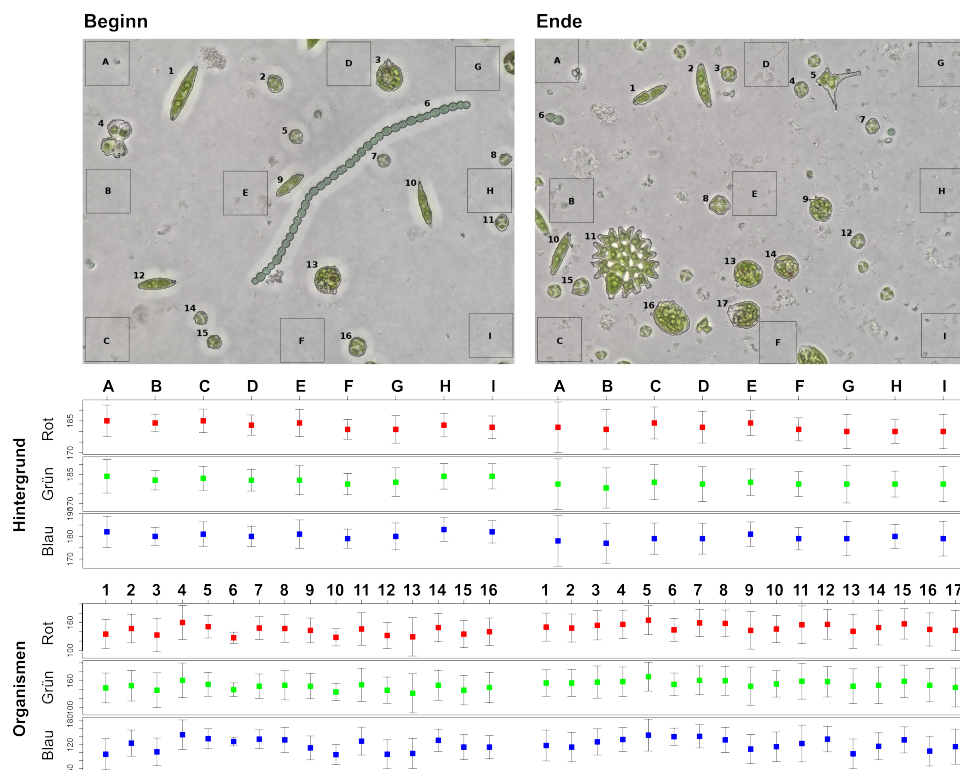


Abbildung 4.3: **Beispielhafter Vergleich der Bilder, die zu Beginn und zum Ende eines Mikroskopiedurchlaufs aufgenommen wurden.** Es wurden die mittlere Helligkeit und Standardabweichung der einzelnen Farbkanäle rot, grün und blau für verschiedene Bereiche des Hintergrundes (A-I) und für verschiedene Organismen im Bild (1-17) ermittelt.

## Fluoreszenz

Über die Fluoreszenzmikroskopie wurde die Autofluoreszenz verschiedener Photopigmente registriert. Chlorophyll a, welches in allen Phytoplanktern vorhanden ist (siehe hierzu der Vergleich in Tabelle 1.3), sollte dabei für eine gute Abgrenzung der



Planktonorganismen von restlichen, in der Probe vorhandenen Partikel genutzt werden. Zur Überprüfung dieser Annahme wurden für die verwendeten Organismen im Trainingsset Mittelwert und Standardabweichung jeder Klasse ermittelt (Abb. 4.4).

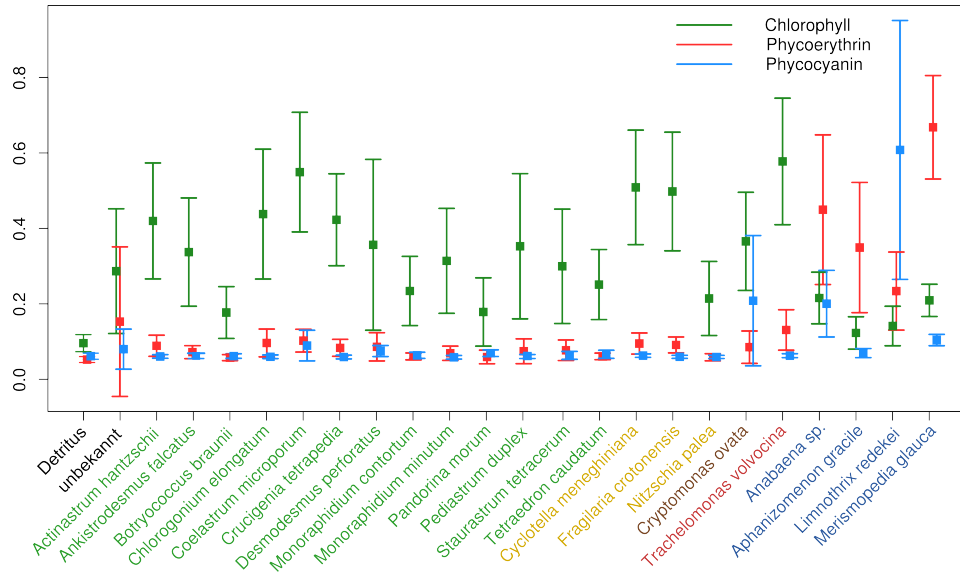


Abbildung 4.4: **Mittlere normierte Fluoreszenzintensität verschiedener Photopigmente für alle verwendeten Phytoplankter.** Es wurde Mittelwert und Standardabweichung der Chlorophyll-, Phycoerythrin- und Phycocyanin-Fluoreszenzintensität aller verwendeten Phytoplankter ermittelt. Hierzu wurden die Daten der erstellten Trainingssets genutzt. Alle Werte wurden auf den Bereich 0 - 1 normiert. Folgende Organismusgruppen sind vertreten und im Namen farblich markiert: Chlorophyceae (grün), Bacillariophyta (gelb), Cryptophyta (braun), Euglenophyta (rot) und Cyanobakterien (blau).

Dabei ist gut zu erkennen, dass die Partikel der Klasse "Detritus", die alle Partikel, die nicht dem Plankton zugeordnet werden können beinhaltet, keine Fluoreszenz aufweist. Für alle Phytoplankter ergaben sich Fluoreszenzsignale, wobei zwischen den einzelnen Arten deutliche Unterschiede in der Intensität festgestellt werden konnten. Hierbei soll angemerkt werden, dass der verwendete Fluoreszenzfilter so gewählt wurde, dass Chlorophyll a, b & c angeregt wurde und somit die Intensität des Signals abhängig von der Pigmentzusammensetzung der Organismen war.

So besitzen Cyanobakterien nur Chlorophyll a in ihrem Photosystem und zeigten im Vergleich zu vielen verwendeten Taxa, die entweder Chlorophyll a & b (Chlorophy-

ceae und Bacillariophyta) oder Chlorophyll a & c (Cryptophyta und Euglenophyta) enthalten, eine geringere Fluoreszenzintensität. Allerdings ergaben sich auch bei *Botryococcus*, *Monoraphidium contortum*, *Monoraphidium minutum*, *Nitzschia*, *Pandorina*, *Staurastrum* und *Tetraedron* vergleichbar geringe Intensitäten, obwohl diese Chlorophyll a & b bzw. c in ihrem Photosystem enthalten. Dies ist wahrscheinlich auf eine nicht durchgehende Pigmentierung der Organismen oder eine geringe Organismusgröße zurückzuführen.

Für eine bessere Unterscheidung innerhalb der Phytoplanktonklassen sollte die unterschiedliche Pigmentierung der verschiedenen Spezies genutzt werden. Dabei wurde die Autofluoreszenz von Phycoerythrin und Phycocyanin während der Mikroskopie aufgenommen. Signale für diese Pigmente konnten in unterschiedlichen Kombinationen und Intensitäten für die Organismusgruppen der Cyanobakterien und Cryptophyceae aufgenommen werden.

Für alle Fluoreszenzsignale wurden zudem innerhalb der jeweiligen Organismusgruppe große Schwankungen zwischen den einzelnen Individuen beobachtet.

Während der Aufnahme der Fluoreszenz führten selbst geringe Belichtungszeiten an einer Position zu einem Ausbleichen der Probe über diese Position hinaus. Um diesen Effekt so gering wie möglich halten zu können, wurde eine minimale Distanz von 500 µm zwischen den Aufnahmepositionen festgelegt. Zusätzlich wurde die Signalverstärkung bei der Bildaufnahme so hoch wie möglich gesetzt (ohne die folgende Bildanalyse durch ein zu hohes Rauschen in den Bildern zu stören), um die Belichtungszeit so gering wie möglich halten zu können. Nach diesen Adaptionen konnten keine weiteren Veränderungen der untersuchten Proben, die die Analyse stören würden, während der Bildaufnahme beobachtet werden.

#### **4.1.2 Bildverarbeitung und Bilderkennung**

##### **Adaption von ImageJ**

Alle Verarbeitungs- und Auswertungsaufgaben wurden in ImageJ integriert. Hierbei handelt es sich um eine Bildverarbeitungssoftware die vorrangig für die Analyse von biologischen Bildern entwickelt wurde und hierfür wichtige Methoden bereits zur Verfügung stellt. Die Funktionalität kann über Plugins erweitert werden und aufgrund eines weit verbreiteten Einsatzes stehen eine Vielzahl von veröffentlichten

Methoden zur Verfügung. Vorteil zu den in der Bildverarbeitung oft verwendeten Scriptsprachen Matlab oder IDL ist, dass es sich um ein Open Source Projekt mit offenem Quellcode handelt, das kostenfrei verwendet werden kann. Durch die Implementierung in Java kann es zudem ohne Mehraufwand auf allen Betriebssystemen eingesetzt werden.

Allerdings ist ImageJ nicht für den Einsatz von komplexen Klassifikationsaufgaben vorgesehen. Um dies zu ermöglichen, mussten zusätzliche Funktionen integriert werden. Hierzu zählt eine Region Growing Segmentierung und die Berechnung verschiedener neuer Merkmale. Zudem stellt ImageJ keine Klassifikationsmethoden zur Verfügung. Hierfür wurde die Encog Bibliothek, die vorrangig neuronale Netzstrukturen zur Verfügung stellt, in das System integriert.

### **Segmentierung**

In diesem Schritt der Bildverarbeitung werden alle Partikel, die im Bild dargestellt sind, vom Hintergrund abgetrennt und anhand ihrer Außenlinie registriert. Beim Vergleich beider Bildtypen (Hellfeld und QFF Bilder) war die Segmentierung der einzelnen Organismen in den QFF Bildern besser möglich (Abb. 4.5). Ursache war vorrangig, dass einzelne Organismen in den Hellfeldbildern aufgrund ihrer Lage oder Größe oft nicht vollständig fokussiert abgebildet wurden. Zudem konnten bei Mischproben oft nicht alle Organismen gleichzeitig scharf im Bild dargestellt werden, da hier zusätzliche Unterschiede im Sedimentationsverhalten auftraten. Durch die häufig unscharfe Abbildung ergaben sich zudem Folgefehler in der Verarbeitung. So konnte auch bei korrekter Segmentierung der Organismen nicht immer eine fehlerfreie Merkmalsberechnung und folgende Klassifikation sicher gestellt werden. Durch die Integration von verschiedenen Schärfeniveaus in ein Bild, konnten diese Folgefehler für die QFF Bilder vermieden werden. Darauf basierend wurden nur noch QFF Bilder für das folgende Training und Testen des Systems verwendet.

### **Merkmalsauswahl und Klassifikation mit neuronalen Netzen**

Nach Segmentierung und Registrierung der einzelnen Partikel wurden verschiedene Merkmale berechnet, die unterschiedliche Charakteristika der Partikel beschreiben. Diese Merkmale werden in der anschließenden Klassifikation zur Unterscheidung der



Abbildung 4.5: **Beispielhafte Ergebnisse der Segmentierung in mikroskopischen Hellfeld und Quick Full Focus Bildern.** Die Bilder zeigen die Segmentierung von Taxa, die eine gewisse dreidimensionale Struktur aufweisen. Der vom Hintergrund abgegrenzte Bereich ist mit einer schwarzen Linie markiert. Es sind Ergebnisse für Hellfeldbilder (linke Seite) und die entsprechenden QFF Bilder (rechte Seite) dargestellt (Schulze et al., 2013).

verschiedenen Planktonspezies genutzt.

Im Weiteren werden die Ergebnisse für ein Testset, bestehend aus zehn verschiedenen Planktontaxa, dargestellt. Hierbei soll angemerkt sein, dass der Einsatz aller berechneten Merkmale für das genutzte Set an Spezies nicht die besten Ergebnisse lieferte. Da nicht ausgeschlossen werden kann, dass diese für die Unterscheidung anderer Spezies gut einsetzbar sind wurde die Integration im ImageJ Plugin jedoch beibehalten.

Für die Klassifikation zeigte sich, dass neben dem Phytoplankton noch weitere Partikel in der Probe unterschieden werden mussten. Hierbei handelte es sich vorrangig um weitere Probenbestandteile wie Detritus oder heterotrophe Bakterien. Zudem enthielten die Proben Phytoplanktonorganismen, die nicht die charakteristischen morphologischen Eigenschaften aufwiesen, ungünstig sedimentierten und eine Einordnung

des Bildes nicht möglich machten oder nicht korrekt segmentiert wurden. Es wurden somit zwölf verschiedene Klassen für die Unterscheidung definiert: je eine Klasse pro Phytoplankton Taxon, eine Klasse für andere Probenbestandteile und eine Klasse für Organismen, die aufgrund der morphologischen Eigenschaften des segmentierten Partikels (unabhängig davon, ob dies durch die Eigenschaften des Organismus selber oder durch Techniken der Bildgebung und Bildverarbeitung verursacht wurde) nicht unterschieden werden können.

Für die Unterscheidung des Testsets wurde zunächst der Einsatz von neuronale Netzen untersucht. Dabei zeigten erste Versuche mit einem Netz, welches für die gleichzeitige Unterscheidung aller Bestandteile genutzt werden sollte, dass unerkennbares Phytoplankton und andere Probenbestandteile nicht gut abgegrenzt werden konnten (Daten nicht gezeigt), da vor allem letztere eine große Variabilität in ihrer Größe, Form und Struktur aufwiesen. Hierauf hin wurde eine zweistufige Unterscheidung eingesetzt. Ein erstes neuronales Netz wurde für die Abtrennung aller Bestandteile, die nicht dem Phytoplankton zugeordnet werden können genutzt. Dabei zeigten diese Partikel keine Fluoreszenzsignale und waren meist transparent oder braun. Somit war die Abtrennung über ein empirisch ermitteltes Merkmalsset, bestehend aus Merkmalen zur Beschreibung von Farbe (Hue-Histogramm), Fluoreszenz (mittlere Helligkeit der Phycoerythrin-, Phycocyanin- und Chlorophyll-Fluoreszenz), Größe (Fläche) und Zirkularität gut möglich (Tabelle 4.1, S. 58, Klasse "det").

Im zweiten Schritt wurde ein weiteres Netz für die eigentliche Unterscheidung der Planktontaxa trainiert. Hierzu wurden zunächst alle vorhandenen Merkmale für ein Training genutzt und anschließend die Signifikanz der Neuronen in der Eingangsschicht bestimmt. Da jedes Neuron dieser Schicht jeweils ein Merkmal in das Netzwerk einspeist, kann so bestimmt werden, wie groß der Einfluss der einzelnen Merkmale auf die Verarbeitung im Netz ist (Abb. 4.6). Höchste Signifikanz zeigten Merkmale für Farbe (Sättigung und Hue) und Helligkeit. Zudem ergaben sich hohe Signifikanzen für Merkmale, die Texturen innerhalb der Partikel (Local Binary Patterns) und unterschiedliche Fluoreszenzen beschreiben. Basierend auf diesen Vorergebnissen, wurden Merkmale von Hand ausgewählt und mit dem Testset evaluiert. Dabei zeigten einige Merkmale, die im initialen Training eine hohe Signifikanz aufwiesen, bei der Evaluierung mit dem Testset nicht die entsprechend guten Klassifikationsra-

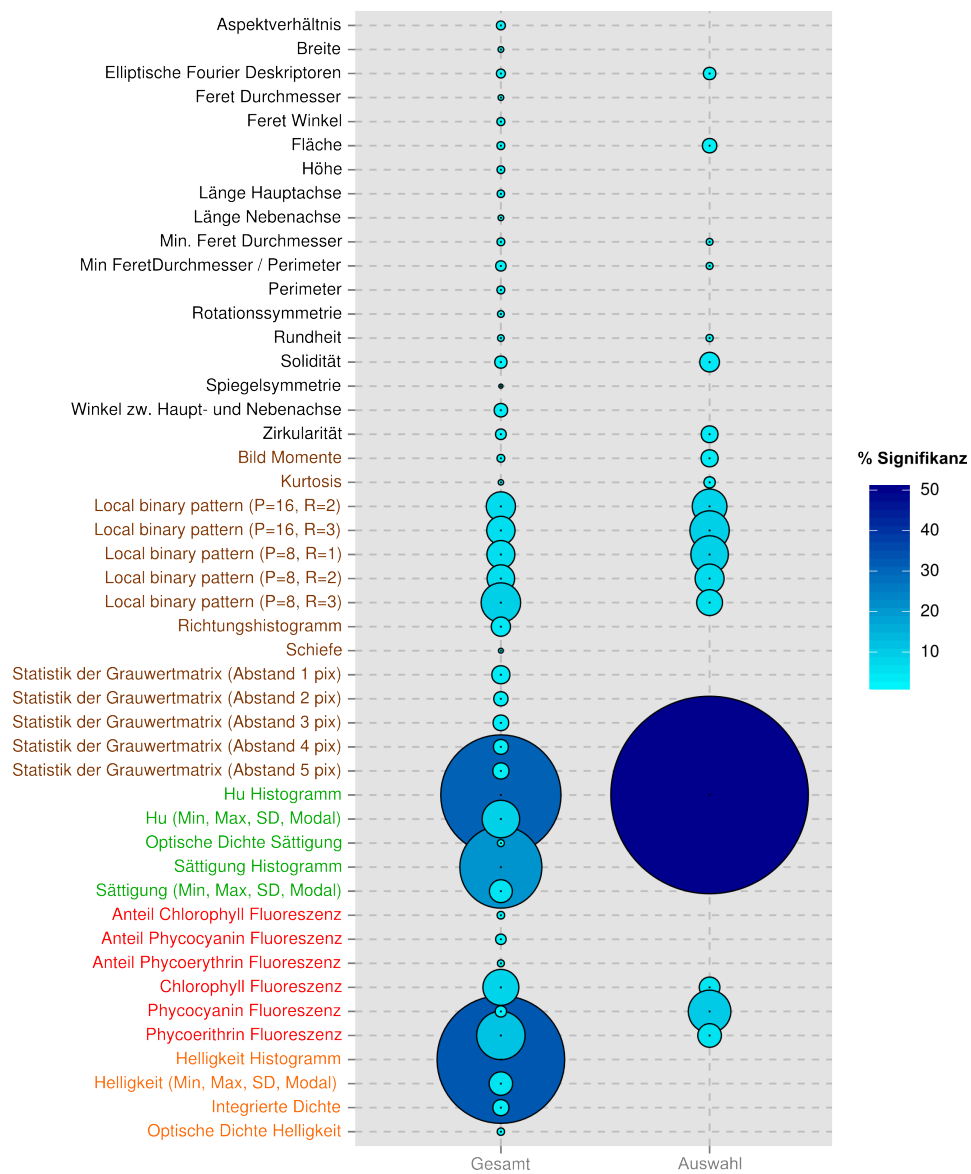


Abbildung 4.6: **Mittlere prozentuale Signifikanz der Input Neuronen geordnet nach jeweiligem Merkmal.** Der Graph zeigt die einzelnen Signifikanzen, die die Input Neuronen eines trainierten Netzes aufwiesen. Dabei können Neuronen mit hoher Signifikanz besser für die Unterscheidung genutzt werden. Bei Merkmalen die mehrere Inputneuronen nutzen, wurden die Daten als Mittelwert zusammengefasst dargestellt. Die linke Spalte (Gesamt) zeigt Ergebnisse für alle vorhandenen Merkmale. Die rechte Spalte (Auswahl) zeigt Ergebnisse für das endgültig ausgewählte Merkmalsset. Die Merkmale sind geordnet nach Beschreibung von Form (schwarz), innerer Textur (braun), Farbe (grün), Fluoreszenz (rot) und Helligkeit (orange).

ten und wurden nicht in das finale Merkmalsset integriert. Das endgültig verwendete Merkmalsset enthielt folgende Charakteristika: innere Textur (über Local Binary Patterns und Image Moment 1), Form (über Elliptische Fourier Deskriptoren 2-13, Zirkularität, Rundheit, Solidität, Minimaler Feretdurchmesser und Perimeter), Größe (über Fläche des Partikels) und Pigmentierung (Farbe über das Hue-Histogramm, Photopigmente über die jeweilige mittlere Helligkeit des Fluoreszenzsignales). Auch hierfür wurde die jeweilige Neuronensignifikanz der entsprechenden Merkmale ermittelt (Abb. 4.6 Spalte 2). Dabei zeigte sich, dass vor allem die Merkmale, die die innere Textur und die Pigmentierung des Organismus beschreiben einen großen Einfluss auf das Klassifikationsergebnis haben. Eine geringere Bedeutung kam Merkmalen zu, die die äußere Form oder Größe beschreiben. Allerdings waren auch diese für die erzielten Klassifikationsergebnisse erforderlich.

Für das ausgewählte Merkmalsset und eine Netzstruktur, bestehend aus 2 versteckten Schichten (erste Schicht mit 50 Neuronen, zweite Schicht mit 30 Neuronen) zeigte die Klassifikation der Testproben eine mittlere Erkennungsrate von 94.7%, die für die einzelnen Klassen im Bereich zwischen 89.9% und 99.9% lag (Tabelle 4.1, S. 58). Die durchschnittliche Falsch Positive Rate lag hier bei 5.5 % (mit einem Bereich zwischen 0% und 13.6%) und wurde durch übereinander liegende oder aggregierte Organismen, Organismen die nicht die typischen morphologischen Charakteristika ausbildeten, falsch segmentierte Partikel und Partikel, die nicht innerhalb des verwendeten Fokusbereichs lagen verursacht.

### **Vergleich mit weiteren Klassifikationsverfahren**

Zur besseren Einschätzung des Klassifikators wurden die erzielten Ergebnisse mit weiteren Klassifikationsverfahren verglichen. Hierzu wurde die Klassifikations-API Weka eingesetzt, da diese eine Vielzahl an verschiedenen Methoden zur Verfügung stellt. Sie basiert auf Java und würde eine gute Integration in ImageJ ermöglichen. Dabei wurden die Daten des bereits eingesetzten Merkmalssets für die Evaluation verschiedener Verfahren verwendet (Abb. 4.7 linke Spalte). Es zeigte sich, dass mit Hilfe des neuronalen Netzes sowohl die beste Erkennungsrate als auch die beste Falsch Positive Rate erzielt werden konnte. Ähnlich gute Ergebnisse ergaben sich mit einer Support Vektor Maschine.

Tabelle 4.1: **Konfusionsmatrix der Klassifikationsergebnisse für Mischproben aus zehn verschiedenen Taxa.** Die Daten wurden in folgende Klassen unterschieden: Detritus (det), unbekannte Plankton Organismen (unb), *Cyclotella* (1), *Anabeana* (2), *Chlorogonium* (3), *Cryptomonas* (4), *Desmodesmus* (5), *Staurastrum* (6), *Botryococcus* (7), *Pediastrum* (8), *Trachelomonas* (9) und *Crucigenia* (10). Das Testset bestand aus Daten von vier verschiedenen Proben. Um die Auswahl eines übertrainierten Klassifikators zu verhindern, wurden die Proben unabhängig voneinander und an unterschiedlichen Tagen vorbereitet, mikroskopiert und analysiert. Die Ergebnisse der automatischen Analyse (Spalten) wurden den Ergebnissen einer manuellen Unterscheidung gegenübergestellt (Zeilen). Korrekte Ergebnisse befinden sich in der Hauptdiagonalen und sind hervorgehoben. Erk. = Erkennung; FP = Falsch Positiv (Schulze et al., 2013)

		Automatische Erkennung												
	det	unb	1	2	3	4	5	6	7	8	9	10	% Erk.	%FP
det	<b>11061</b>	0	0	1	0	0	0	0	0	0	0	0	99.99	0.01
unb	0	<b>371</b>	0	2	0	4	3	0	7	3	0	13	92.06	8.09
1	0	3	<b>91</b>	0	0	0	0	0	1	0	1	0	94.79	6.59
2	0	5	3	<b>182</b>	0	0	0	0	2	0	0	1	94.30	3.85
3	0	1	0	0	<b>57</b>	1	1	0	0	0	0	0	95.00	0.00
4	0	0	0	2	0	<b>115</b>	0	0	0	0	0	0	98.29	8.70
5	0	1	0	0	0	0	<b>124</b>	2	8	3	0	0	89.86	4.03
6	0	7	0	2	0	0	0	<b>94</b>	0	0	1	0	90.38	2.13
7	0	2	1	0	0	0	0	0	<b>253</b>	2	1	2	96.93	7.51
8	0	0	0	0	0	0	0	0	0	<b>59</b>	3	0	95.16	13.56
9	0	1	0	0	0	5	0	0	0	0	<b>127</b>	0	95.49	4.72
10	0	10	2	0	0	0	1	0	1	0	0	<b>240</b>	94.49	6.25
∅													94.73	5.45



Weiterhin wurden für eine vollautomatische Merkmalsauswahl der Ranking Algorithmus ReliefF verwendet. Hierüber sollte zum einen ermittelt werden, inwiefern eine andere Zusammensetzung des Merkmalsset die Klassifikationsleistung der verschiedenen Verfahren beeinflusst. Zum anderen sollte dies einen Überblick darüber geben, wie gut die zuvor erfolgte halbautomatische Merkmalsauswahl durch ein automatisches Verfahren ersetzt werden kann (Abb. 4.7 rechte Spalte).

Dabei ergab sich, dass die Klassifikationsleistung nur bei wenigen Methoden leicht verbessert wurde und im Vergleich zu den Ergebnissen des neuronalen Netzes keine Verbesserung darstellte. Weiterhin zeigte sich, dass im Abhängigkeit vom verwendeten Klassifikationsverfahren auch mit der automatischen Merkmalsauswahl gute Klassifikationsergebnisse möglich waren.

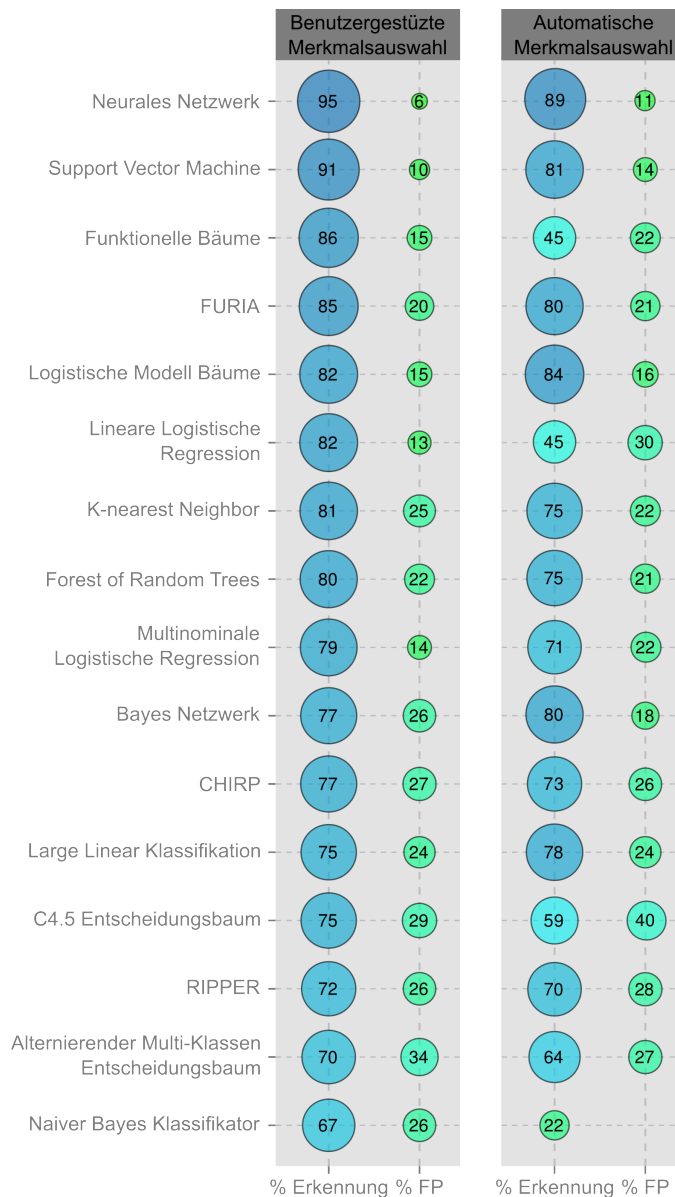


Abbildung 4.7: **Vergleich verschiedener Klassifikationsmethoden für die Unterscheidung von zehn verschiedenen Taxa.** Die Abbildung zeigt die Erkennungsrate (% Erkennung) und die Falsch Positiv Rate (% FP) für das gleiche Trainings- und Testset für verschiedene Klassifikationsmethoden. Die linken Spalte zeigt Ergebnisse für das Merkmalsset, mit dem die in Tabelle 4.1 dargestellten Ergebnisse erzielt wurden. Die rechte Spalte zeigt Ergebnisse für ein in Weka automatisch aus allen vorhandenen Merkmalen ausgewähltes Subset.

### 4.1.3 Vergrößerung des Testsets und Integration weiterer Merkmale

Nachdem die erste Unterscheidung von zehn verschiedenen Taxa möglich war, wurde das Testset um zwölf weitere Spezies erweitert. Hierbei wurde darauf geachtet Taxa zu verwenden, die in unterschiedlichen Charakteristika Ähnlichkeiten zu bereits verwendeten Arten aufweisen oder während ihres Wachstums unterschiedliche Strukturen ausbilden, um hier mögliche Schwachpunkte des Systems finden zu können. Für die Unterscheidung wurde ein neuronales Netz und eine Support Vektor Maschine trainiert. Die Support Vektor Maschine wurde dabei aufgrund der mit dem neuronalen Netz vergleichbar guten Klassifikationsrate in den vorigen Versuchen ausgewählt. Bei der Verwendung des zuvor ermittelten Merkmalsset ergaben sich für beide Verfahren keine nutzbare Erkennungsrate (Daten nicht gezeigt).

Die bereits erfolgte Einstufungen der unterschiedlichen Merkmale nach ihrer Wichtigkeit zeigte, dass vor allem Merkmale zur Beschreibung von Textur und Pigmentierung gut für die Unterscheidung genutzt werden konnten (siehe Abb. 4.6). In Folge dessen wurde die Merkmalsberechnung um verschiedene Methoden erweitert, die vor allem diese Charakteristika gut beschreiben (siehe Methoden - Abschnitt Merkmalsberechnung ab S. 25).

Mit der halb empirischen Anpassung des Merkmalsset war es nur noch schwer möglich gute Klassifikationsergebnisse zu erzielen und es wurde eine vollautomatische Merkmalsauswahl eingesetzt. Dabei wurde auf die bereits genutzte ReliefF Methode zurückgegriffen, da diese in vorhergehenden Ergebnissen ähnlich gute Klassifikationsergebnisse wie für das halbautomatisch ausgewählte Merkmalsset ermöglichte. Die ausgewählten Merkmale und entsprechende Qualitätsfaktoren sind in Tabelle 4.2 zusammengefasst. Aufgrund bereits bestehender Ergebnisse wurden Merkmale, die nur auf der Pixel - Helligkeit oder Sättigung beruhen nicht weiter betrachtet.

Weiterhin zeigte sich, dass mit steigender Klassenzahl die Position der Klasse in der Ausgabeschicht des Netzwerkes einen zunehmenden Einfluss auf die Ergebnisse hatte. Um diesen Effekt zu verhindern, wurde die Zuordnung zu den jeweiligen Klassen in der Ausgabeschicht des Netzwerkes equilateral normalisiert. Über ein Pruning des Netzwerkes ergab sich mit den automatisch ausgewählten Merkmalen eine optimale Netzwerkstruktur von einer versteckten Schicht mit 50 Neuronen. Mit den erwähnten Anpassungen und Erweiterungen wurde für das neurale Netz eine mittlere

Tabelle 4.2: **Qualitätswerte für die Merkmalsauswahl.** Mit Hilfe der ReliefF Methode wurden alle ermittelten Merkmale nach ihrer Wichtigkeit eingestuft und über einen Schwellwert ausgewählt. Um bei Merkmale, die sich aus mehreren Ziffern zusammensetzen eine bessere Auswahl zu ermöglichen, wurde hier der negativ gewichtete Anteil ermittelt. Grau gekennzeichnete Merkmale wurden anschließend für das Klassifikatortraining genutzt.

Merkmale	Qualität	Neg. Anteil [%]	Merkmale	Qualität	Neg. Anteil [%]
Rundheit	0.0733		Moment <sub>1</sub>	0.0034	
Zirkularität	0.0624		CEDD	0.0034	54
Minimum Feret / Perimeter	0.0555		Feret Durchmesser	0.0032	
CHL Fluoreszenz	0.0434	0	Opponent Histogramm	0.0029	80
Länge der Nebenchse	0.0350		FCTH	0.0028	77
Schiefe	0.0260		Joint Opponent Histogramm	0.0022	83
Einfache Merkmale (LIRE)	0.0253	13	Joint Histogramm	0.0020	77
Richtungshistogramm	0.0248	0	Aspektverhältnis	0.0019	
LBP <sub>(R=1, P=8)</sub>	0.0242	0	Farb-Autokorrelogramm	0.0018	65
Solidität	0.0220		RGB-Histogramm	0.0015	83
LBP <sub>(R=2, P=8)</sub>	0.0217	0	Elliptische Fourier Deskriptoren	0.0013	14
Kanten Histogramm	0.0213	0	Höhe	0.0010	
LBP <sub>(R=3, P=8)</sub>	0.0201	0	Gabor Textur Merkmale	0.0007	50
Minimum Feret	0.0188		HSB-Histogramm	0.0004	98
LBP <sub>(R=2, P=16)</sub>	0.0179		Fuzzy Opponent Histogramm	0.0002	98
DCT Koeffizienten Histogramm	0.0176	4	Breite	0.0002	
LBP <sub>(R=3, P=16)</sub>	0.0169	0	Punktsymmetrie	0.0000	
Achsensymmetrie	0.0129		Moment <sub>5</sub>	-0.0006	
Hue Histogramm	0.0115	2	Moment <sub>7</sub>	-0.0009	
PC Fluoreszenz	0.0105	0	Moment <sub>8</sub>	-0.0012	
Tamura Merkmale	0.0099	0	Moment <sub>6</sub>	-0.0012	
PE Fluoreszenz	0.0074	29	Feret Winkel	-0.0013	
Kurtosis	0.0065		Moment <sub>3</sub>	-0.0014	
Länge der Hauptachse	0.0059		Moment <sub>4</sub>	-0.0016	
Fläche	0.0058		Moment <sub>2</sub>	-0.0019	
Integrierte Dichte	0.0047		Zernik Momente	-0.0029	90
Perimeter	0.0047				

Erkennungsrate von 86.81 % erreicht, die für die einzelnen Klassen im Bereich von 70.19% und 97.42% lag. Die durchschnittliche Falsch Positiv Rate lag bei 11.89 %. Mit Ausnahme von *Fragilaria* (45.61 %) lagen die Werte für die jeweiligen Klassen im Bereich von 22.71 - 2.06 % (Tabelle 4.3). Für die Support Vektor Maschine ergab sich eine ähnliche mittlere Erkennungsrate von 82,6 %. Die Falsch Positiv Rate war im Vergleich zum neuronalen Netz mit 55.3 % deutlich erhöht (Tabelle 4.4).

Wie bereits zuvor dargestellt, erfolgte die Abtrennung von Detritus gut und wurde in einem vorgestellten Schritt mit einem dafür trainierten Netzwerk automatisch durchgeführt.

Tabelle 4.3: **Konfusionsmatrix der Klassifikationsergebnisse für die Erweiterung der Testdaten, die mit Hilfe eines neuronalen Netzes erzielt wurden.** Die Daten wurden in folgende Klassen unterschieden: Unbekannte Plankton Organismen (unb), *Cyclotella* (1), *Anabeana* (2), *Chlorogonium* (3), *Cryptomonas* (4), *Desmodesmus* (5), *Staurastrum* (6), *Botryococcus* (7), *Pediastrum* (8), *Trachelomonas* (9), *Crucigenia* (10), *Limnothrix* (11), *Actinastrum* (12), *Coelastrum* (13), *Fragilaria* (14), *Monoraphidium cont.* (15), *Tetraedron* (16), *Ankistrodesmus falcatus* (17), *Merismopedia* (18), *Nitzschia* (19), *Pandorina* (20), *Monoraphidium min.* (21) und *Aphanizomenon* (22). Die Ergebnisse der automatischen Analyse (Spalten) wurden den Ergebnissen einer manuellen Unterscheidung gegenübergestellt (Zeilen). Korrekte Ergebnisse befinden sich in der Hauptdiagonalen und sind hervorgehoben. Erk. = Erkennung; FP = Falsch Positiv

	Automatische Erkennung																						% Erk.	%FP	
	unb	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21			22
unb	1747	0	20	2	0	30	12	9	7	2	14	7	7	3	22	17	2	3	9	0	9	15	4	90.01	19.52
1	2	140	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	95.24	2.86
2	0	0	297	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	4	0	0	4	6	94.59	18.52
3	2	0	0	97	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92.38	2.06
4	1	3	0	0	101	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	86.32	9.9
5	16	0	0	0	0	295	1	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	93.06	22.71
6	5	0	0	0	0	1	139	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	93.92	10.07
7	32	0	1	0	10	1	0	254	3	11	0	0	0	0	1	0	0	0	0	0	3	0	0	80.38	7.48
8	2	0	0	0	0	0	0	0	151	2	0	0	0	0	0	0	0	0	0	0	0	0	0	97.42	6.62
9	15	0	0	0	0	8	0	0	0	105	0	0	0	0	0	0	0	0	0	0	5	0	0	78.95	14.29
10	46	0	0	0	0	0	0	0	0	0	198	0	0	0	0	0	2	0	0	1	0	3	0	79.2	7.58
11	9	0	0	0	0	0	0	0	0	0	0	246	0	0	9	1	0	10	0	0	0	0	1	89.13	5.28
12	7	0	1	0	0	0	1	0	0	0	0	0	130	0	1	5	0	2	0	0	1	0	0	87.84	5.38
13	14	0	0	0	0	0	0	0	0	0	0	0	0	108	0	0	16	0	0	0	0	7	0	74.48	3.7
14	0	1	0	0	0	2	0	0	0	0	0	0	0	0	114	0	0	0	0	0	0	0	0	97.44	45.61
15	20	0	1	0	0	0	0	0	0	0	1	0	0	0	0	222	0	1	0	0	0	1	0	90.24	14.41
16	11	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	139	0	0	0	0	0	0	92.05	17.27
17	38	0	2	0	0	0	0	0	0	0	0	3	0	0	5	3	0	210	0	0	0	0	1	80.15	7.62
18	9	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	170	0	0	0	0	91.4	8.82
19	11	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0	0	82.35	1.79
20	7	0	3	0	0	17	0	7	0	0	0	0	0	0	0	0	0	0	0	0	109	0	0	76.22	17.43
21	87	0	1	0	0	0	0	0	0	0	0	0	0	0	4	4	0	0	0	0	226	0	0	70.19	17.7
22	7	0	19	0	0	0	0	0	0	0	0	3	0	0	3	0	0	0	2	0	0	0	174	83.65	6.9
∅																								86.81	11.89

Tabelle 4.4: **Konfusionsmatrix der Klassifikationsergebnisse für die Erweiterung der Testdaten, die mit Hilfe einer Support Vektor Maschine erzielt wurden.** Die Daten wurden in folgende Klassen unterschieden: Unbekannte Plankton Organismen (unb), *Cyclotella* (1), *Anabeana* (2), *Chlorogonium* (3), *Cryptomonas* (4), *Desmodesmus* (5), *Staurastrum* (6), *Botryococcus* (7), *Pediastrum* (8), *Trachelomonas* (9), *Crucigenia* (10), *Limnothrix* (11), *Actinastrum* (12), *Coelastrum* (13), *Fragilaria* (14), *Monoraphidium cont.* (15), *Tetraedron* (16), *Ankistrodesmus falcatus* (17), *Merismopedia* (18), *Nitzschia* (19), *Pandorina* (20), *Monoraphidium min.* (21) und *Aphanizomenon* (22). Die Ergebnisse der automatischen Analyse (Spalten) wurden den Ergebnissen einer manuellen Unterscheidung gegenübergestellt (Zeilen). Korrekte Ergebnisse befinden sich in der Hauptdiagonalen und sind hervorgehoben. Erk. = Erkennung; FP = Falsch Positiv

	Automatische Erkennung																						% Erk.	%FP	
	unb	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21			22
unb	767	44	20	20	21	66	32	49	38	13	56	48	89	31	19	35	91	103	32	158	65	116	28	39.52	17.47
1	1	133	0	0	0	4	0	0	0	1	0	0	0	1	3	0	0	1	0	0	3	0	0	90.48	52.63
2	5	0	264	0	0	0	1	0	2	0	4	2	0	0	0	0	1	0	18	1	0	11	5	84.08	17.05
3	1	0	0	95	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	5	0	0	0	90.48	23.16
4	1	0	0	0	84	5	0	3	0	22	0	0	0	0	0	0	0	0	0	0	2	0	0	71.79	57.14
5	8	2	0	1	0	291	0	1	3	1	0	0	1	0	0	0	2	0	0	1	6	0	0	91.8	40.89
6	2	0	0	0	0	2	137	0	2	0	0	0	5	0	0	0	0	0	0	0	0	0	0	92.57	29.2
7	12	2	0	0	11	14	1	219	5	8	2	0	0	0	0	0	2	0	0	1	39	0	0	69.3	29.22
8	3	0	0	0	0	1	0	2	145	0	0	0	0	0	1	0	1	0	0	0	2	0	0	93.55	37.93
9	0	0	0	0	15	1	0	0	1	113	0	0	0	0	0	0	0	0	0	0	3	0	0	84.96	45.13
10	28	9	0	0	0	1	0	1	0	2	183	0	0	3	0	0	5	0	0	0	1	17	0	73.2	39.34
11	3	0	2	0	0	0	2	0	0	1	0	245	1	0	0	0	0	7	0	2	0	0	13	88.77	35.1
12	0	0	0	0	0	1	0	0	1	0	0	0	133	0	0	1	0	12	0	0	0	0	0	89.86	84.21
13	3	3	0	0	0	1	0	0	0	2	3	0	0	123	0	0	5	0	0	0	4	1	0	84.83	32.52
14	12	2	0	0	0	5	2	1	2	0	0	3	1	0	81	0	0	3	0	5	0	0	0	69.23	29.63
15	6	1	0	0	0	1	0	0	0	0	0	1	6	0	0	215	0	8	0	5	0	3	0	87.4	20
16	2	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	143	0	0	0	0	2	0	94.7	83.22
17	1	0	2	0	0	0	1	0	0	0	0	13	6	0	0	4	0	229	0	1	0	0	5	87.4	59.83
18	8	0	14	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	153	1	0	6	1	82.26	34.64
19	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	60	0	0	0	88.24	305
20	5	0	0	0	0	16	0	6	1	0	0	0	0	0	0	0	1	0	0	0	114	0	0	79.72	109.65
21	26	5	0	1	1	1	0	1	0	1	7	0	0	3	0	1	10	0	0	1	0	264	0	81.99	59.09
22	0	2	7	0	0	0	0	0	0	0	0	19	0	0	1	0	0	0	3	2	0	0	174	83.65	29.89
∅																								82.60	55.30

## 4.2 Entwicklung eines Viabilitätsassays

Ziel dieser Arbeiten war es, eine einfache mikroskopische Methode zur Viabilitätsanalyse von Phytoplankton, die auf dem Zell- bzw. Organismuslevel arbeitet, zu etablieren. Hierfür sollte die rote Autofluoreszenz des Chlorophylls und eine von Tang und Dobbs (2007) für verschiedene Algen beschriebene unspezifische grüne Autofluoreszenz gleichzeitig in der Zelle angeregt und detektiert werden.

Durch eine Automatisierung der Analyse sollte zum einen der Arbeitsaufwand verringert und zum anderen eine gute Reproduzierbarkeit der Ergebnisse gewährleistet werden. Hierfür wurden Fluoreszenzbilder über eine automatisierte Mikroskopieroutine aufgenommen und anschließend mit einer in ImageJ implementierten Methode automatisiert ausgewertet. Dabei wurden Zellkonzentration und Anteil an viablen und nicht viablen Zellen ermittelt.

Zur Überprüfung der Methode wurden zunächst eine tote und eine viable Kultur mit bekannter Zellzahl in unterschiedlichen Verhältnissen gemischt und mit der Fluoreszenzmethode analysiert. Anschließend wurden für einen Vergleich mit heutigen Standardmethoden der Chlorophyllgehalt, das Absorptionsspektrum und der OD<sub>750</sub>-Wert für die jeweiligen Proben ermittelt und der Anteil an viablen Zellen über Ausplattieren und Koloniezählung bestimmt.

### 4.2.1 Einsatz der Fluoreszenzmikroskopie

Grundprinzip des neu etablierten Assays war die Erkenntnis, dass eine rote Chlorophyll - Fluoreszenz nur in viablen Zellen vorhanden sein kann. Durch einen Rückgang des roten Fluoreszenzsignales während des Sterbens der Zelle wird eine grüne unspezifische Fluoreszenz sichtbar, welche bei gleicher Wellenlänge angeregt wird aber sonst vom roten Signal überstrahlt ist. Die gleichzeitige Erfassung von beiden Signalen in einem Bild sollte dabei durch den Einsatz eines Langpassfilters für die Emission ermöglicht werden. Hierfür wurden verschiedene Kombinationen von Anregungs- und Emissionsfiltern getestet (Daten nicht gezeigt). Mit der in Abbildung 4.8 dargestellte Kombination konnten grünes und rotes Signal am besten gleichzeitig abgebildet werden. Die Fluoreszenzbilder der zur Validierung verwendeten Mischproben zeigen dabei, dass dies möglich war und beide Signale in einem Bild festgehalten werden



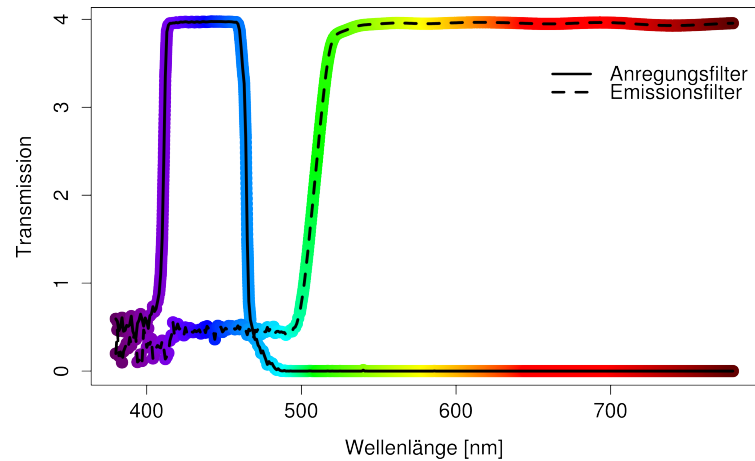


Abbildung 4.8: **Transmission des eingesetzten Filtersets für die Fluoreszenzanalyse.** Mit Hilfe des Anregungsfilters, der für Wellenlängen im Bereich von 415 - 460 nm durchlässig ist, kann sowohl die rote Chlorophyll-Fluoreszenz als auch die grüne unspezifische Fluoreszenz angeregt werden. Durch den Einsatz eines Langpassfilters, der für alle Wellenlängen nach 510 nm durchlässig ist, können rotes und grünes Emissionssignal gleichzeitig im Bild dargestellt werden. Die jeweilige Farbe der Graphen entspricht der Farbe des Lichtes bei entsprechender Wellenlänge.

konnten (Abb. 4.9). Zudem war wie erwartet erkennbar, dass mit abnehmender Anzahl an viablen Zellen der Anteil an rot fluoreszierenden Zellen abnimmt bzw. der Anteil an grün fluoreszierenden Zellen zunimmt. Die Überlagerung von Fluoreszenz- und Phasenkontrastbildern bestätigte weiterhin, dass beide Fluoreszenzsignale von *Synechocystis* ausgehen. Die Intensität des Chlorophyllsignals lag dabei deutlich über der Intensität der grünen unspezifischen Fluoreszenz. Dies hatte aber keinen weiteren Einfluss auf die Analyse.

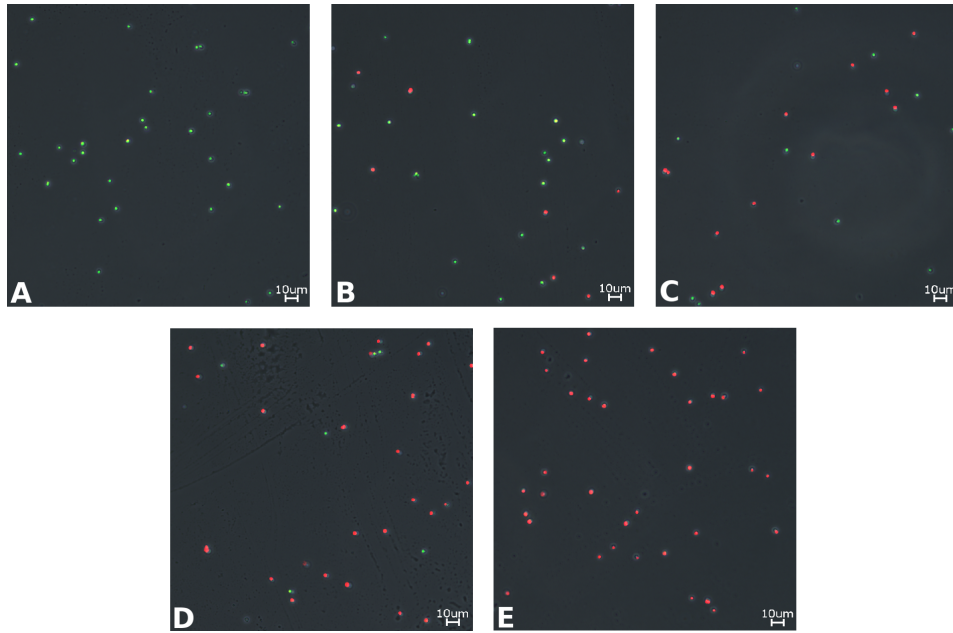


Abbildung 4.9: Mikroskopische Aufnahmen der verschiedenen Mischproben, die aus einer viablen und einer nicht viablen *Synechocystis* Kultur erstellt wurden. Der Anteil an viablen Zellen betrug A: 0%, B: 25%, C: 50%, D: 75% und E: 100%. Die Bilder zeigen die Überlagerung von Fluoreszenzaufnahme und Phasenkontrastaufnahme. Das rote Fluoreszenzsignal wird durch Anregung von Chlorophyll verursacht und kommt nur in viablen Zellen vor. Nicht viable Zellen zeigen ein grünes Fluoreszenzsignal, dessen Ursprung weitgehend unbekannt ist (Schulze et al., 2011).

#### 4.2.2 Automatisierung

Eine Automatisierung der Mikroskopie war gut möglich. Über die AutoRangePhoto Funktion des Keyence Mikroskops konnten scharfe Positionen innerhalb des Scanbereiches interpoliert werden, wodurch das Abscannen des vordefinierten Bereiches einfach möglich war. Für die Automatisierung der Bildauswertung wurde ein neues Plugin für ImageJ erstellt. Eine Übersicht der Vorgehensweise ist beispielhaft in Abbildung 4.10 dargestellt.

Aufgrund der Intensitätsunterschiede von rotem und grünen Signal musste die Segmentierung der Zellen vom Hintergrund separat erfolgen. Über die große Differenz der mittleren grünen Intensität der Zellen und den Einsatz eines einfachen Schwellwertes konnte die Unterscheidung in roten und grünen Zellen sehr präzise erfolgen (Abb. 4.9 unten).

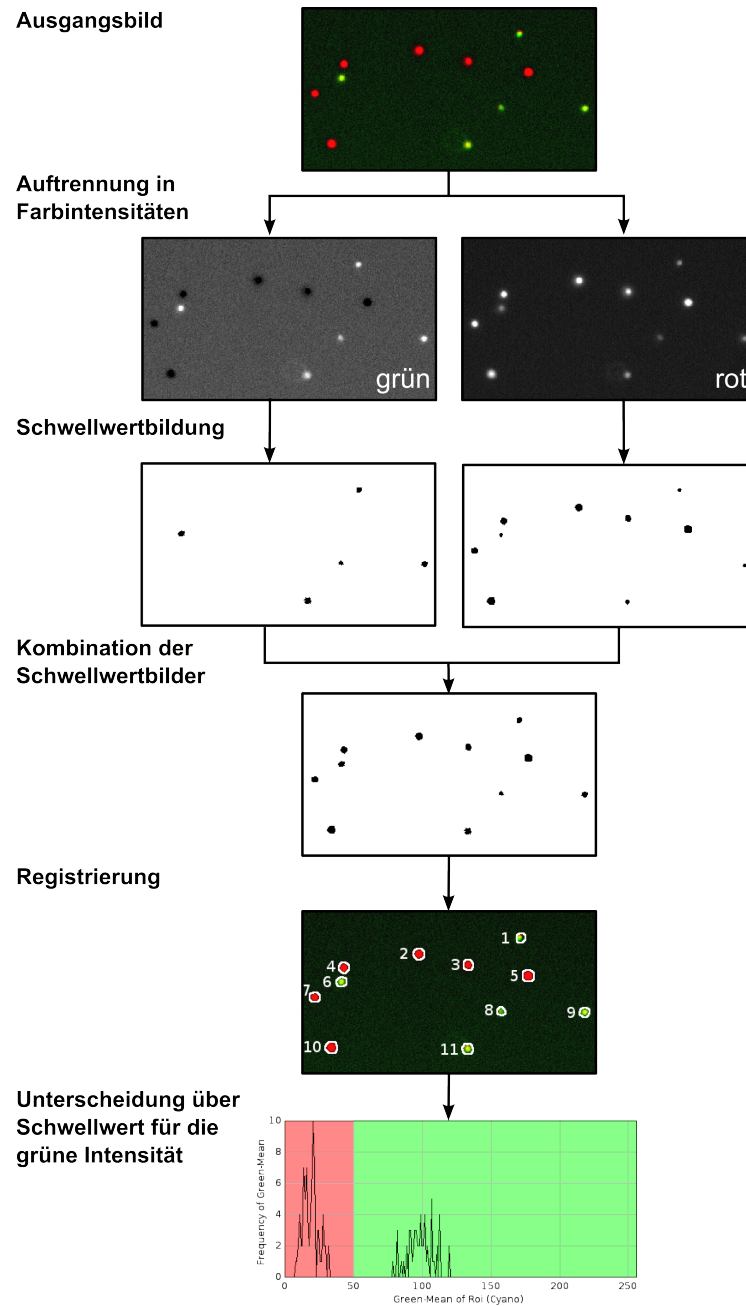
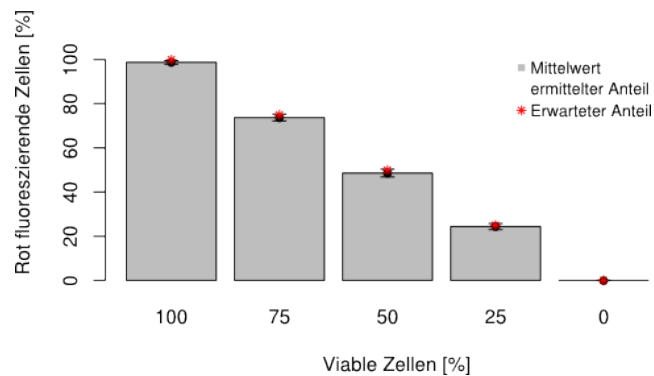


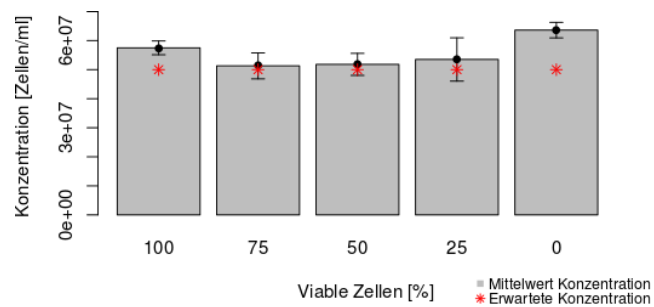
Abbildung 4.10: **Übersicht über die automatisierte Bildauswertung der Viabilitätsanalyse.** Das Ausgangsbild wurde in roten und grünen Farbkanal aufgetrennt, um die schwächeren Signale der grünen Zelle gut erfassen zu können. Für beide Kanäle wurde ein Schwellwertbild erstellt. Beide Bilder wurden wieder kombiniert, da einige Zellen in beiden Bildern erfasst wurden. Anschließend wurden alle Zellen registriert und anhand ihrer mittleren Intensität im grünen Kanal in rot und grün unterschieden. Für eine bessere Darstellung wurden Kontrast und Helligkeit der Originalbilder angepasst.

### 4.2.3 Reproduzierbarkeit

Über die automatisierte Mikroskopie und Bildauswertung war die Bestimmung der Anteile an roten (viablen) und grünen (nicht viablen) Zellen in den Mischproben mit großer Genauigkeit möglich (Abb. 4.11 a). Die berechnete Zell-Konzentration korrelierte für die Proben mit einem Anteil an 25%, 50% und 75% viablen Zellen gut mit der eingestellten Zellzahl von  $5 \cdot 10^7$  Zellen/ml. Proben, die 0% und 100% viable Zellen enthielten, zeigten einen signifikanten Unterschied. Hier lagen die Werte im Bereich von  $5.57 \cdot 10^7$  -  $6.65 \cdot 10^7$  Zellen/ml (Abb. 4.11 b).



(a)



(b)

Abbildung 4.11: **Ermittelter Anteil (a) und ermittelte Konzentration (b) an rot fluoreszierenden Zellen für die Mischproben mit unterschiedlichem Anteil an viablen Zellen.** Die Proben wurden drei mal unabhängig voneinander mikroskopiert und ausgewertet (Schulze et al., 2011).

#### **4.2.4 Validierung der Methode**

##### **Ausplattieren und Koloniezahlbestimmung**

Das Ausplattieren ist heutige Standardmethode zur Bestimmung des viablen Zellanteils einer Kultur. Hierbei wird die Kultur auf einer Agarplatte ausplattiert und nach mehrtägigem Wachstum die Anzahl an Kolonie bildenden Einheiten (KbE) bestimmt. Die verschiedenen Mischproben zeigten mit abnehmender Zahl an viablen Zellen auch eine Abnahme in der Anzahl an KbE's. Eine lineare Regression zwischen normalisierter Anzahl an gewachsenen Kolonien und dem Anteil an viablen Zellen, der mit Hilfe der Fluoreszenzmethode ermittelt wurde, ergab einen Korrelationskoeffizienten von 0.995 (Abb. 4.12).

##### **Chlorophyllgehalt**

Die Chlorophyll-Konzentration wird oft als Anhaltspunkt für die Viabilität einer phototrophen Kultur genutzt. Hier ergab sich eine Abnahme mit abnehmender Anzahl an viablen Zellen (Abb. 4.13).

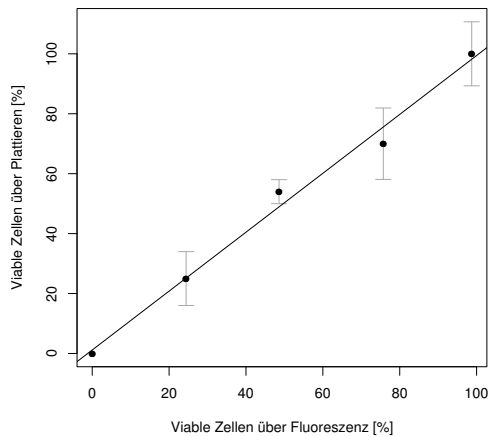


Abbildung 4.12: **Korrelation zwischen der Anzahl an viablen Zellen, die über Zellplattierung und den Fluoreszenzassay bestimmt wurden.** Die Koloniezahl und die Anzahl an fluoreszierenden Zellen wurde in Triplikaten ermittelt. Eine lineare Regression ergab für die Gerade einen Anstieg von 0,9865, einen Achsenschnittpunkt von 0,4618, und einen Korrelationskoeffizienten von 0,9956 (Schulze et al., 2011).

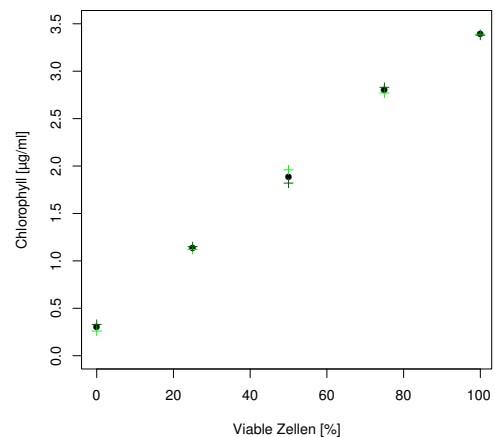


Abbildung 4.13: **Ermittelte Chlorophyll Konzentration bei steigender Anzahl an viablen Zellen in der Probe.** Die Chlorophyll Konzentration wurde für jede Probe in Duplikaten ermittelt. Es sind die Werte beider Einzelmessungen (grünes und dunkelgrünes Kreuz) und der Mittelwert beider Messungen (schwarzer Punkt) dargestellt (Schulze et al., 2011).

### Absorptionsspektrum

Absorptionsspektren dienen als schnelle und weit verbreitete Methode, um eine Übersicht über die Quantität und die Verteilung von Pigmenten in der gesamten Kultur zu erhalten. In der Probe mit 100% viablen Zellen konnten die typischen Peaks für Phycocyanin bei 630 nm, Chlorophyll a bei 450 und 680 nm und die Karotenoide im Bereich von 470-520 nm gut beobachtet werden (Abb. 4.14). Mit steigender Anzahl an nicht viablen Zellen ergab sich eine Abnahme für alle Peaks. In der Probe ohne viable Zellen konnten keine Peaks beobachtet werden, was einen Abbau aller Pigmente anzeigt.

## Optische Dichte

Die Bestimmung der optischen Dichte ist eine sehr schnelle Methode und korreliert meist gut mit der Zellzahl der Kultur. Obwohl alle Proben auf eine Zellzahl von  $5 \times 10^7$  Zellen/ml eingestellt wurden, ergab sich für den OD<sub>750</sub> Wert eine Abnahme von 0,971 auf 0,649 mit einer abnehmenden Anzahl an viablen Zellen in der Kultur (Abb. 4.15).

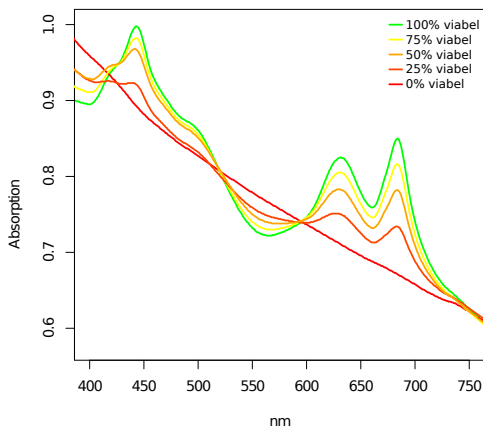


Abbildung 4.14: **Absorptionsspektren für Proben mit einem abnehmenden Anteil an viablen Zellen.** Alle Proben wurde auf eine optische Dichte von 0,65 eingestellt (Schulze et al., 2011).

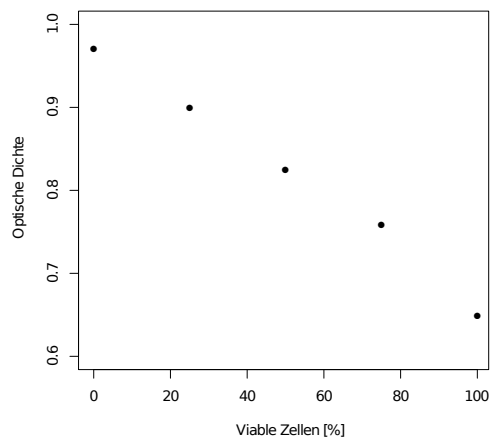


Abbildung 4.15: **Optische Dichte bei 750 nm für Proben mit einem abnehmenden Anteil an viablen Zellen.** Die Messung erfolgte in Einfachbestimmung (Schulze et al., 2011).

### 4.2.5 Anwendungsbeispiel

Die Fluoreszenzanalyse kam bei Untersuchungen zur Kulturoptimierung mittels Randonmutagenese zum Einsatz (Tillich et al., 2012). Ziel war es hier, *Synechocystis* Kulturen im Hinblick auf ihre Temperaturstabilität zu optimieren. Zellen einer Kultur wurde hierfür zunächst mutagenisiert und anschließend über eine schrittweise Erhöhung der Kultivierungstemperatur selektiert. Dabei wurde der Prozess für eine Mutageneserunde mit Hilfe des Fluoreszenzassays überwacht (Abb. 4.16). Es wurde eine Startkultur, mit einem Anteil an 90% viablen Zellen, in zwei Kulturen (Kontrolle und mutagenisierte Kultur) aufgeteilt und entsprechend behandelt. Unter den

selektiven Bedingungen konnte ein schneller Rückgang an viablen Zellen für beide Kulturen beobachtet werden. Während einer folgenden Erholungsphase mit gesenkter Temperatur ergab sich ein Anstieg an viablen Zellen in der mutagenisierten Kultur. Die Kontrollkultur verblieb hingegen bei einem Anteil nahe 0% viabler Zellen. Nach weiterer Zunahme des Anteils an viablen Zellen konnte die Temperatur in mehreren Schritten über die anfängliche maximale Kultivierungstemperatur hinaus erhöht werden, ohne dass ein Rückgang in der Viabilität der Kultur festgestellt werden konnte.

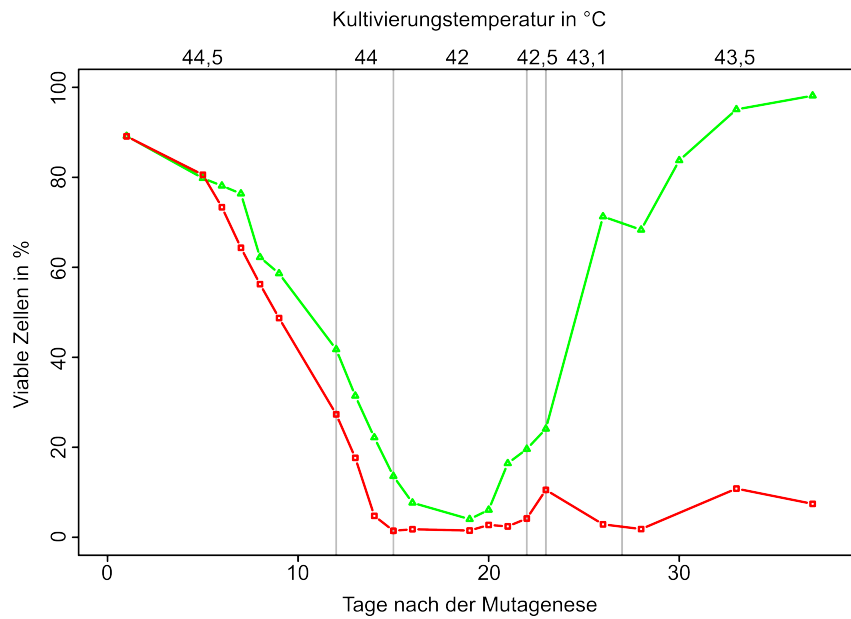


Abbildung 4.16: **Viabilitätsmonitoring einer mutagenisierten *Synechocystis* Kultur.** Es ist die zeitliche Veränderung des prozentuale Anteils an viablen Zellen bei veränderten Kultivierungstemperaturen für eine mutagenisierte (grün; Dreiecke) und eine Kontrollkultur (rot, Viereck) dargestellt (Tillich et al., 2012).



#### 4.2.6 Kombination mit Absorptionsspektren zur Vitalitätsanalyse auf Zellebene

Absorptionsspektren geben einen guten Überblick über den Pigmentgehalt einer Kultur und werden deshalb oft dazu genutzt, die Vitalität einer Kultur zu überprüfen. Allerdings können aufgrund der Herunterregulierung von Pigmenten in Stresssituationen verschiedene Zustände einer Kultur zu den gleichen Absorptionsspektren führen. Zum Beispiel kann eine Kultur, die nur gestresste aber viable Zellen enthält ein ähnliches Spektrum aufweisen, wie eine Kultur die aus einer heterogenen Mischung von nicht gestressten, viablen Zellen und nicht viablen Zellen besteht.

Aufgrund dessen sollte überprüft werden, ob mit Hilfe des Viabilitätsassays ermittelt werden kann, wie viele Zellen zur Gesamtpigmentierung der Kultur beitragen. Somit soll es möglich werden, die durchschnittliche Pigmentierung und somit die Vitalität der viablen Zellen in der Kultur zu ermitteln.

Hierfür wurde für eine temperaturadaptierte aber gestresste *Synechocystis* Kultur<sup>1</sup> die Anzahl an viablen Zellen bestimmt und ein Absorptionsspektrum aufgenommen. Diese Daten wurden anschließend mit den bereits gewonnenen Daten der optimal kultivierten Kultur verglichen. In der gestressten Kultur lag die Anzahl an viablen Zellen bei 89% und zeigte somit einen Unterschied von geringen 9% (98% viable Zellen) (siehe Abb. 4.11 a). Der Vergleich der Peakhöhen beider Absorptionsspektren zeigte deutliche Unterschiede im Gesamtpigmentgehalt (Abb. 4.17). So ergaben sich, ausgehend vom Normalisierungswert bei 750 nm ( $OD = 0.625$ ), für Chlorophyll Unterschiede von 47% (450 nm) und 40% (680 nm). Für Phycocyanin konnte ein Unterschied von 25% (630nm) beobachtet werden.

Aufgrund der relativ ähnlichen Anzahl an viablen Zellen in beiden Kulturen ist somit in der gestressten Kultur der Anteil an Pigmenten (vor allem Chlorophyll) pro Zelle deutlich herunter reguliert. Hierüber kann die Aussage getroffen werden, dass die lebendigen Zellen dieser Kultur im Durchschnitt eine geringere Vitalität als die Zellen der optimal kultivierten Kultur aufwiesen.

---

<sup>1</sup>Die Kultivierung wurde in einem Bioreaktor durchgeführt - Tag (12 Stunden): 44°C, 330  $\mu\text{E}/\text{m}^2\text{s}$ , pH-reguliert  $\text{CO}_2$  (10%) (pH = 7.3); Nacht (12 Stunden): 26°C; 10 ml/min geblubberte Luft (Tillich et al., 2012)

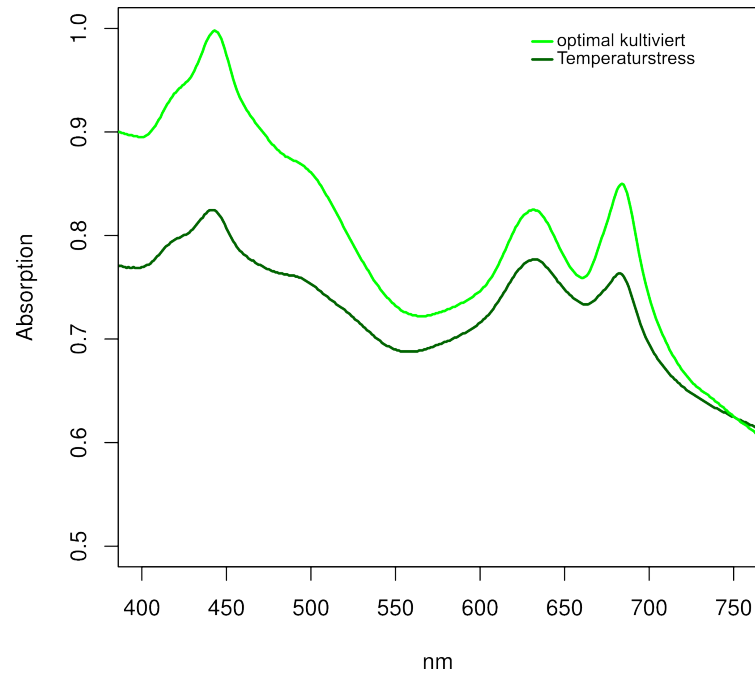


Abbildung 4.17: **Absorptionsspektrum einer unter Temperaturstress kultivierten und einer optimal kultivierten *Synechocystis* Kultur.** Die Viabilitätsanalyse zeigte für beide Kulturen einen Unterschied von 9% an viablen Zellen (98% für die optimal kultivierte Kultur und 89% für die Kultur unter Temperaturstress). Ausgehend vom Normalisierungswert bei 750 nm ergab sich für Chlorophyll Unterschiede von 47% (450 nm) und 40% (680 nm) in der Peakhöhe. Für Phycocyanin ergab sich ein Unterschied von 25% (630nm) (Schulze et al., 2011).

# 5 Diskussion

## 5.1 Automatisierte Identifikation von Süßwasser - Phytoplankton

Das in dieser Arbeit dargestellte System ermöglicht die automatisierte Identifikation von Phytoplankton und soll dabei helfen, diesen sonst sehr zeit- und arbeitsintensiven Prozess zu vereinfachen und die Wasserqualitätsanalyse zu verbessern. Dabei wurden sowohl schon beschriebene Verfahren (Rodenacker et al., 2006; Hense et al., 2008) des Systems PLASA, als auch neue Aspekte in die Methode integriert. Der Ablauf der Analyse kann dabei in zwei generelle Prozesse eingeteilt werden: Eine automatisierte Mikroskopie der Wasserprobe und die daran anschließende Bildverarbeitung und Erkennung des Phytoplanktons, die im folgenden diskutiert werden.

### 5.1.1 Mikroskopie

Für die Mikroskopie wurde die Probe in einer Utermöhl Kammer sedimentiert und über eine definierte Routine im Mikroskop abgescannt. Um eine präzise Analyse der Proben zu ermöglichen, wurden an jeder Position verschiedene Hellfeld- und Fluoreszenzbilder aufgenommen. Neben normalen Hellfeldaufnahmen erwiesen sich Quick Full Focus Bilder (QFF), die neu in die Analyse eingebunden wurden, als gut nutzbar. Diese integrieren mehrere Fokusebenen in ein Bild und haben mehrere Vorteile: Zum einen können Organismen mit einer großen Ausbreitung über die Fokusebenen auch bei hohen Vergrößerungen vollständig mikroskopiert werden, wodurch die Genauigkeit der folgenden Bildverarbeitung und Erkennung verbessert werden kann. Zum anderen kann ein Verlust von ganzen Organismusgruppen, die z. B. aufgrund ihrer Form (wie Stacheln, die sich um den Organismus herum befinden oder transparente Kapseln, in die der Organismus eingebettet ist) in anderen Fokusebenen sedimen-

tieren, besser verhindert werden. Hierdurch ergibt sich wahrscheinlich ein geringerer systematischer Fehler für das System. Zudem können Fehler des Autofokus besser ausgeglichen werden, da die besten Fokusebenen trotzdem erfasst und in das Bild integriert werden.

Da die QFF Bilder mit einer proprietären Software von Keyence aufgenommen wurden, konnte nicht genau ermittelt werden, wie die Bilder generiert wurden. Dies könnte die Adaption des Systems an andere Mikroskopieplattformen erschweren, da diese wahrscheinlich ähnliche aber nicht unbedingt genau die gleichen Funktionen zur Verfügung stellen. Um solche Probleme zu verhindern, sollte, wenn möglich, auch bei der Kontrolle von Mikroskopen auf Open Source Lösungen zurückgegriffen werden. Eine denkbare Möglichkeit wäre hier der Einsatz der Software Micro manager (Edelstein et al., 2001). Micro manager basiert auf ImageJ und ermöglicht die Aufnahme von Z-Stacks. Hier wäre die Integrierung von bereits bestehenden oder neu erstellten Plugins sehr einfach denkbar. Die derzeit unterstützte Hardware der Micro manager Software umfasst neben einzelnen Komponenten auch voll motorisierte Mikroskope. Leider wurde die in dieser Arbeit verwendete Hardware nicht unterstützt und die Software konnte somit nicht eingesetzt werden.

Weiterer Nachteil der proprietären Software war die notwendige Autofokus Funktion, die sehr langsam arbeitete und nicht variierbar in Hinblick auf Scanbereich und Abstand der Scanebenen war. Zudem wurde aufgrund eines sehr großen z-Abstandes zwischen den einzelnen analysierten Positionen häufiger nicht die optimale Fokusebene ausgewählt. Allerdings konnte dies durch den Einsatz der QFF-Funktion gut ausgeglichen werden.

Die Aufnahme der Fluoreszenzsignale der verschiedenen Phytoplankter ermöglichte eine bessere Abbildung der unterschiedlich vorhandenen Photo-Pigmente in den einzelnen Taxa. Dabei konnte ausgenutzt werden, dass verschiedene Pigmente bei unterschiedlichen Anregungswellenlängen Autofluoreszenzsignale zeigen. Hier wurde neben den in bereits bestehenden Systemen verwendeten Signalen von Chlorophyll und Phycoerythrin zusätzlich die Fluoreszenz von Phycocyanin aufgenommen, die ein zusätzliches Merkmal zur Unterscheidung liefert.

Wie von Hense et al. (2008) beschrieben, konnte bei einer Anregung mit handels-

üblichen Chlorophyllfiltern, die nur Chlorophyll b anregen, nicht für alle Plankter ein Signal aufgenommen werden. Dies ist darauf zurückzuführen, dass Chlorophyll b im Photosystem vieler Phytoplankton Spezies nicht vorhanden ist (Vergleich Tabelle 1.3). Durch eine Anpassung des Filtersets wurden Chlorophyll a, b und c angeregt. Dadurch war es möglich, Signale für alle Phytoplankter aufzunehmen und somit für die Unterscheidung zu nicht Planktern zu nutzen. Weitere Überlegung war es, die aufgenommene Signalintensität für eine Differenzierung der Chlorophyll - Pigmentierung nutzen zu können. Zwar war es möglich Unterschiede zwischen verschiedenen Organismusgruppen festzustellen, allerdings machen weitere Einflussfaktoren wie Größe und Verteilung der Pigmentierung innerhalb des Organismus, sowie große Schwankungen innerhalb einer Klasse die Differenzierung nicht trivial.

Für Phycoerythrin und Phycocyanin konnte in den Gruppen der Cyanobakterien (*Anabaena*, *Limnothrix*, *Merismopedia*, *Aphanizomenon*) und Cryptophyta (*Cryptomonas*) Signale aufgenommen werden, die zudem innerhalb der Gruppen unterschiedlich ausgeprägt waren. Somit konnten hierüber wertvolle Informationen für eine Unterscheidung gewonnen werden.

Weiterhin ergaben sich großen Schwankungen in den Fluoreszenzintensitäten einer Art. Diese sind vorrangig darauf zurückzuführen, dass für die Fluoreszenz nur eine Fokusebene aufgenommen wurde und somit nicht immer der gleiche Fokuspunkt abgebildet werden konnte. Zudem wurden die Kulturen bei Tageslicht gehalten, was eine Anpassung der Photosysteme auf sich verändernde Lichtbedingungen und teilweise auch eine unterschiedliche Vitalität der Kultur bewirkt. Die Aufnahme von QFF Bildern war hier nicht möglich, da dies zu großen Bleicheffekten führten.

### **5.1.2 Bildverarbeitung und Bilderkennung**

#### **Segmentierung**

Für die Segmentierung wurde ein Region Growing Ansatz verwendet. Vorteil ist hier, dass nur Informationen des Hintergrundes in die Berechnung mit einbezogen werden (Adams und Bischof, 1994). Im Vergleich zu Histogramm-basierten Verfahren, bei denen alle im Bild enthaltenen Informationen die Segmentierung beeinflussen, ermöglicht dieser Ansatz eine robustere Segmentierung, wenn, wie in diesem Fall, sehr unterschiedliche Partikel erfasst werden müssen. Zur guten Erfassung von Organis-

men mit transparenten Anteilen wurde eine Kantenerkennung (für die Helligkeit) in die Segmentierung integriert. Hier konnten allerdings nur scharf abgebildete Organismen gut abgetrennt werden, da Kanten nur an steilen Intensitätsunterschieden gut ermittelt werden können. Dies stellt allerdings keinen Nachteil dar, da falsch fokussierte Organismen durch ihre ungenaue Abbildung zu einer reduzierten Genauigkeit bzw. verfälschten Merkmalsberechnung führen und somit die Gesamtanalyse negativ beeinflusst hätten. Somit scheint die zusätzliche Verwendung von Kanteninformationen im Region Growing Verfahren für diese Anwendung besser geeignet zu sein.

### **Merkmalsberechnung und Auswahl**

Für die Charakterisierung der einzelnen Partikel wurde verschiedene Merkmale berechnet, die Form, Textur, Farbe, Helligkeit und Fluoreszenz der Partikel auf unterschiedliche Weise beschreiben. Ziel war es aus der Gesamtzahl an zur Verfügung stehenden Merkmalen eine Auswahl zu ermitteln, die im aufgebauten System gut für die Unterscheidung von Phytoplanktern genutzt werden kann.

Hierzu wurden mit dem Trainingsset, bestehend aus zehn verschiedenen Taxa, zunächst ein Training mit allen verfügbaren Merkmalen durchgeführt und anschließend die Signifikanz der dem Merkmal entsprechenden Neuronen bestimmt (Abb. 4.6). Anhand dieser Orientierungshilfe sollte ein geeignetes Merkmalsset ausgewählt und über eine Klassifikation des Testsets überprüft werden. Hierbei erwiesen sich verschiedene Merkmale zur Beschreibung der Farbe, Local Binary Patterns und verschiedene Fluoreszenzmerkmale als gut geeignet.

Werte für die Helligkeit und Sättigung zeigten zunächst gute Signifikanzen, führten aber bei der Testklassifikation zu keinen guten Ergebnissen. Gleiches gilt für verschiedene Statistikwerte der Grauwertmatrix, die jedoch schon beim initialen Training geringe Bedeutung zeigten. Diese Merkmale sind dabei nicht invariant zur Beleuchtung des Bildes. Hier kann ein Altern der verwendeten Lampe oder eine variierende Trübungen des Probenwassers mögliche Unterschiede hervorrufen. Diese konnten trotz konstant gehaltener Beleuchtungsdauern und durchgeführter Ausleuchtungskorrektur nicht vermieden werden.

Weiterhin war auffällig, dass vor allem Merkmale zur Beschreibung der Form der Par-

tikel nur geringe Signifikanzen zeigten. Dies war sowohl im initialen Training mit allen Merkmalen, als auch im ausgewählten Merkmalsset zu beobachten. Um gute Klassifikationsergebnisse zu erzielen, war trotzdem die Integration von mehreren Merkmalen zur Formbeschreibung erforderlich. Mögliche Ursache hierfür könnten die häufig sehr großen Formschwankungen für die einzelnen Organismen einer Spezies sein. So konnte z. B. die Fourier Deskriptoren 1 - 12 gut für die Klassifikation genutzt werden. Bei Verwendung von den Deskriptoren 13 - 29 kam es eher zu Verschlechterungen der Klassifikationsleistung. Da die Form mit zunehmender Anzahl an Deskriptoren immer genauer repräsentiert wird, wurden die auftretenden Schwankungen zu genau beschrieben und verhinderten so eine exakte Unterscheidung. Somit lieferte die Auswahl von Merkmalen über die Neuronensignifikanz nur einen groben Überblick und konnte die Auswahl von Hand nicht ersetzen.

Einige Merkmale wurden trotz guter Signifikanz nicht mit der Testklassifikation überprüft, da bereits gute Klassifikationsergebnisse erzielt wurden und bei Erweiterung des Merkmalssets nicht ersichtlich war, ob diese eine Verbesserung ermöglichen. Diese wurden bei der Erweiterung des Testsets überprüft.

In weiteren Versuchen wurde die ReliefF - Methode zur automatischen Merkmalsauswahl von Weka verwendet, die gut bleibende Unterscheidungsraten ermöglichte. Gerade im Hinblick auf eine Erweiterung der Daten, bei der die Merkmalsauswahl immer komplexer wird, stellt dies Vorgehensweise eine bessere Alternative dar.

## **Klassifikation**

Die Klassifikation des initialen Testsets mit zehn Organismen wurde zunächst mit Hilfe von neuronalen Netzen durchgeführt. Zur Unterscheidung in die einzelnen Gruppen wurde ein Netz mit zwölf verschiedenen Ausgabemöglichkeiten (zehn für die jeweiligen Taxa, eine für Detritus und eine für restliche unbekannte Partikel) gewählt und mit einem ausgewählten Set an Merkmalen trainiert. Hier wäre es auch denkbar gewesen, für jede Gruppe ein Netz mit speziell optimiertem Merkmalsset zu trainieren und in der Klassifikation zu prüfen, ob ein Partikel mit dem jeweiligen Netz zugeordnet werden kann oder nicht. Allerdings wären Training und Merkmalsselektion deutlich komplexer und nach Erweiterung des Systems um weitere Gruppen, müsste jedes Netz neu geprüft und wenn nötig adaptiert werden. Da die Klassifikation mit

einem Netz gute Ergebnisse zeigte, wurde dieser Ansatz aufgrund seiner Einfachheit und guten Erweiterbarkeit bevorzugt.

Trotz der sehr guten durchschnittlichen Klassifikationsrate von ca. 95% kann ein Übertrainieren des Netzes ausgeschlossen werden, da das verwendete Testset Bilder von unterschiedlichen Proben und Aufnahmezeitpunkten enthielt. Zudem wurde eine relativ gleichmäßige Erkennungsrate ohne bestimmten Bias zu einer Klasse beobachtet.

Der Vergleich der Klassifikationsergebnisse mit weiteren veröffentlichten Systemen ist schwer möglich, da hier sowohl Unterschiede in der Klassenanzahl als auch in der morphologischen Ähnlichkeit der zu unterscheidenden Organismen bestehen. Zusätzlich wurden die meisten beschriebenen Systeme für eine Unterscheidung von marinem Plankton verwendet. Dennoch wurden ein Vergleich mehrere Systeme von Culverhouse (2007) zusammengestellt. Dabei sind die unterschiedlichen Klassifikationsergebnisse wie folgt dargestellt:

- Das System ADIAC (37 Taxa, 75–90% Erkennungsrate) zur automatisierten Analyse von Diatomeenpräparaten (Du Buf und Bayer, 2002)
- Die Plattform Zooscan (29 Gruppen, 75–85% Erkennungsrate), bei der mit Hilfe von wasserdichten Scannern erzeugte Bilder von Mesozooplankton analysiert werden (Grosjean et al., 2004; Gorsky et al., 2010)
- Die Systeme SIPPER (5 Gruppen, 75–90% Erkennungsrate) und VPR (7 Gruppen, 72% Erkennungsrate), wobei es sich um Systeme zur Analyse von Mesozooplankton handelt und Bilder der Organismen während der Probennahme aufgenommen werden (Remsen et al., 2004; Q. Hu und Davis, 2005)
- DiCANN (3–23 Spezies, 70–87% Erkennungsrate) zur Klassifikation von Dinoflagellaten (Culverhouse et al., 1996)
- Das System Cytosense (30 Gruppen, 91% Erkennungsrate), bei dem ein Flow Cytometer zum Einsatz kam (Dubelaar et al., 2007)

Zudem gibt es das System PLASA, welches als einziges für die Analyse von Phytoplankton entwickelt wurde und hierfür mikroskopische Aufnahmen nutzt (Rodenacker et al., 2006). Die durchschnittliche Erkennungsrate von PLASA (ca. 94% bei



5 Klassen) liegt dabei über allen Erkennungsraten der vorher genannten Systeme, jedoch noch unter der Erkennungsrate des hier vorgestellten Systems (ca. 95% für 10 Klassen). Großer Nachteil von PLASA ist, dass die Software in IDL entwickelt wurde und somit nicht frei zur Verfügung steht.

Weiterhin muss angemerkt werden, dass die drei besten Systeme ADIAC, Cytosense und SIPPER wahrscheinlich nur schlecht für eine Unterscheidung von Phytoplankton genutzt werden könnten. Das System SIPPER wurde für die Erkennung von Mesozooplankton entwickelt und kann somit nicht den Größenbereich von Phytoplankton erfassen, welches sich hauptsächlich aus Mikro- und Nanoplankton im Größenbereich von 2 -200  $\mu\text{m}$  zusammensetzt. Der flow cytometrische Ansatz Cytosense hat den Nachteil, dass Partikel nur auf Grundlage von Charakteristika, die mit dem Photomultiplier Rohr eines Flowcytometers erfasst werden (vorrangig Lichtstreuung und Fluoreszenz) unterschieden werden müssen. Hierdurch wird nur eine Unterscheidung in Gruppen ohne größere taxonomische Auflösung möglich. Das ADIAC System wurde für die Analyse von präparierten Diatomeenslides entwickelt. Da diese kaum Ähnlichkeiten zu Phytoplanktonproben aufweisen scheint ein Einsatz hierfür auch ungeeignet.

Weiterhin war es möglich, die Falsch-Positiv Rate (ca. 6%) im Vergleich zu publizierten Ergebnissen der anderen Systeme (PLASA 20%; VPR 55% (Q. Hu und Davis, 2005), Zooscan 26% (Gorsky et al., 2010)) deutlich zu verringern. Dabei wurden Fehler meist durch Falschklassifikation von dem System unbekanntem Organismen oder anderen Partikeln wie Detritus verursacht. Im dargestellten System konnte diese Fehlerquelle wahrscheinlich durch den Einsatz des optimierten Chlorophyll-Fluoreszenzfilter deutlich reduziert werden. Hierdurch konnten vor allem nicht-planktonische Partikel sehr gut von der Analyse ausgeschlossen werden (Fehlerrate  $<1\%$ ). Zudem konnten unbekannte Partikel (hier handelte es sich vor allem um Partikel die falsch abgebildet oder segmentiert wurden, Organismen die nicht die typischen morphologischen Merkmale aufwiesen und Organismen die mit anderen Partikel überlagert bzw. aggregiert waren) mit einer Genauigkeit von ca. 92% identifiziert werden, obwohl diese große Unterschiede in Form, Farbe und Fluoreszenzeigenschaften zeigten. Dies ist wahrscheinlich auf den Einsatz von neuronalen Netzen zurückzuführen, da diese besonders gut für die Modellierung von komplexen Frage-

stellungen geeignet sind.

Es konnten sowohl Taxa die innerhalb ihrer Gruppe eine sehr ähnliche Morphologie zeigten (z.B. *Cryptomonas* und *Trachelomonas* mit einer runden und stabil bleibenden Form) als auch Taxa, die deutliche morphologische Unterschiede innerhalb ihrer Gruppe aufwiesen (z.B. haben *Botryococcus* und *Anabeana* eine grundlegende Zellstruktur, bei der sich allerdings die Gesamtstruktur durch die Anzahl an Zellen und deren Anordnung während des Wachstums verändert) gut voneinander unterschieden werden.

Beim Vergleich der durchschnittlichen Klassifikationsrate des dargestellten Systems von ca. 95% mit menschlicher Genauigkeit zeigt sich eine etwas bessere Fehlerrate als für Personal das Routineauswertungen durchführt (84-95% Genauigkeit). Für geschulte aber nicht im Routineeinsatz tätige Personen konnte eine erkennbar bessere Rate festgestellt werden (67-83% Genauigkeit für eine Person und 43% Übereinstimmung zwischen verschiedenen Personen) (Culverhouse et al., 2003). Weiterer Vorteil der automatischen Lösungen ist dabei auch, dass die Probenanalyse durch den Menschen nur schwer reproduzierbare Fehlerraten ergibt.

### **Vergleich mit weiteren Klassifikationsverfahren**

Der Vergleich mit weiteren Klassifikationsverfahren zeigte, dass die guten Ergebnisse auch auf den Einsatz eines neuronalen Netzes zurückzuführen sind. Dies ist damit zu erklären, dass es durch die innere Vernetzung möglich ist, komplexe nichtlineare Beziehungen zwischen verschiedenen Merkmalen zu detektieren. Durch die Fähigkeit komplexe nichtlineare Funktionen auch modellieren zu können, können diese detektierten Beziehungen auch für eine Unterscheidung genutzt werden (Tu, 1996). Hierdurch sind sie vor allem bei komplexen bzw. realistischen Fragestellungen, wie auch dem hier bearbeitete Problem, anderen Methoden überlegen. Allerdings muss auch angemerkt werden, dass die Verfahrensweise eines trainierten Netzes nur schwer nachvollzogen werden kann. Hierdurch wird es z.B. schwer zu erkennen, ob mit dem Training lokale Minima oder optimale Ergebnisse erzielt wurden. So kann das gleiche Merkmalsset bei gleicher Fehlerrate im Training zu unterschiedlichen Klassifikationsergebnissen führen und die Auswahl von geeigneten Parametern erschweren und weitere Trainingsdurchgänge notwendig machen.

Als einziges weiteres Verfahren zeigte die Support Vektor Maschine eine ähnlich gute Erkennungsrate. Hierbei handelt es sich um ein Klassifikationsverfahren, welches über den Einsatz von verschiedenen Kernelfunktionen die Unterscheidung von komplexen Daten über jegliche nichtlineare Funktionen ermöglicht. Somit sind sie sehr gut, ähnlich wie neuronale Netze, für den Einsatz bei komplexen Fragestellungen geeignet. Alle weiteren Klassifikationsverfahren zeigten im Vergleich deutlich schlechtere Erkennungs- und Falsch-Positiv Raten und scheinen für diese Problemstellung weniger geeignet zu sein. Aufgrund dieser Ergebnisse wurde bei der Erweiterung des Test- und Trainingsset, neben den neuronalen Netzen, auch auf Support Vektor Maschinen als mögliche Klassifikationsmethode zurückgegriffen.

### 5.1.3 Vergrößerung des Testsets und Integration weiterer Merkmale

Bei der Erweiterung der Daten wurde darauf geachtet Planktonarten in die Analyse zu integrieren, die in unterschiedlichen Charakteristika Ähnlichkeiten zu anderen Arten aufweisen oder während ihres Wachstums unterschiedliche Strukturen ausbilden, um hier mögliche Schwachpunkte des Systems besser erkennen zu können. Neben den jeweilig sehr ähnlichen Pigmentierungen und Fluoreszenzcharakteristika in den einzelnen Organismusgruppen (Chlorophyceae, Bacillariophyta und Cyanobakterien) wiesen innerhalb der Gruppen einzelne Plankter weitere Ähnlichkeiten auf. So zeigen zum Beispiel die Cyanobakterien *Anabaena*, *Aphanizomenon* und *Limnothrix* eine ähnliche längliche Form, die aus kleineren Segmenten aufgebaut wird. *Anabaena* und *Aphanizomenon* wiesen zusätzlich noch eine ähnliche Pigmentierung auf. Bei den Grünalgen zeigen *Monoraphidium contortum*, *Monoraphidium minutum* und *Ankistrodesmus* eine längliche und gekrümmte Form, wobei Größe und Krümmungsgrad variieren. Vor allem bei den Arten *Coelastrum*, *Actinastrum*, *Pandorina*, *Botryococcus*, *Merismopedia*, *Fragilaria* und *Anabaena* war eine Variation der Morphologie in verschiedenen Wachstumsphasen zu beobachten, da diese alle aus verschiedenen Grundstrukturen während des Wachstums aufgebaut werden. Zudem traten bei *Anabaena* und *Merismopedia* in einzelnen Phasen Ähnlichkeiten auf, da beide kugelige Grundstrukturen aufweisen. Dieser Effekt konnte auch bei *Pandorina* und *Botryococcus* beobachtet werden, wobei hier die jeweilige Grundstruktur kreisförmig um den Ursprung herum erweitert werden. Weiterhin sind in frühen Wachstumsstadien von

*Actinastrum* und *Fragilaria* längliche Strukturen vorhanden, die jeweils *Ankistrodesmus* bzw. *Cyclotella* ähneln können.

Trotz dieser Erweiterungen zeigten die Ergebnisse, dass auch hier die Unterscheidung der einzelner Plankter zwar leicht verringert aber immer noch gut möglich war. Dabei ist die geringe Klassifikationsleistung vorrangig auf die Unterscheidung zu unbekanntem bzw. nicht erkennbaren Partikeln zurückzuführen. Hier muss berücksichtigt werden, dass die Daten mit Hilfe von Laborkulturen gewonnen wurden. Dabei bilden nicht immer alle Mitglieder einer Population die charakteristischen morphologischen Merkmale aus und es können zudem morphologische Veränderungen über den Kultivierungszeitraum auftreten. Dieses Problem trat vorrangig für die neu in die Analyse integrierten Arten auf. Somit kam es hier zu einer deutlich größeren Diversität und Anzahl an Partikeln, die in diese Datenklasse unterschieden werden mussten. Im Hinblick auf die Analyse von echten Gewässerproben sollte dieses Problem geringer sein, da der Großteil der Plankter morphologisch wichtige Merkmale ausbildet. Hier stellen eher die Verunreinigung mit nicht planktonischen Partikeln ein Problem dar, die jedoch mit dem vorgestellten System auch bei Erweiterung der Daten noch gut abgetrennt werden konnten.

Im Vergleich mit den bestehenden Systemen, die bereits im Abschnitt **Klassifikation** aufgeführt wurden, wurde eine gleich gute Erkennungsrate erzielt. Die Falsch Positiv Rate, die vor allem durch unbekannte Partikel beeinflusst wird, war weiterhin verbessert.

Bei den bereits erwähnten Arten, die größere Ähnlichkeiten aufwiesen oder verschiedene Wachstumsstadien durchlaufen, konnte nur für *Anabaena* und *Aphanizomenon* eine erhöhte Fehleranfälligkeit festgestellt werden. Somit scheint es mit dem System auch gut möglich zu sein, Klassen mit geringen Unterschieden voneinander abzutrennen. Zudem können auch Veränderungen in der Morphologie in einer Klasse gut erfasst werden.

Weitere Häufungen von Fehlklassifikationen konnten jeweils zwischen den Klassen von *Monoraphidium minutum* und *Desmodesmus*, *Ankistrodesmus* und *Fragilaria* sowie *Botryococcus* und *Trachelomonas* festgestellt werden. Hier können die Fehler allerdings nicht auf grundsätzliche Ähnlichkeiten zurückgeführt werden. Da die Unterscheidung zwischen diesen Klassen für den Großteil der Testdaten möglich war,

ist es wahrscheinlich, dass die jeweilige Klassen des Trainingssets noch nicht alle möglichen phänotypischen Stadien der Organismen erfassten und hier noch erweitert werden müssen.

Weiterhin besteht auch die Möglichkeit, dass durch die vollautomatische Merkmalsauswahl eine noch nicht ganz optimale Selektion stattfand. Hier wäre für das neuronale Netz als alternative Möglichkeit zum Beispiel die Nutzung des NEAT (NeuroEvolution of Augmenting Topologies) Verfahrens denkbar (Stanley und Miikkulainen, 2002). Besonders an dieser Methode ist, dass die Netzstruktur nicht vorgegeben werden muss, sondern während des Trainings aufgebaut wird. Somit können z.B. alle Merkmale dem Netz zur Verfügung gestellt werden und nur dann mit dem Netz verknüpft werden, wenn diese von Nutzen sind (Whiteson et al., 2005).

Die Ergebnisse zeigen weiterhin, dass eine Vorhersage, wie gut einzelne Organismen mit Hilfe von maschinellen Methoden unterschieden werden können, nur schwer möglich ist. Hier bestehen zu große Unterschiede zwischen der menschlichen und maschinellen Verarbeitung des Problems.

Der Vergleich zwischen den beiden hier eingesetzten Klassifikationsverfahren zeigt, dass durch die Erweiterung der Daten bei der Support Vektor Maschine vor allem bei der Abtrennung von unbekanntem Bestandteilen größere Probleme auftraten und dies vorrangig zu einer deutlich erhöhten Falsch Positiv Rate führte. Somit ist das neuronale Netz deutlich flexibler, um vor allem sehr diverse Klassen abzugrenzen und somit für die hier behandelte Fragestellung besser geeignet.

## 5.2 Entwicklung eines Viabilitätsassays

Bei der Analyse von Autofluoreszenzsignalen unterschiedlicher Planktonarten und Mikroalgen konnte von Tang und Dobbs (2007) und Hense et al. (2008) eine unspezifische grüne Fluoreszenz beobachtet werden. Diese trat vorrangig dann auf, wenn Autofluoreszenzsignale des Photosystems nicht mehr vorhanden waren. Überlegung dieser Arbeiten war es, den Effekt für eine einfache und schnelle Viabilitätsanalyse zu nutzen. Einsatzmöglichkeit ist zum einen die sehr zeitnahe Überwachung von Laborkulturen, aber auch die Nutzung als neuer Parameter in der Analyse von Phytoplanktongemeinschaften ist naheliegend und könnte hier genauere Einblicke ermöglichen.

Für höhere Pflanzen und Algen ist bekannt, dass ein Gleichgewicht zwischen Aufbau und Degradation des D1- Proteins des Photosystems, in dem das Chlorophyll lokalisiert ist, mit hohen Umsatzraten von ca. 1 Stunde besteht. Bei Seneszenz oder Sterben wird dieses Gleichgewicht stark in Richtung Abbau des D1-Proteins verschoben (Sundby et al., 1993; Yao et al., 2012; Beck und Scheibe, 2003; Hörtensteiner, 2006). Somit kommt es zu einem Rückgang der roten Chlorophyllfluoreszenz, die als Marker für viable Zellen dient. Gleichzeitig wird die unspezifische grüne Fluoreszenz sichtbar und kann als Marker für nicht viable Zellen genutzt werden.

Für die Etablierung der Methode wurde das Cyanobakterium *Synechocystis* sp. PCC6803 ausgewählt. Hierbei handelt es sich um einen einzelligen und weit genutzten Modellorganismus für phototrophe Organismen. *Synechocystis* ist offiziell als Süßwasserstamm eingeordnet, ist aber sehr tolerant gegenüber Salz und marinem Medium (Marin et al., 2002). Das Genom von *Synechocystis* wurde erstmals im Jahre 1996, als erstes Genom eines photosynthetisch aktiven Organismus sequenziert (Kaneko et al., 1996).

Die Filterauswahl für den Assay orientierte sich an der Kombination von Hense et al. (2008). Mit dieser konnte jedoch nur Chlorophyll b angeregt werden und war somit für Arten, die dieses Pigment nicht enthalten unbrauchbar. Als Modifikation wurde der Anregungsbereich so erweitert, dass die Anregung von Chlorophyll a, b und c möglich wurde. Da alle photosynthetisch aktiven Organismen Chlorophyll a und/oder b bzw. c enthalten, ist ein Einsatz für andere Arten durch die Anpassung

der Filter somit nicht mehr begrenzt. Gerade im Hinblick auf eine Integration der Methode in die Analyse von Phytoplanktongemeinschaften stellt dies eine wichtige Weiterentwicklung dar.

Im Gegenteil zu Methoden mit langen Vorbereitungszeiten oder Wachstumsphasen, handelt es sich bei der neu etablierten Methode um eine sehr einfache und schnelle Möglichkeit eine unmittelbare Analyse auf Zellebene durchzuführen. Mit Hilfe der Automatisierung ergaben sich bei der Bestimmung an viablen und nicht viablen Zellen gut reproduzierbare Ergebnisse und es konnte somit eine präzise Analyse sichergestellt werden. Im Vergleich zu Färbemethoden wird eine genauere Analyse möglich, da Kontaminationen (vorrangig heterotrophe Bakterien) aus verschiedenen Kulturen bei der genutzten Anregungswellenlänge keine Fluoreszenzsignale zeigten und somit nicht berücksichtigt wurden (Daten nicht gezeigt).

Zur Überprüfung der Methode wurden die erzielten Ergebnisse mit der Koloniezahl nach Ausplattieren, der Chlorophyll-Konzentration und gewonnenen Absorptionsspektren verglichen. Hierbei handelt es sich um gängige Methoden, die als Standardmethoden für die Kulturüberwachung eingesetzt werden. Die starke Korrelation zwischen den Ergebnissen des Fluoreszenzassays und der Kolonieanzahl nach Ausplattieren der Kultur bestätigt die Beobachtungen und zeigt, dass eine genaue Analyse der viablen Zellzahl möglich wird. Auch Absorptionsspektrum und Chlorophyllkonzentration stimmen mit den Ergebnissen des Assays überein.

Diese Ergebnisse lassen zunächst vermuten, dass sowohl Absorptionsspektren als auch Chlorophyllkonzentration dafür genutzt werden können, den Anteil an viablen Zellen zu bestimmen. Hier muss jedoch beachtet werden, dass diese Werte immer nur einen Durchschnitt aller Zellen der Kultur darstellen. Hier kann nicht unterschieden werden, ob ein Teil der Zellen tot ist oder ob alle Zellen ihren Pigmentgehalt reduziert haben. Mögliche Ursachen hierfür sind z.B. Stress durch hohe Lichtintensitäten oder aber das Sterben von Teilen der Kultur durch Alterungsprozesse. Somit sind Absorptionsspektren und die Chlorophyllkonzentration nicht für die Bestimmung der Anzahl an viablen Zellen einsetzbar. Dies kann nur Methoden auf Zellebene ermöglicht werden.

Wie bereits erwähnt, geben Absorptionsspektren und die Chlorophyllkonzentration einen guten quantitativen Überblick über den Gesamtgehalt an Pigmenten in der

Kultur. Eine Kombination mit der neu entwickelten binären Methode scheint dabei von Vorteil, um bessere Rückschlüsse auf die Vitalität von Kulturen treffen zu können<sup>1</sup>. Der Vergleich der Absorptionsspektren einer temperaturgestressten und einer nicht gestressten Kultur zeigte dabei eine deutlich höhere Pigmentierung für die nicht gestresste Kultur, obwohl in der Anzahl an viablen Zellen ein Unterschied von nur 9% festgestellt werden konnte. Dies lässt vermuten, dass die gestresste Kultur einen deutlich geringeren Pigmentgehalt pro viabler Zelle aufwies und somit die Zellen eine geringere Vitalität besaßen. Die Kombination von neu entwickelter Methode und Absorptionsspektren scheint es somit möglich zu machen, die durchschnittliche Pigmentierung der viablen Zellen in der Kultur besser bestimmen zu können, da nun ermittelt werden kann, wie viele Zellen zum Signal des Absorptionsspektrums beitragen.

Zusätzlich wurden die Ergebnisse mit OD<sub>750</sub> Werten der Proben verglichen. Die optische Dichte einer Probe wird oft dazu verwendet, die Zellkonzentration einer Kultur zu bestimmen. Obwohl alle Proben auf die gleiche Zellkonzentration eingestellt wurden, kam es bei sinkendem Anteil an viablen Zellen zu einer Abnahme der optischen Dichte. Für die Kultur, die nur nicht viable Zellen enthielt, wurde eine etwa 30% geringere Zellkonzentration als tatsächlich erwartet bestimmt. Wahrscheinliche Ursache hierfür ist der Verlust der Pigmentierung, der eine Veränderung der Absorption und Lichtstreuung in der Kultur bewirkt. Für genauere Analysen sollte somit auf eine Zellzählung zurückgegriffen werden.

Die automatisierte Bestimmung der Zellkonzentration zeigte, dass die ermittelten Werte aller Proben gut mit den erwarteten Werten korrelierten.

Der Einsatz der Methode zur Kulturüberwachung bei Mutageneseversuchen ermöglichte eine gute Überwachung und Dokumentation des Prozesses. Sie stellte eine optimale Anpassung der Kultivierungsbedingungen an den Zustand der Kultur sicher und gewährleistete somit auch den Erfolg der Optimierung. Im Verlauf des Versuches konnte über die gewonnenen Daten gut verfolgt werden, wie es in der beginnenden Phase der Selektion sowohl in der mutagenisierten Kultur als auch in der Kontrollkultur zu einer starken Abnahme an viablen Zellen kam. Dabei konnte für die mutageni-

---

<sup>1</sup>In dieser Arbeit wird die Vitalität einer Kultur/Zelle als Pigmentgehalt pro Kultur/Zelle definiert, da Cyanobakterien ausbleichen, wenn sie unter Stressbedingungen kultiviert werden.



sierte Kultur eine verringerte und langsamere Sterberate beobachtet werde. Dies war insofern erstaunlich, als dass die verwendete Mutagen dosis eine hohe Letalitätsrate von 90% für *Synechocystis* aufweist (Tillich et al., 2012). Zudem konnte sich die geringe Anzahl an noch viablen Zellen in der mutierten Kultur (4% an Tag 19) im Vergleich zur Kontrollkultur gut erholen. Dies zeigt, dass wahrscheinlich ein nur geringer Teil der Zellen einen selektiven Vorteil gewonnen hatte, sich anschließend aber stark vermehren konnte.

## 5.3 Fazit und Ausblick

Im ersten Teil der Arbeit wurde durch die Nutzung von neuen Verfahren in der Mikroskopie und der Bilderkennung ein System aufgebaut, das die automatische Erkennung unterschiedlichster Phytoplankter ermöglicht. Im Bereich der Mikroskopie wurde durch die Aufnahme von dreidimensionalen Informationen und den Einsatz von neuen und verbesserten Fluoreszenzfiltern eine konsistente und gute Abbildung unterschiedlichster Phytoplanktern möglich. Dies ist gerade im Hinblick auf die sehr variable Morphologie des Phytoplanktons entscheidend.

In der Bilderkennung wurde durch den Gebrauch von neuronalen Netzen und dem Einsatz einer Vielzahl und teilweise zuvor noch nicht für diese Aufgabenstellung genutzten Charakteristika eine sehr hohe Erkennungsrate für das System ermöglicht. Im Vergleich zu bestehenden Verfahren erlaubte die sehr gute Erkennung von Detritus und unbekanntem Partikeln deutlich verbesserte Falsch-Positiv-Raten und machte die Analyse somit deutlich spezifischer.

In Verlauf der Arbeiten konnte die Anzahl an zu unterscheidenden Klassen deutlich erweitert werden und es konnte noch keine Limitierung auf eine bestimmte Anzahl an zu unterscheidbaren Klassen festgestellt werden. Eine weitere Erhöhung der Klassenanzahl mit den hier vorgestellten Methoden scheint somit machbar und ein Einsatz für die Analyse von echten Gewässerproben ist gut denkbar, wurde jedoch im Rahmen dieser Arbeiten noch nicht überprüft.

Anzumerken ist hier, dass das einmalige Training eines Klassifikators, der für eine Unterscheidung aller definierter Indikatortaxa Deutschlands (mehrere 100 Arten) genutzt werden könnte, einen sehr hohen Rechenaufwand benötigt. So wurde für das Training für die Unterscheidung von zehn Organismen ca. 30 min mit einem Desktop-PC (Intel Xeon W3520; 4 x 2.67GHz; 12 GB DDR3 RAM) benötigt. Für die Unterscheidung der 22 Organismen wurde aufgrund von Strukturanpassungen des Netzes und der Erweiterung des Merkmalssets bereits mehrere Stunden benötigt. Da zudem viele Arten unterschiedliche Stadien in ihrem Wachstum durchlaufen, steigt mit weiterer steigender Klassenzahl die Komplexität des Problems deutlich stärker. Hier sollte in Zukunft für das einmal nötige Training des Klassifikators auf einen rechenstarken Server zurückgegriffen werden. Die eigentliche Klassifikation,

die in allen Analysen durchgeführt wird, ist dabei größtenteils unabhängig von der Klassenzahl und kann auch weiterhin auf Desktop-PC's durchgeführt werden.

Im Hinblick auf die Routineanalyse scheint zudem, gerade zur Verringerung der Komplexität und zum sicherstellen einer optimalen Erkennung, eine an die regional vorherrschenden Bedingungen angepasste Auswahl an Taxa sinnvoll. Hier ist auch anzumerken, dass die Definition von geeigneten Indikatoren ein noch nicht abgeschlossener Prozess ist. Zum Beispiel zeigte die Analyse von ausgewählten Seen Europas, die fast ideale Referenzbedingungen aufwiesen, dass nur 29 Arten echte Indikatoren für die vorherrschenden Bedingungen darstellten. Weiterhin wurde festgestellt, dass eine regional angepasste Auswahl von Taxa, die nicht als Indikator definiert sind, ein verlässlicher Parameter sein könnte (Järvinen et al., 2013). Somit wäre es zum Beispiel denkbar einen Katalog von Trainingsdaten zu erstellen und diesen dann nach entsprechender Auswahl für das Training eines angepassten Klassifikators zu nutzen. Naheliegend wäre hier auch eine mehrstufige Unterscheidung, bei der die Daten zunächst in größere Gruppen vorsortiert werden.

Eine weitere Möglichkeit wäre auch die Kombination mit anderen Techniken, wie z.B. Flowsortern, die eine grobe Vorsortierung der Probe in definierte Klassen ermöglichen (Dubelaar und Jonker, 2000; Erickson et al., 2012). Anschließend könnte auch hier eine genauere Unterscheidung der vorsortierten Proben über an die Gruppen angepasste Klassifikatoren erfolgen.

Da die neuronalen Netze im entwickelten System einfach mit neuen Daten trainiert werden können, ist die Integration von neuen Taxa in die Analyse gut möglich. Nach Einrichtung und Adaption des Systems könnte so die sonst sehr zeitaufwendige Routineanalyse, für die Spezialisten erforderlich sind, automatisiert werden und auch von technischem Personal übernommen werden. Somit könnten die heute schon nur in begrenzter Zahl zur Verfügung stehenden Spezialisten in ihrer Arbeit entlasten und für andere Aufgaben genutzt werden.

Im Hinblick auf die europäische Rahmengesetzgebung ist bei Phytoplanktonuntersuchungen im Bereich des Gewässermonitorings weiterhin mit einem steigenden Bedarf zu rechnen. Dieser könnte durch die mit dem entwickelten System besser gedeckt werden. Dabei kann der Einsatz des Systems sowohl im Bereich der Auftragsanalyse als auch an staatlichen Stellen, Universitäten und Forschungseinrichtungen erfolgen.

Hierbei ist die Software leicht und kostenfrei zugänglich, da sie als Open Source Lösung entwickelt wurde.

Durch den möglichen Einsatz an unterschiedlichen Stellen ist auch eine Verbesserung im Hinblick auf die Qualitätssicherung denkbar. Durch die Nutzung gleicher Verfahren (in den Arbeiten eingesetzte Mikroskopie und Bildauswertung) würden Ergebnisse für unterschiedliche Proben und Standorte besser vergleichbar. Dabei ist die Software (insbesondere der trainierte Klassifikator) leicht zugänglich und könnte zwischen den einzelnen Stellen einfach ausgetauscht werden.

Im zweiten Abschnitt der Arbeit wurde die Etablierung eines neuen Verfahrens dargestellt, welches für die schnelle und akkurate Quantifizierung von viablen und nicht viablen *Synechocystis* Zellen genutzt werden kann. Im Gegensatz zu bereits bestehenden Methoden, ermöglicht es ohne weitere Probenvorbereitung einen sofortigen Einblick auf Zellebene und hat praktisch keine laufenden Kosten. Durch eine Kombination mit Absorptionsspektren wird zudem ein schneller Einblick in die Vitalität von Zellen möglich. Der Einsatz bei Mutageneseversuchen zeigte, dass präzise Einschätzungen über den Zustand der Kultur getroffen werden konnten und erlaubte eine optimale Überwachung des Prozesses.

Neben der Überwachung von Laborkulturen ist es auch gut denkbar, die Methode in das System zur Planktonerkennung zu integrieren, um auch hier genauere Einblicke in den Zustand der einzelnen Phytoplankter zu ermöglichen. Als neuer Parameter in der Wasserqualitätsanalyse könnte dies zu besseren Einblicken in das Populationsverhalten von Phytoplankton verhelfen. Dabei wurde die Methode im Rahmen der Arbeit noch nicht für weitere Arten überprüft. Allerdings wurde die grüne Fluoreszenz für unterschiedlichste Phytoplankter (Cyanobakterien, Grünalgen, Diatomeen, Dinoflagellaten und Raphidophyceae) beschrieben (Tang und Dobbs, 2007) und konnte auch während der eigenen Arbeiten nicht nur in Cyanobakterien, sondern auch bei verschiedenen Grün- und Kieselalgen beobachtet werden. Somit scheint auch für andere Klassen des Phytoplanktons ein Einsatz gut möglich, müssten jedoch mit ähnlichen Validierungsversuchen überprüft werden.

Während der Versuche konnten im Verlauf des Absterbens der Kultur auch orange

fluoreszierende Zellen in den Proben beobachtet werden. Dieses Signal ergab sich als Mischfarbe aus rotem und grünen Fluoreszenzsignal und zeigt einen Abbau der Photopigmente an. Hier ist es denkbar, den Farbverlauf zwischen beiden Signalen in Zukunft auch als Vitalitätsfaktor auf Zellebene nutzen zu können.

Eine einfache und umfangreich anwendbare Erweiterung der Methodik wäre auch hier die Adaption an ein Flow Cytometer mit Sorter. Ohne die bis jetzt notwendigen Färbemethoden, welche zum Sterben der Zellen führen, könnten so viable Zellen einfach aus Proben isoliert werden. Dies kann beispielsweise eine effizientere Optimierung von Kulturen ermöglichen. Eine weitere hilfreiche Anwendung kann der Einsatz zur Abtrennung von Kontaminationen sein, da diese in Kultivierungsprozessen häufig Probleme hervorrufen. Zudem könnte das Verfahren hierdurch auch zum einfachen vereinzeln von nur viablen Zellen einer Kultur genutzt werden.

Eine zukünftiger Einsatz der Bildverarbeitung ist auch direkt im Feld denkbar. Hier wurden bereits verschiedene Systeme vorgestellt, bei denen durch geschickte Linsenauswahl, Bilder mit Auflösungen von bis zu 400x über ein Mobiltelefon aufgenommen werden konnten (Smith et al., 2011; Arpa et al., 2012). Auch die Aufnahme von Fluoreszenzbildern ist bereits durch eine miniaturisierte Anordnung von Filtern und LEDs in mobilen Systemen möglich (Wei et al., 2013). Hier ist es gut denkbar, die dargestellten Methoden der Bilderkennung in einer mobilen Anwendung zu überführen. Um Auswertungsergebnisse direkt vor Ort erhalten zu können, wären Cloudanwendungen zwischen dem mobilen Gerät zur Aufnahme und einem zentralen Server zur Bildauswertung gut einsetzbar.

Obwohl in diesen Arbeiten Methoden zur Charakterisierung und Unterscheidung verschiedener Phytoplankter erstellt wurde, ist der Einsatz nicht auf diese beschränkt, da die eingesetzte Vorgehensweisen universell anwendbar sind. Zudem können die integrierten Klassifikationsverfahren mit selbst erstellten Trainingsdaten leicht trainiert werden. Somit sind prinzipiell alle Anwendungen denkbar, bei denen bestimmte Objekte in (mikroskopischen) Bildern erkannt werden müssen. Gerade in Kombination mit mobilen Anwendungen könnten sich hier neue Anwendungsfelder entwickeln. Denkbare Einsatzgebiete wäre z.B. die Qualitätssicherung für die Lebensmittelpro-

duktion, die Unterscheidung verschiedener Pollen zur Warnung von Allergikern oder aber auch die medizinische Diagnostik vor Ort. Hiermit könnten vor allem logistische Probleme verringert und Analyse öfter und zeitnaher erfolgen.

# Zusammenfassung

Zentrales Ziel dieser Arbeit war es, Methoden der Mikroskopie, Bildverarbeitung und Bilderkennung für die Charakterisierungen verschiedener Phytoplankter zu nutzen, um deren Analyse zu verbessern und zu vereinfachen.

Der erste Schwerpunkt der Arbeit lag auf der Analyse von Phytoplanktongemeinschaften, die im Rahmen der Überprüfung der Süßwasserqualität als Marker dienen. Die konventionelle Analyse ist dabei sehr aufwendig, da diese noch immer vollständig von Hand durchgeführt wird und hierfür speziell ausgebildetes Personal eingesetzt werden muss. Ziel war es, ein System zur automatischen Erkennung aufzubauen, um die Analyse vereinfachen zu können. Mit Hilfe von automatischer Mikroskopie war es möglich Plankter unterschiedlicher Ausdehnung durch die Integration mehrerer Schärfestufen besser in einem Bild aufzunehmen. Weiterhin wurden verschiedene Fluoreszenzeigenschaften in die Analyse integriert.

Mit einem für ImageJ erstellten Plugin können Organismen vom Hintergrund der Aufnahmen abgetrennt und eine Vielzahl von Merkmalen berechnet werden. Über das Training von neuronalen Netzen wird die Unterscheidung von verschiedenen Gruppen von Planktontaxa möglich. Zudem können weitere Taxa einfach in die Analyse integriert und die Erkennung erweitert werden.

Die erste Analyse von Mischproben, bestehend aus 10 verschiedenen Taxa, zeigte dabei eine durchschnittliche Erkennungsrate von 94.7% und eine durchschnittliche Falsch-Positiv Rate von 5.5%. Im Vergleich mit bestehenden Systemen konnte die Erkennungsrate verbessert und die Falsch Positiv Rate deutlich gesenkt werden. Bei einer Erweiterung des Datensatzes auf 22 Taxa wurde darauf geachtet, Arten zu verwenden, die verschiedene Stadien in ihrem Wachstum durchlaufen oder höhere Ähnlichkeiten zu den bereits vorhandenen Arten aufweisen, um evtl. Schwachstellen des Systemes erkennen zu können. Hier ergab sich eine gute Erkennungsrate (86.8%), bei der der Ausschluss von nicht-planktonischen Partikeln (11.9%) weiterhin verbes-

sert war.

Der Vergleich mit weiteren Klassifikationsverfahren zeigte, dass neuronale Netze anderen Verfahren bei dieser Problemstellung überlegen sind. Ähnlich gute Klassifikationsraten konnten durch Support Vektor Maschinen erzielt werden. Allerdings waren diese bei der Unterscheidung von unbekanntem Partikeln dem neuronalen Netz deutlich unterlegen.

Der zweite Abschnitt stellt die Entwicklung einer einfachen Methode zur Viabilitätsanalyse von Cyanobakterien, bei der keine weitere Behandlung der Proben notwendig ist, dar. Dabei wird die rote Chlorophyll - Autofluoreszenz als Marker für lebende Zellen und eine grüne unspezifische Fluoreszenz als Marker für tote Zellen genutzt. Der Assay wurde mit dem Modellorganismus *Synechocystis* sp. PCC 6803 etabliert und validiert. Die Auswahl eines geeigneten Filtersets ermöglicht es beide Signale gleichzeitig anzuregen und zu beobachten und somit direkt zwischen lebenden und toten Zellen zu unterscheiden. Die Ergebnisse zur Etablierung des Assays konnten durch Ausplattieren, Chlorophyllbestimmung und Bestimmung des Absorptionsspektrums bestätigt werden. Durch den Einsatz von automatisierter Mikroskopie und einem neu erstellten ImageJ Plugin wurde eine sehr genaue und schnelle Analyse der Proben möglich. Der Einsatz beim Monitoring einer mutagenisierten Kultur zur Erhöhung der Temperaturtoleranz ermöglichte genaue und zeitnahe Einblicke in den Zustand der Kultur. Weitere Ergebnisse weisen darauf hin, dass die Kombination mit Absorptionsspektren es ermöglichen können bessere Einblicke in die Vitalität der Kultur zu erhalten.



# Abstract

Central goal of this work was to improve and simplify the characterization of different phytoplankton by the use of automated microscopy, image processing and image analysis.

The first part of the work dealt with the analysis of phytoplankton communities, which are used as a marker for the determination of fresh water quality. The current routine analysis, is very time consuming and expensive, as it is carried out manually by trained personnel. Thus the goal of this work was to develop a system for automating the analysis. With the use of automated microscopy different focal planes could be integrated into one image, which made it possible to image plankton of different focus levels simultaneously. Additionally it allowed the integration of different fluorescence characteristics into the analysis. An image processing routine, developed in ImageJ, allows the segmentation of organisms from the image background and the calculation of a large range of features. Neural networks are then used for the classification of previously defined groups of plankton taxa. The program allows easy integration of additional taxa and expansion of the recognition targets.

The analysis of samples containing 10 different taxa showed an average recognition rate of 94.7% and an average error rate of 5.5%. The obtained recognition rate was better than those of existing systems and the exclusion of non-plankton particles could be greatly improved. After extending the data set to 22 different classes of (more demanding) taxa a still good recognition (86.9 %) and still improved error rate (11.9 %) were obtained. This extended set was specifically selected in order to target potential weaknesses of the system. It contained mainly taxa that showed strong similarities to each other or taxa that go through various different morphological stages during their growth. The obtained recognition rates were comparable or better than those of existing systems and the exclusion of non-plankton particles could be greatly improved.

A comparison of different classification methods showed, that neural networks are superior to all other investigated methods when used for this specific task. While similar recognition rates could be achieved with the use of support vector machines they were vastly inferior for the differentiation of unknown particles.

The second part focused on the development of a simple live - dead assay for unicellular cyanobacteria without the need of sample preparation. The assay uses red chlorophyll fluorescence, corresponding to viable cells, and an unspecific green autofluorescence, that can only be observed in non viable cells.

The assay was established and validated for the model organism *Synechocystis* sp. PCC 6803. With the selection of a suitable filter-set both signals could be excited and observed simultaneously, allowing a direct classification of viable and non-viable cells. The results were confirmed by plating/colony count, absorption spectra and chlorophyll measurements. The use of an automated fluorescence microscope and an ImageJ based image analysis plugin allows a very precise and fast analysis. The monitoring of a random mutagenized culture undergoing selection for improved temperature tolerance allowed an accurate and prompt insight into the condition of the culture. Further results indicate that a combination of the new assay with absorption spectra or chlorophyll concentration measurements allows the estimation of the vitality of cells.

# Publikationen

## Primärpublikationen

**Schulze K**, Tillich UM, Dandekar T and Frohme M (2013). PlanktoVision - an automated analysis system for the identification of phytoplankton, BMC Bioinformatics 2013, 14:115

Tillich UM, Lehmann S, **Schulze K**, Dühring U, Frohme M (2012) The Optimal Mutagen Dosage to Induce Point-Mutations in Synechocystis sp. PCC6803 and Its Application to Promote Temperature Tolerance, PLoS ONE 7(11): e49467.

**Schulze K**, López DA, Tillich UM und Frohme M (2011). A Simple Live / Dead Analysis for unicellular Cyanobacteria using a new Autofluorescence Assay, automated Microscopy, and ImageJ, BMC Biotechnology 2011, 11:118 (30 November 2011)

## Konferenzbeiträge

**Schulze K**, López DA, Tillich UM and Frohme M (2013). A simple viability analysis for unicellular cyanobacteria. 8. Deutsches Biosensor Symposium, Wildau (*Poster*)

**Schulze K**, López DA, Tillich UM, Gatomski G, Krüger P, Foitzik A, Mischke U, Frohme M (2012). Characterization of microalgae and phytoplankton samples. BioHyTec Workshop: Bioanalytik und Biosystemtechnik, Wildau (*Talk*)

# Literatur

- Adams, R. und L. Bischof (1994). “Seeded region growing”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.6, S. 641–647.
- Aha, D. W., D. Kibler und M. K. Albert (1991). “Instance-based learning algorithms”. *Machine learning* 6.1, S. 37–66.
- Ahonen, T., J. Matas, C. He und M. Pietikäinen (2009). “Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features”. In: *Image Analysis*. Hrsg. von A.-B. Salberg, J. Y. Hardeberg und R. Jenssen. Lecture Notes in Computer Science 5575. Springer Berlin Heidelberg, S. 61–70.
- Altınok, A., E. Kiris, A. J. Peck, S. C. Feinstein, L. Wilson, B. S. Manjunath und K. Rose (2007). “Model based dynamics analysis in live cell microtubule images”. *BMC Cell Biology* 8 (Suppl 1), S4.
- Arpa, A., G. Wetzstein, D. Lanman und R. Raskar (2012). “Single lens off-chip cellphone microscopy”. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), S. 23–28.
- Beck, E. und R. Scheibe (2003). “Senescence and ageing in plants and cyanobacteria”. *Physiologia Plantarum* 119.1, S. 1–4.
- Benfield, M., P. Grosjean, P. Culverhouse, X. Irigolen, M. Sieracki, A. Lopez-Urrutia, H. Dam, Q. Hu, C. Davis, A. Hanson, C. Pilskaln, E. Riseman, H. Schulz, P. Utgoff und G. Gorsky (2007). “RAPID: Research on Automated Plankton Identification”. *Oceanography* 20.2, S. 172–187.
- Beutler, M., K. H. Wiltshire, B. Meyer, C. Moldaenke, C. Lüring, M. Meyerhöfer, U.-P. Hansen und H. Dau (2002). “A fluorometric method for the differentiation of algal populations in vivo and in situ”. *Photosynthesis Research* 72.1, S. 39–53.

- Bouckaert, R. R. (2004). *Bayesian network classifiers in weka*. Department of Computer Science, University of Waikato.
- Boudier, T. und B. Tupper (2009). *ImageJ - Elliptic Fourier Descriptor Plugin*. URL: [http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:fourier\\_shape\\_analysis:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:fourier_shape_analysis:start).
- Breiman, L. (2001). “Random Forests”. *Machine Learning* 45.1, S. 5–32.
- Cabrera, J. E. (2006). *ImageJ - Texture Analyzer Plugin*. URL: <http://rsbweb.nih.gov/ij/plugins/texture.html>.
- Carpenter, A. E., T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland und D. M. Sabatini (2006). “CellProfiler: image analysis software for identifying and quantifying cell phenotypes”. *Genome Biology* 7.10, R100.
- Carson, J. P., T. Ju, H.-C. Lu, C. Thaller, M. Xu, S. L. Pallas, M. C. Crair, J. Warren, W. Chiu und G. Eichele (2005). “A Digital Atlas to Characterize the Mouse Brain Transcriptome”. *PLoS Comput Biol* 1.4, e41.
- Carvalho, L., S. Poikane, A. L. Solheim, G. Phillips, G. Borics, J. Catalan, C. De Hoyos, S. Drakare, B. J. Dudley, M. Järvinen, C. Laplace-Treyture, K. Maileht, C. McDonald, U. Mischke, J. Moe, G. Morabito, P. Nöges, T. Nöges, I. Ott, A. Pasztaleniec, B. Skjelbred und S. J. Thackeray (2013). “Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes”. *Hydrobiologia* 704.1, S. 127–140.
- Cellamare, M., A. Rolland und S. Jacquet (2010). “Flow cytometry sorting of freshwater phytoplankton”. *Journal of applied phycology* 22.1, S. 87–100.
- Cessie, S. L. und J. C. V. Houwelingen (1992). “Ridge Estimators in Logistic Regression”. *Applied Statistics* 41.1, S. 191.
- Chang, C.-C. und C.-J. Lin (2011). “LIBSVM: a library for support vector machines”. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.3, S. 27.
- Chatzichristofis, S. A. und Y. S. Boutalis (2008a). “CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval”. In: *Computer Vision Systems*. Hrsg. von A. Gasteratos, M. Vincze und J. K. Tsotsos. Lecture Notes in Computer Science 5008. Springer Berlin Heidelberg, S. 312–322.

- (2008b). “Fcth: Fuzzy color and texture histogram—a low level feature for accurate image retrieval”. In: *Ninth International Workshop on Image Analysis for Multimedia Interactive Services, 2008. WIAMIS’08*. IEEE, S. 191–196.
- Cohen, W. W. (1995). “Fast Effective Rule Induction”. In: *Twelfth International Conference on Machine Learning*. Bd. 95, S. 115–123.
- Culverhouse, P. F. (2007). “Human and machine factors in algae monitoring performance”. *Ecological Informatics* 2.4, S. 361–366.
- Culverhouse, P. F., R. G. Simpson, R. Ellis, J. A. Lindley, R. Williams, T. Parisini, B. Reguera, I. Bravo, R. Zoppoli und G. Earnshaw (1996). “Automatic classification of field-collected dinoflagellates by artificial neural network”. *Marine ecology progress series* 139.1, S. 281–287.
- Culverhouse, P. F., R. Williams, B. Reguera, V. Herry und S. González-Gil (2003). “Do experts make mistakes? A comparison of human and machine identification of dinoflagellates”. *Marine Ecology Progress Series* 247, S. 17–25.
- Du Buf, H. und M. M. Bayer (2002). *Automatic diatom identification*. Bd. 51. World Scientific.
- Dubelaar, G. B. J. und R. R. Jonker (2000). “Flow cytometry as a tool for the study of phytoplankton”. *Scientia Marina* 64.2, S. 135–156.
- Dubelaar, G. B. J., R. Casotti, G. A. Tarran und I. C. Biegala (2007). “Phytoplankton and their Analysis by Flow Cytometry”. In: *Flow Cytometry with Plant Cells*. Hrsg. von J. Doležal, J. Greilhuber und J. Suda. Wiley-VCH Verlag GmbH & Co. KGaA, S. 287–322.
- Edelstein, A., N. Amodaj, K. Hoover, R. Vale und N. Stuurman (2001). “Computer Control of Microscopes Using  $\mu$ Manager”. In: *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.
- Elliott, D. L. (1993). “A Better Activation Function for Artificial Neural Networks”. *ISR Technical Report*.
- Erickson, J. S., N. Hashemi, J. M. Sullivan, A. D. Weidemann und F. S. Ligler (2012). “In situ phytoplankton analysis: there’s plenty of room at the bottom”. *Analytical chemistry* 84.2, S. 839–850.
- Euopäische Union (2000). “Richtlinie 2000/60/EG des europäischen Parlaments und des Rates vom 23. Oktober 2000 zur Schaffung eines Ordnungsrahmens für Maß-

- nahmen der Gemeinschaft im Bereich der Wasserpolitik". *Amtsblatt Nr. L 327*, S. 0001–0073.
- Fan, R.-E., K.-W. Chang, C.-J. Hsieh, X.-R. Wang und C.-J. Lin (2008). "LIBLINEAR: A library for large linear classification". *The Journal of Machine Learning Research* 9, S. 1871–1874.
- Fischer, S., M. Binkert und H. Bunke (2000). "Symmetry based indexing of diatoms in an image database". In: *15th International Conference on Pattern Recognition, 2000. Proceedings*. 15th International Conference on Pattern Recognition, 2000. Proceedings. Bd. 2, 895–898 vol.2.
- Flusser, J. (2000). "On the independence of rotation moment invariants". *Pattern Recognition* 33.9, S. 1405–1410.
- Flusser, J. und T. Suk (2006). "Rotation moment invariants for recognition of symmetric objects". *IEEE Transactions on Image Processing* 15.12, S. 3784–3790.
- Gama, J. (2004). "Functional trees". *Machine Learning* 55.3, S. 219–250.
- Gorsky, G., M. D. Ohman, M. Picheral, S. Gasparini, L. Stemmann, J.-B. Romagnan, A. Cawood, S. Pesant, C. García-Comas und F. Prejger (2010). "Digital zooplankton image analysis using the ZooScan integrated system". *Journal of Plankton Research* 32.3, S. 285–303.
- Graham, L., J. Graham und L. W. Wilcox (2009). *Algae*. 2. Aufl. Pearson Education Inc, San Francisco, California.
- Grosjean, P., M. Picheral, C. Warembourg und G. Gorsky (2004). "Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system". *ICES Journal of Marine Science: Journal du Conseil* 61.4, S. 518–525.
- Hall, M. und G. Holmes (2003). "Benchmarking attribute selection techniques for discrete class data mining". *IEEE Transactions on Knowledge and Data Engineering* 15.6, S. 1437–1447.
- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann und I. H. Witten (2009). "The WEKA Data Mining Software: An Update". *ACM SIGKDD Explorations Newsletter* 11.1, S. 10–18.

- Haralick, R., K. Shanmugam und I. Dinstein (1973). "Textural Features for Image Classification". *IEEE Transactions on Systems, Man and Cybernetics* SMC-3.6, S. 610–621.
- Heaton, J. (2010). *Programming neural networks with Encog 2 in Java*. Heaton Research Inc.
- Hense, B. A., P. Gais, U. Jütting, H. Scherb und K. Rodenacker (2008). "Use of fluorescence information for automated phytoplankton investigation by image analysis". *Journal of Plankton Research* 30.5, S. 587–606.
- Hensen, V. (1887). *Über die Bestimmung des Planktons oder des im Meere treibenden Materials an Pflanzen und Tieren*. Schmidt & Klaunig.
- Hoehn, E. (1998). *Erfassung und Bewertung von Planktonorganismen*. Oldenbourg.
- Holmes, G., B. Pfahringer, R. Kirkby, E. Frank und M. Hall (2002). "Multiclass alternating decision trees". In: *Machine Learning: ECML 2002*. Springer, S. 161–172.
- Hörtensteiner, S. (2006). "Chlorophyll Degradation During Senescence\*". *Annual Review of Plant Biology* 57.1, S. 55–77.
- Hsu, C.-W. und C.-J. Lin (2002). "A comparison of methods for multiclass support vector machines". *IEEE Transactions on Neural Networks* 13.2, S. 415–425.
- Hu, M.-K. (1962). "Visual pattern recognition by moment invariants". *IRE Transactions on Information Theory* 8.2, S. 179–187.
- Hu, Q. und C. Davis (2005). "Automatic plankton image recognition with co - occurrence matrices and support vector machine". *Marine Ecology Progress Series* 295, S. 21–31.
- Hu, X., R. Su, F. Zhang, X. Wang, H. Wang und Z. Zheng (2010). "Multiple excitation wavelength fluorescence emission spectra technique for discrimination of phytoplankton". *Journal of Ocean University of China* 9.1, S. 16–24.
- Huang, J., S. R. Kumar, M. Mitra, W.-J. Zhu und R. Zabih (1997). "Image indexing using color correlograms". In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE, S. 762–768.
- Hühn, J. und E. Hüllermeier (2009). "FURIA: an algorithm for unordered fuzzy rule induction". *Data Mining and Knowledge Discovery* 19.3, S. 293–319.



- Igathinathane, C., L. O. Pordesimo, E. P. Columbus, W. D. Batchelor und S. R. Methuku (2008). “Shape identification and particles size distribution from basic shape parameters using ImageJ”. *Computers and Electronics in Agriculture* 63.2, S. 168–182.
- Igel, C. und M. Hüsken (2000). “Improving the Rprop learning algorithm”. In: *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*, S. 115–121.
- Järvinen, M., S. Drakare, G. Free, A. Lyche-Solheim, G. Phillips, B. Skjelbred, U. Mischke, I. Ott, S. Poikane, M. Søndergaard, A. Pasztaleniec, J. V. Wichelen und R. Portielje (2013). “Phytoplankton indicator taxa for reference conditions in Northern and Central European lowland lakes”. *Hydrobiologia* 704.1, S. 97–113.
- John, G. H. und P. Langley (1995). “Estimating continuous distributions in Bayesian classifiers”. In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., S. 338–345.
- Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda und S. Tabata (1996). “Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions”. *DNA Research* 3.3, S. 109–136.
- Kapur, J. N., P. K. Sahoo und A. K. C. Wong (1985). “A new method for gray-level picture thresholding using the entropy of the histogram”. *Computer Vision, Graphics, and Image Processing* 29.3, S. 273–285.
- Keerthi, S. S., S. K. Shevade, C. Bhattacharyya und K. R. K. Murthy (2001). “Improvements to Platt’s SMO algorithm for SVM classifier design”. *Neural Computation* 13.3, S. 637–649.
- Keller, A., M. Wolf und T. Dandekar (2010). “Ribosomal RNA phylogenetics: the third dimension”. *Biologia* 65.3, S. 388–391.
- Kuhl, F. P. und C. R. Giardina (1982). “Elliptic Fourier features of a closed contour”. *Computer Graphics and Image Processing* 18.3, S. 236–258.

- Lalli, C. und T. R. Parsons (1997). *Biological Oceanography: An Introduction*. Butterworth - Heinemann.
- Landwehr, N., M. Hall und E. Frank (2005). "Logistic model trees". *Machine Learning* 59.1, S. 161–205.
- Lee, T. J., K. Nakano und M. Matsumura (2000). "A new method for the rapid evaluation of gas vacuoles regeneration and viability of cyanobacteria by flow cytometry". *Biotechnology Letters* 22.23, S. 1833–1838.
- Loman, N. J., C. Constantinidou, J. Z. Chan, M. Halachev, M. Sergeant, C. W. Penn, E. R. Robinson und M. J. Pallen (2012). "High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity". *Nature Reviews Microbiology* 10.9, S. 599–606.
- Long, F., H. Peng, X. Liu, S. K. Kim und E. Myers (2009). "A 3D digital atlas of *C. elegans* and its application to single-cell analyses". *Nature Methods* 6.9, S. 667–672.
- Longair, M. H., D. A. Baker und J. D. Armstrong (2011). "Simple Neurite Tracer: open source software for reconstruction, visualization and analysis of neuronal processes". *Bioinformatics* 27.17, S. 2453–2454.
- Lux, M. und S. A. Chatzichristofis (2008). "Lire: lucene image retrieval: an extensible java cbir library". In: *Proceedings of the 16th ACM international conference on Multimedia*. ACM, S. 1085–1088.
- Maileht, K., T. Nöges, P. Nöges, I. Ott, U. Mischke, L. Carvalho und B. Dudley (2013). "Water colour, phosphorus and alkalinity are the major determinants of the dominant phytoplankton species in European lakes". *Hydrobiologia* 704.1, S. 115–126.
- Marin, K., J. Huckauf, S. Fulda und M. Hagemann (2002). "Salt-Dependent Expression of Glucosylglycerol-Phosphate Synthase, Involved in Osmolyte Synthesis in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803". *Journal of Bacteriology* 184.11, S. 2870–2877.
- Metzker, M. L. (2010). "Sequencing technologies — the next generation". *Nature Reviews Genetics* 11.1, S. 31–46.
- Mischke, U. (2006). "Bundesweiter Praxistest eines Bewertungsverfahrens für Phytoplankton in Fließgewässern Deutschlands zur Umsetzung der EU - Wasser-

- rahmenrichtlinie - Verfahrensvereinfachung und -überprüfung mit Handbuchtentwurf". *LAWA-Projekt O 3*.
- Mischke, U., H. Behrendt, J. Köhler und D. Opitz (2005). *Entwicklung eines Bewertungsverfahrens für Fließgewässer mittels Phytoplankton zur Umsetzung der EU-Wasserrahmenrichtlinie*. Leibniz-Institut für Gewässerökologie und Binnenfischerei (IGB).
- Mischke, U. und W. H. Kusber (2006). "Die harmonisierte Taxaliste des Phytoplanktons für Seen und Flüsse in Deutschland. Excel-Datei-Liste mit ausführlichen Anmerkungen." *H. Kusber (2006): Die harmonisierte Taxaliste des Phytoplanktons für Seen und Flüsse in Deutschland*.
- Mischke, U. und B. Nixdorf (2008). *Gewässerreport (Nr. 10): Bewertung von Seen mittels Phytoplankton zur Umsetzung der EU-Wasserrahmenrichtlinie*. BTU, Fak. Umweltwissenschaften und Verfahrenstechnik.
- Mischke, U., M. Venohr und H. Behrendt (2011). "Using Phytoplankton to Assess the Trophic Status of German Rivers". *International Review of Hydrobiology* 96.5, S. 578–598.
- Nabors, M. W. (2007). *Botanik*. Pearson Deutschland GmbH.
- Nixdorf, B., U. Mischke, E. Hoehn, U. Riedmüller, U. Rücker und I. Schönfelder (2006). "Leitbildorientierte Bewertung von Seen anhand der Teilkomponente Phytoplankton im Rahmen der Umsetzung der EU-Wasserrahmenrichtlinie". *BTU Cottbus, Lehrstuhl Gewässerschutz*.
- Noble, W. S. (2006). "What is a support vector machine?" *Nature Biotechnology* 24.12, S. 1565–1567.
- Ojala, T., M. Pietikainen und T. Maenpaa (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7, S. 971–987.
- Peng, H. (2008). "Bioimage informatics: a new area of engineering biology". *Bioinformatics* 24.17, S. 1827–1836.
- Pietzsch, T., S. Preibisch, P. Tomančák und S. Saalfeld (2012). "ImgLib2—generic image processing in Java". *Bioinformatics* 28.22, S. 3009–3011.
- Platt, J. (1998). "Sequential minimal optimization: A fast algorithm for training support vector machines". *Technical Report MSR-TR-98-14, Microsoft Research*.

- Prewitt, J. M. S. und M. L. Mendelsohn (1966). “The Analysis of Cell Images\*<sup>†</sup>”. *Annals of the New York Academy of Sciences* 128.3, S. 1035–1053.
- Prodanov, D. (2008). *ImageJ Image Moment Plugin*. URL: <http://rsbweb.nih.gov/ij/plugins/image-moments.html>.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Rasband, W. S. (1997). *ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA*. URL: <http://rsb.info.nih.gov/ij/>.
- Remsen, A., T. L. Hopkins und S. Samson (2004). “What you see is not what you catch: a comparison of concurrently collected net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder data from the northeast Gulf of Mexico”. *Deep Sea Research Part I: Oceanographic Research Papers* 51.1, S. 129–151.
- Reynolds, C. S. (1984). *The ecology of freshwater phytoplankton*. Cambridge University Press.
- Rezakhaniha, R., A. Agianniotis, J. T. C. Schrauwen, A. Griffa, D. Sage, C. V. C. Bouten, F. N. van de Vosse, M. Unser und N. Stergiopoulos (2012). “Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy”. *Biomechanics and modeling in mechanobiology* 11.3, S. 461–473.
- Riedmiller, M. und H. Braun (1993). “A direct adaptive method for faster backpropagation learning: the RPROP algorithm”. In: , *IEEE International Conference on Neural Networks, 1993*. , IEEE International Conference on Neural Networks, 1993, 586–591 vol.1.
- Robnik-Sikonja, M. und I. Kononenko (1997). “An adaptation of Relief for attribute estimation in regression”. In: *Fourteenth International Conference on Machine Learning*. Hrsg. von D. H. Fisher. Morgan Kaufmann, S. 296–304.
- Rodenacker, K., B. Hense, U. Jütting und P. Gais (2006). “Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation”. *Microscopy Research and Technique* 69.9, S. 708–720.
- Sato, M., Y. Murata, M. Mizusawa, H. Iwahashi und S.-i. Oka (2004). “A simple and rapid dual-fluorescence viability assay for microalgae”. *Microbiol Cult Coll* 20.2, S. 53–59.

- Schmid, B., J. Schindelin, A. Cardona, M. Longair und M. Heisenberg (2010). “A high-level 3D visualization API for Java and ImageJ”. *BMC Bioinformatics* 11.1, S. 274.
- Schulze, K., D. A. López, U. M. Tillich und M. Frohme (2011). “A simple viability analysis for unicellular cyanobacteria using a new autofluorescence assay, automated microscopy, and ImageJ”. *BMC Biotechnology* 11.1, S. 118.
- Schulze, K., U. M. Tillich, T. Dandekar und M. Frohme (2013). “PlanktoVision - an automated analysis system for the identification of phytoplankton”. *BMC Bioinformatics* 14.1, S. 115.
- Shendure, J. und H. Ji (2008). “Next-generation DNA sequencing”. *Nature Biotechnology* 26.10, S. 1135–1145.
- Sikora, T. (2001). “The MPEG-7 visual standard for content description-an overview”. *IEEE Transactions on Circuits and Systems for Video Technology* 11.6, S. 696–702.
- Smith, Z. J., K. Chu, A. R. Espenson, M. Rahimzadeh, A. Gryshuk, M. Molinaro, D. M. Dwyre, S. Lane, D. Matthews und S. Wachsmann-Hogiu (2011). “Cell-Phone-Based Platform for Biomedical Device Development and Education Applications”. *PLoS ONE* 6.3, e17150.
- Solimini, A. G., A. C. Cardoso und A.-S. Heiskanen (2006). “Indicators and methods for the ecological status assessment under the Water Framework Directive”. *European Commission, Joint Research Centre: Ispra. EUR report 22314*.
- Sommer, U. (2005). *Biologische Meereskunde*. Springer DE.
- Stanley, K. O. und R. Miikkulainen (2002). “Evolving Neural Networks through Augmenting Topologies”. *Evolutionary Computation* 10.2, S. 99–127.
- Streble, H. und D. Krauter (2006). *Das Leben im Wassertropfen, Mikroflora und Mikrofauna des Süßwassers, ein Bestimmungsbuch. 10*. Kosmos Naturführer.
- Sumner, M., E. Frank und M. Hall (2005). “Speeding Up Logistic Model Tree Induction”. In: *Knowledge Discovery in Databases: PKDD 2005*. Hrsg. von A. M. Jorge, L. Torgo, P. Brazdil, R. Camacho und J. Gama. Lecture Notes in Computer Science 3721. Springer Berlin Heidelberg, S. 675–683.

- Sundby, C., S. McCaffery und J. M. Anderson (1993). "Turnover of the photosystem II D1 protein in higher plants under photoinhibitory and nonphotoinhibitory irradiance." *Journal of Biological Chemistry* 268.34, S. 25476–25482.
- Taiz, L. und E. Zeiger (2010). *Plant Physiology*. 5. Aufl. Sinauer Associates, Inc.
- Tamura, H., S. Mori und T. Yamawaki (1978). "Textural Features Corresponding to Visual Perception". *IEEE Transactions on Systems, Man and Cybernetics* 8.6, S. 460–473.
- Tandeau de Marsac, N. (1977). "Occurrence and nature of chromatic adaptation in cyanobacteria." *Journal of Bacteriology* 130.1, S. 82–91.
- Tang, Y. Z. und F. C. Dobbs (2007). "Green autofluorescence in dinoflagellates, diatoms, and other microalgae and its implications for vital staining and morphological studies". *Applied and environmental microbiology* 73.7, S. 2306–2313.
- Teague, M. R. (1980). "Image analysis via the general theory of moments\*". *JOSA* 70.8, S. 920–930.
- Tillich, U. M., S. Lehmann, K. Schulze, U. Dühring und M. Frohme (2012). "The Optimal Mutagen Dosage to Induce Point-Mutations in *Synechocystis* sp. PCC6803 and Its Application to Promote Temperature Tolerance". *PLoS ONE* 7.11, e49467.
- Tinevez, J.-Y. (2010). *Imagej Plugin Directionality*. URL: <http://fiji.sc/Directionality>.
- Tu, J. V. (1996). "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes". *Journal of Clinical Epidemiology* 49.11, S. 1225–1231.
- Uchida, S. (2013). "Image processing and recognition for biological images". *Development, growth & differentiation* 55.4, S. 523–549.
- Utermöhl, H. (1958). "Zur Vervollkommnung der quantitativen Phytoplankton Methodik". *Mitteilungen. Internationale Vereinigung fuer Theoretische und Angewandte Limnologie* 9, S. 1–38.
- Van De Sande, K. E., T. Gevers und C. G. Snoek (2010). "Evaluating color descriptors for object and scene recognition". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.9, S. 1582–1596.

- Walker, R. F., P. Jackway und I. D. Longstaff (1995). “Improving co-occurrence matrix feature discrimination”. In: *DICTA '95, 3rd Conference on Digital Image Computing: Techniques and Application*, S. 643–648.
- Wei, Q., H. Qi, W. Luo, D. Tseng, S. J. Ki, Z. Wan, Z. Göröcs, L. A. Bentolila, T.-T. Wu und R. Sun (2013). “Fluorescent imaging of single nanoparticles and viruses on a smart phone”. *ACS nano* 7.10, S. 9147–9155.
- White, A. M., D. S. Daly, A. R. Willse, M. Protic und D. P. Chandler (2005). “Automated Microarray Image Analysis Toolbox for MATLAB”. *Bioinformatics* 21.17, S. 3578–3579.
- Whiteson, S., P. Stone, K. O. Stanley, R. Miikkulainen und N. Kohl (2005). “Automatic Feature Selection in Neuroevolution”. In: *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*. GECCO '05. New York, NY, USA: ACM, S. 1225–1232.
- Wilkinson, L., A. Anand und D. N. Tuan (2011). “CHIRP: A New Classifier Based on Composite Hypercubes on Iterated Random Projections”. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '11. New York, NY, USA: ACM, S. 6–14.
- Wollman, R. und N. Stuurman (2007). “High throughput microscopy: from raw images to discoveries”. *Journal of cell science* 120.21, S. 3715–3722.
- Yan, Q. und Y. Yu (2011). “Metagenome-based analysis: A promising direction for plankton ecological studies”. *Science China Life Sciences* 54.1, S. 75–81.
- Yao, D. C., D. C. Brune und W. F. Vermaas (2012). “Lifetimes of photosystem I and II proteins in the cyanobacterium *Synechocystis* sp. PCC 6803”. *FEBS letters* 586.2, S. 169–173.
- Zetsche, E.-M. und F. J. R. Meysman (2012). “Dead or alive? Viability assessment of micro- and mesoplankton”. *Journal of Plankton Research* 34.6, S. 493–509.

# Erklärung

Hiermit erkläre ich diese Dissertation selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Die Dissertation wurde weder in gleicher noch in ähnlicher Form in einem Prüfungsverfahren vorgelegt. Ich erkläre zudem, dass ich bisher keinen akademischen Grad erworben habe oder versucht habe zu erwerben.

Würzburg, den

---

Katja Schulze