

# Aktives Lernen zur Detektion von Pflanzenkrankheiten in hyperspektralen Bildern

PETER SCHMITTER<sup>1</sup>, JAN BEHMANN<sup>1</sup>, JÖRG STEINRÜCKEN<sup>1</sup>, ANNE-KATRIN MAHLEIN<sup>2</sup>, ERICH-CHRISTIAN OERKE<sup>2</sup> & LUTZ PLÜMER<sup>1</sup>

*Zusammenfassung: Aktives Lernen ermöglicht durch die gezielte Wahl von Trainingsdaten eine Reduzierung des Annotationsaufwands. Ausgehend von einem Auswertemodell auf Basis einer kleinen Menge von Trainingsdaten wird diese Menge durch einen Anwender interaktiv und iterativ erweitert. In diese Arbeit wird Aktives Lernen für die Detektion und Unterscheidung von Pflanzenkrankheiten (Cercospora Blattflecken und Echter Mehltau) in hyperspektralen Bildern genutzt. Es wird gezeigt, dass Aktives Lernen im Vergleich zu einem als fehlerfrei angenommenen Referenzmodell bereits mit einer geringen Anzahl an Trainingsdaten sehr hohe Klassifikationsgenauigkeiten erreicht. Die Betrachtung der relativen Klassenhäufigkeiten in Zeitreihen hyperspektraler Bilder ermöglicht die Detektion und Differenzierung der Pflanzenkrankheiten.*

## 1 Einleitung und Problemstellung

Die Auswertung hyperspektraler Bilder von Pflanzen ermöglicht die Detektion gestresster Pflanzen und die Differenzierung des auslösenden Stressfaktors. Stress ist die Reaktion der Pflanze auf Abweichungen von optimalen Wachstumsbedingungen, die ein tolerierbares Maß übersteigen (z.B. Trockenheit oder Krankheitserreger). Die rechtzeitige Erkennung von gestressten Pflanzen ermöglicht eine gezielte Behandlung und damit die Minimierung von Ertragsverlusten.

Die Auswertung hyperspektraler Bilder ist herausfordernd. In den letzten Jahren wurden erfolgreich Methoden des Maschinellen Lernens eingesetzt, eine Übersicht geben z.B. BEHMANN et al. (2014). Für die Auswertung mit überwachten Lernverfahren (z.B. Klassifikation) ist eine ausreichende Anzahl annotierter Trainingsdaten notwendig. Da die Annotation meist Expertenwissen erfordert und dementsprechend mit hohen Kosten verbunden ist, wird häufig nur eine geringe Anzahl der Pixel annotiert. Es stehen damit nur wenige Trainingsdaten zur Verfügung, die sich noch dazu nicht optimal auf dem Bild verteilen. Insbesondere Randbereiche sind häufig unterrepräsentiert. Die resultierenden Auswertemodelle modellieren die Trainingsdaten, liefern aber auf neuen Daten unbefriedigende Ergebnisse. Um die Auswertemodelle zu verbessern, werden ergänzende Trainingsdaten benötigt.

Aktives Lernen ist eine Strategie zur *gezielten* Selektion und Annotation von Daten, die der Menge der Trainingsdaten hinzugefügt werden (LEWIS et al. 1994). Durch die Erweiterung der Trainingsdaten kann iterativ ein robusteres Modell gelernt werden, woraus eine verbesserte Prädiktion für neue, nicht annotierte Daten folgt. Die Strategie zur Datenselektion ist abhängig

---

<sup>1</sup> Universität Bonn, Institut für Geodäsie und Geoinformation, Meckenheimer Allee 172, 53115 Bonn; E-Mail: [schmitter, behmann, steinbruecken, pluemer]@igg.uni-bonn.de

<sup>2</sup> Universität Bonn, Institut für Nutzpflanzenwissenschaften und Ressourcenschutz, Professur für Pflanzenkrankheiten u. Pflanzenschutz (INRES)-Phytophthora, Meckenheimer Allee 166a, 53115 Bonn

vom gewählten Klassifikator (WUTTKE et al. 2014). Für die Annotation der selektierten Daten ist der Eingriff eines Anwenders mit Expertenwissen nötig. Im Fall von hyperspektralen Bildern muss dazu neben der räumlichen Lage des Pixels im Bild das dazugehörige Spektrum betrachtet werden. Häufig ist erst durch die kombinierte Information von Pixelposition und elektromagnetischem Spektrum eine eindeutige Klassenzuweisung möglich.

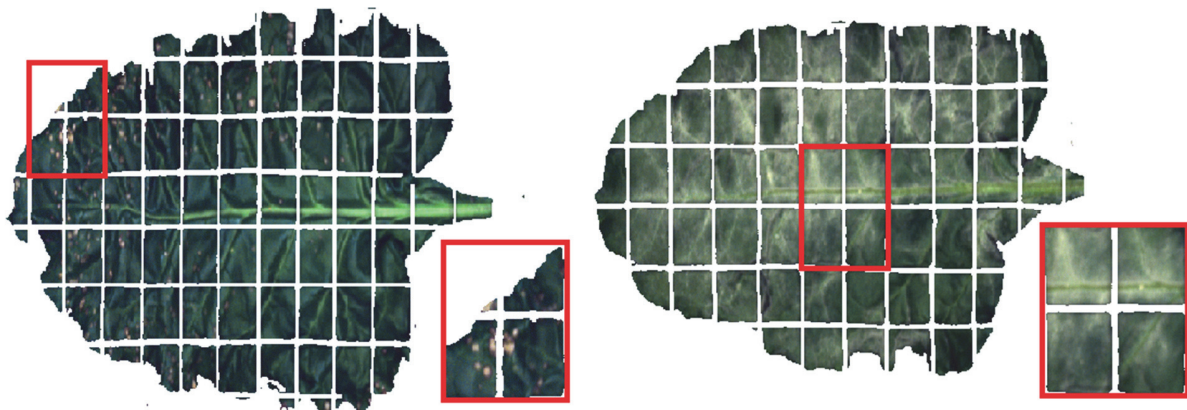


Abb. 1: Aus hyperspektralen Aufnahmen abgeleitete RGB-Darstellung eines mit Cercospora-Blattflecken (links) und eines mit Echtem Mehltau befallenen Zuckerrübenblattes.

In dieser Arbeit wird Aktives Lernen für die Detektion und Unterscheidung von Pflanzenkrankheiten (Cercospora Blattflecken und Echter Mehltau) in hyperspektralen Bildern von Zuckerrübenblättern genutzt (Abb. 1). Als Klassifikator wird Support Vector Machine verwendet. Es wird gezeigt, dass Aktives Lernen im Vergleich zu einem als fehlerfrei angenommenen Referenzmodell bereits mit einer geringen Anzahl an Trainingsdaten sehr hohe Klassifikationsgenauigkeiten erreicht. Die Betrachtung der relativen Klassenhäufigkeiten in Zeitreihen hyperspektraler Bilder ermöglicht die Detektion und Differenzierung der Pflanzenkrankheiten.

## 2 Aktives Lernen

### 2.1 Support Vector Machine

Support Vector Machine (SVM) ist ein überwachtes Lernverfahren zur Klassifikation, welches eine Menge von Daten mittels Hyperebene in zwei Klassen teilt (BURGES 1998). Für die Klassifikation wird die Hyperebene bestimmt, welche den Margin maximiert. Der Margin ist hier definiert als Abstand der Ebene zu den annotierten Daten. Die Lage der Hyperebene wird durch die Support Vektoren beschrieben, welche eine Teilmenge der annotierten Trainingsdaten sind. Die Anzahl der Support Vektoren in Relation zur Menge der annotierten Trainingsdaten dient als obere Grenze für den *leave-one-out* Fehler, welcher den Generalisierungsfehler des Klassifikators schätzt (CHAPELLE et al. 2002). Neue Daten werden anhand ihrer Lage zur Hyperebene einer Klasse zugewiesen. Für eine bessere Interpretation der Distanzen der Daten zur Hyperebene, werden diese mit Hilfe einer Sigmoid-Funktion in Wahrscheinlichkeiten transformiert (PLATT 1999).

Mehrklassenprobleme lassen sich mit Hilfe dieses binären Klassifikators lösen, indem das Mehrklassenproblem in mehrere Zweiklassenprobleme zerlegt wird. In dieser Arbeit wird dafür die „one-vs-one SVM“ verwendet (FÜRNKRANZ 2002).

## 2.2 Aktives Lernen

Aktives Lernen bezeichnet im maschinellen Lernen eine Selektionsstrategie zur Erweiterung der Menge der annotierten Trainingsdaten  $L$  (LEWIS et al. 1994). Ziel des Aktiven Lernen ist es mit minimalem Aufwand die Menge der Trainingsdaten gezielt so zu erweitern, dass ein besseres und robusteres Auswertemodell gelernt werden kann. Die zu annotierenden Daten werden aus einer Menge nicht annotierter Daten  $U$  selektiert (*pool-based* aktives Lernen, SETTLES 2009).

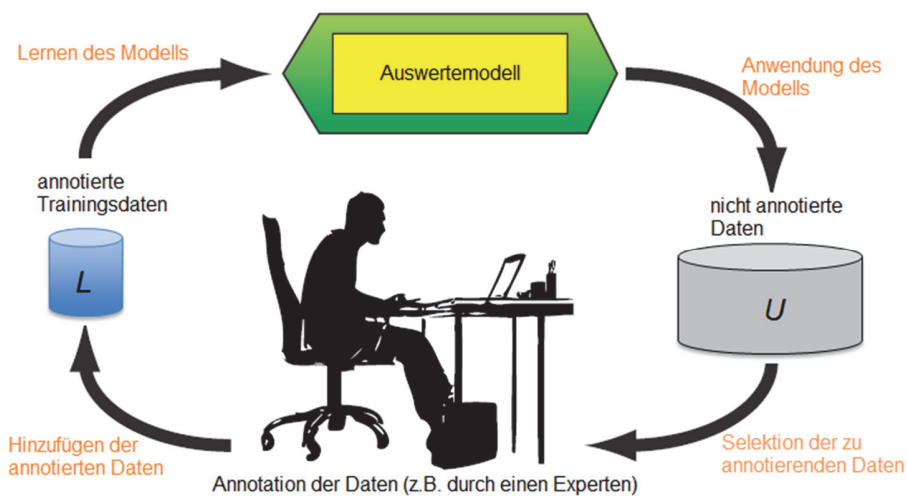


Abb. 2: Schematische Darstellung des pool-based aktiven Lernens (modifiziert nach SETTLES 2009)

Im Verlauf des aktiven Lernens wird zunächst auf Basis der bereits vorhandenen Trainingsdaten  $L$  ein Auswertemodell gelernt. Dieses Modell wird anschließend auf die Menge der nicht annotierten Daten  $U$  angewendet. Basierend auf dem Klassifikationsergebnis werden Daten, welchen ein Label hinzugefügt werden soll, aus  $U$  selektiert. Nachdem diesen Daten ein Label zugewiesen wurde, werden sie der Menge an Trainingsdaten  $L$  hinzugefügt. Anschließend wird ein neues Auswertemodell gelernt, welches auf eine größere Menge an Trainingsdaten zurückgreift. Dieses Vorgehen wird wiederholt bis das Auswertemodell ein zufriedenstellendes Ergebnis liefert.

Für die Auswahl der zu annotierenden Daten gibt es verschiedene Selektionsstrategien, die abhängig vom gewählten Auswertemodell sind. Die beiden bekanntesten Strategien sind das *Uncertainty Sampling* und das *Query by Committee*. Beim *Uncertainty Sampling* werden die Daten selektiert, welche die größte Klassifikationsunsicherheit aufweisen (LEWIS et al. 1994). Wenn keine Wahrscheinlichkeiten vorliegen, kann *Query by Committee* verwendet werden. Dabei werden mehrere Klassifikationsmodelle gelernt, welche jeweils auf einer Teilmenge der Trainingsdaten basieren und jeweils Prädiktionen für  $U$  bestimmen. Die unsichersten Daten sind dann definiert durch die höchste Unreinheit der Prädiktionen (SEUNG et al. 1992). Da die ständige Interaktion mit einem Experten im Aktiven Lernen hohe Anforderungen an die

Performanz der Algorithmen stellt, wird in dieser Arbeit auf das *Uncertainty Sampling* zurückgegriffen.

Um die Anzahl der Iterationen und somit das aufwendige Lernen neuer Modell zu reduzieren, lassen sich mehrere Instanzen pro Iteration selektieren und annotieren (XU et al. 2007). Dabei gilt es zu beachten, dass bei großen Datensätzen, wie hyperspektralen Bildern, die unsichersten Daten eine hohe räumliche und spektrale Ähnlichkeit aufweisen. Um zu vermeiden, dass in einer Iteration mehrere Instanzen mit redundanten Informationen den Trainingsdaten hinzugefügt werden, werden aus der Menge der unsicheren Trainingsdaten die repräsentativen Instanzen bestimmt (XU et al. 2003). Dazu werden bei  $k$  hinzuzufügenden Trainingsdaten die  $l$  unsichersten Daten mit Hilfe des unüberwachten Clusterverfahrens  $k$ -Means (BISHOP 2006) in  $k$  Gruppen aufgeteilt. Innerhalb jedes Clusters wird anschließend das Medoid bestimmt, welcher dem Zentrum der Gruppe am nächsten liegt und demnach die Gruppe repräsentiert. Das Medoid jeder Gruppe wird anschließend selektiert, annotiert und der Menge der Trainingsdaten hinzugefügt. Die Anzahl  $l$  der unsichersten Daten und die Anzahl  $k$  der zu annotierenden Daten wird abhängig von der Menge der nicht annotierten Daten gewählt.

### 3 Aktives Lernen für hyperspektrale Bilder von Pflanzen

#### 3.1.1 Pflanzenmaterial und Pathogene

Zuckerrübenpflanzen der Sorte Paulette (KWS Einbeck) wurden unter kontrollierten Bedingungen in einem Gewächshaus bei 23/20°C (Tag/Nacht),  $60 \pm 10\%$  relativer Luftfeuchte (RH) und einer Photoperiode von 16 h pro Tag kultiviert. Hyperspektrale Messungen wurden an Einzelblättern im Entwicklungsstadium BBCH<sup>3</sup> 16 durchgeführt, welches der Blattentwicklung des Hauptsprosses entspricht. Die Messungen wurden an gesunden Zuckerrübenpflanzen und an Zuckerrübenpflanzen, inokuliert mit *Cercospora beticola* (Erreger der Cercospora Blattflecken) und mit *Erysiphe betae* (Erreger des Echten Mehltaus der Rübe) durchgeführt. Die Inokulation erfolgte wie in MAHLEIN et al. (2012) beschrieben. Pro Variante wurden 3 Blätter täglich bis 21 Tage nach Inokulation (dai) hyperspektral beobachtet.

#### 3.1.2 Hyperspektrale Messungen und Vorprozessierung der Bilddaten

Hyperspektrale Bilder wurden mit dem hyperspektralen, bildgebenden Linienscanner System ImSpector V10E (Spectral Imaging Ltd., Oulu, Finnland) aufgenommen. Der Linienscanner misst 211 Kanalbänder in einem spektralen Bereich von 400 – 1000 nm, mit einer spektralen Auflösung von bis zu 2,8 nm. Die maximale Bildgröße ist durch die Anzahl der Linienbildpixel (1600 Pixel pro Linie) und durch die Sensorpixelgröße von 0,0074 mm charakterisiert. Bedingt durch den Abstand zwischen Kamera und Objekt wurde in den zu Grunde liegenden Messungen eine räumliche Auflösung von 0,019 mm pro Pixel erreicht. Zur Aufnahme 2-dimensionaler Bildsätze mit einer spektralen Dimension wurde das Objekt mit Hilfe eines Spiegelscanners erfasst. Die Steuerung und Einstellung der Kamera erfolgte mit der Software SpectralCube

---

<sup>3</sup> BBCH (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie) bezeichnet eine Skala zur einheitlichen Codierung der Entwicklungsstadien von Pflanzen (MEIER 2001)

(Spectral Imaging Ltd., Oulu, Finnland). Alle Messungen wurden in einer Dunkelkammer bei einer konstanten Beleuchtung durch 6 ASD Pro-Lamps (Analytic Spectral Devices, Boulder, USA) durchgeführt. Für die hyperspektralen Messungen wurden die Zuckerrübenblätter in einem Rahmen fixiert und auf mobilen Tischen platziert. Für die nachfolgende Normalisierung der Bilddaten wurde zunächst ein Weißstandard (Bariumsulfat) gemessen, anschließend wurde der dazugehörige Dunkelstrom der Kamera erfasst. Abschließend wurde eine Aufnahme der Zuckerrübenblätter durchgeführt. Die Berechnung der Reflektion, relativ zum Weißstandard und zum Dunkelstrom sowie die Maskierung des Hintergrundes erfolgte in ENVI<sup>4</sup> 4.6 + IDL<sup>5</sup> 7.0 (ITT Visual Information Solutions, Boulder, USA).

### 3.2 Anwendung auf hyperspektrale Bilder von Pflanzen

Für das Lernen der Auswertemodelle wurde pro Krankheit eine hyperspektrale Aufnahme mit deutlich sichtbaren Symptomen ausgewählt (Abb. 1). Die beiden Bilder wurden am Tag 17 nach der Innokulation der Pflanzenkrankheiten aufgenommen. Für die Erstellung eines als fehlerfrei angenommenen Referenzmodells wurden 300 Instanzen manuell ausgewählt und manuell den Klassen (Gesund, Cercospora Mitte, Cercospora Rand und Mehltau) zugewiesen. Dies entspricht dem traditionellen Vorgehen zur Erstellung von Trainingsdaten. Die Klasse Gesund ist mit 150 Instanzen deutlich ausgeprägter als die restlichen Klassen (Cercospora Mitte: 50, Cercospora Rand: 50, Mehltau: 50), um alle Strukturen des gesunden Blattes (helle Bereiche, dunkle Bereich, Stängel) ausreichend zu repräsentieren. Es wurden mehr Instanzen annotiert als notwendig (Abb. 4), um sicherzustellen, dass das Referenzmodell keine signifikanten Defizite aufweist. Die Prädiktionen für die beiden Bilder durch das Referenzmodell wurden als fehlerfreie Solllösungen angenommen und zur Validierung der Ergebnisse des Aktiven Lernens verwendet.

Die Auswahl und Annotation der Trainingsdaten basierte auf Vorwissen. Dazu wurden die Pixelposition und das dazugehörige Spektrum parallel visualisiert, sodass auf Basis der räumlichen und spektralen Information die manuelle Klassenzuweisung vorgenommen werden konnte. Für die Visualisierung wurden aus den hyperspektralen Aufnahmen RGB-Bilder abgeleitet. Dies erlaubte es dem Anwender die lokale Umgebung der entsprechenden Instanz zu betrachten und daraus erste Schlüsse über die Klassenzuweisung zu ziehen. Um eine endgültige Entscheidung über die Klassenzuweisung treffen zu können wurde der Anwender durch das visualisierte Spektrum der selektierten Instanz unterstützt. Diese Mehrinformation ermöglichte, insbesondere bei Übergängen zwischen Klassen, eine eindeutigere Klassenzuweisung.

Die Basis für das Aktive Lernen war ein Ausgangsmodell, welches auf 20 zufällig gewählten Trainingsdaten basiert. Dazu wurden aus der Menge der 300 Referenztrainingsdaten zufällig fünf Instanzen pro Klasse ausgewählt. Anschließend wurde das gelernte SVM-Modell auf die beiden Bilder angewendet. Solange die Klassifikationsgenauigkeit im Vergleich zur Referenzlösung unter 95% lag, wurden anschließend durch *uncertainty sampling* und Anwendung des *k*-Means (*k*=5) weitere Daten selektiert. Die selektierte Instanz wurde wie oben beschrieben visualisiert

---

<sup>4</sup> ENVI (Environment for Visualizing Images) ist eine Analysesoftware für Geodaten.

<sup>5</sup> IDL (Interactive Data Language) ist eine Programmiersprache für die Datenanalyse.

und annotiert. Auf diese Weise wurden die Trainingsdaten in jeder Iteration um jeweils fünf, möglichst informative, Instanzen ergänzt.

## 4 Ergebnisse

Das Ausgangsmodell auf Basis der 20 Trainingsdaten liefert für beide Aufnahmen keine zufriedenstellenden Ergebnisse (Abb. 3a, e). Insbesondere die Randbereiche des mit *Cercospora* innokulieren Bildes werden fälschlich der Klasse *Cercospora* Rand zugewiesen (Abb. 3a). Außerdem werden einige Bereiche des mit Mehltau innokulieren Blattes ebenfalls der Klasse *Cercospora* Rand zugewiesen (Abb. 3e). Nach der ersten Iteration des Aktiven Lernens werden die Randbereich des mit *Cercospora* innokulieren Blattes korrekt klassifiziert (Abb. 3b), jedoch werden noch große Bereich des mit Mehltau innokulierten Bildes einer falschen Klasse zugewiesen. Diese Fehlklassifikationen treten nach einer weiteren Iteration des Aktiven Lernen nicht mehr auf (Abb. 3g), jedoch werden kleine Bereiche des mit *Cercospora* innokulierten Blattes als Mehltau klassifiziert (Abb. 3c). Nach dem Hinzufügen von weiteren fünf Daten zu der Menge der Trainingsdaten wird auf beiden Bildern ein zufriedenstellendes Klassifikationsergebnis erreicht (Abb. 3d, h). In beiden Bildern werden die *Cercospora* Blattflecken und der Echte Mehltau erfolgreich detektiert.

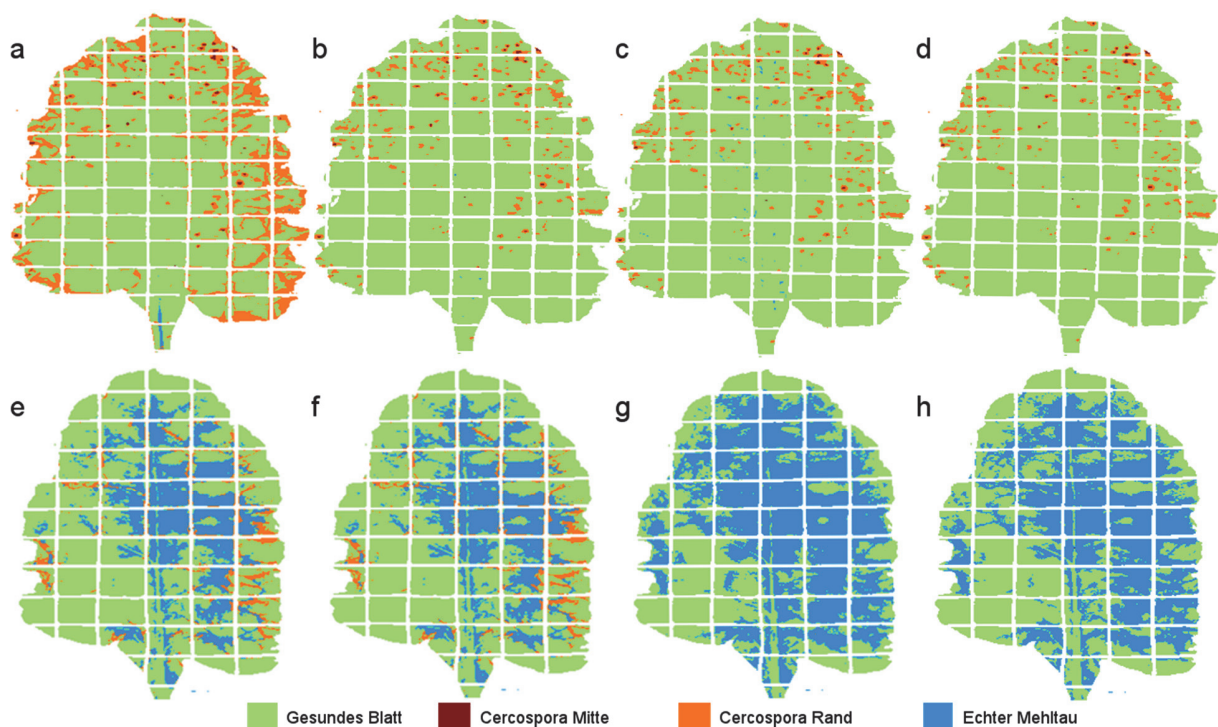


Abb. 3: Klassifikationsergebnisse eines mit *Cercospora* Blattflecken (a-d) und eines mit Echem Mehltau (e-h) innokulierten Blattes. Die Ausgangslage (a,e) wird durch das jeweilige hinzufügen von fünf Instanzen verbessert. Das Aktive Lernen wird abgebrochen nachdem ein Klassifikationsergebnis mit 95% Übereinstimmung zur Referenzlösung erreicht wurde (d, h).

Durch die drei Iterationen des Aktiven Lernens wurden der Menge der annotierten Daten 15 weitere Instanzen hinzugefügt, sodass insgesamt 35 annotierte Daten vorlagen. Auf Basis dieser 35 annotierten Daten wurde das SVM-Modell gelernt, welches zu dem in Abbildung 2d, h dargestellten Ergebnis führt. Das SVM-Modell basiert auf 25 Support Vektoren. Es wurde eine Klassifikationsübereinstimmung von 95% zum Sollergebnis erreicht. Zum Vergleich wurden Modelle auf Basis einer zufälligen Auswahl aus den 300 manuell annotierten Trainingsdaten gelernt. Dazu wurde die Anzahl der Trainingsdaten ausgehend von 20 um jeweils fünf zufällig ausgewählte Instanzen erweitert. Das Prädiktionsergebnis wurde anschließend mit dem Sollergebnis verglichen. Eine Klassifikationsübereinstimmung von 95% wird mit 65 zufällig annotierten Trainingsdaten erreicht (Abb. 4), wobei das Modell auf 29 Support Vektoren basiert.

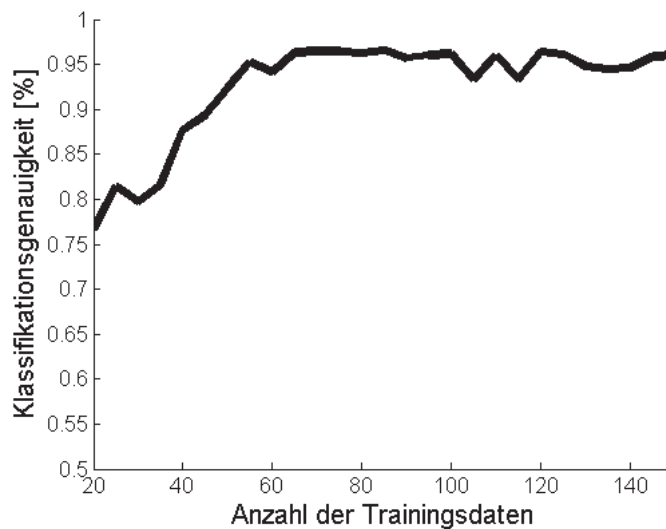


Abb. 4: Klassifikationsergebnis bei zufälliger Auswahl der Trainingsdaten

Das Referenzmodell, basierend auf allen 300 manuell annotierten Trainingsdaten besteht aus 46 Support Vektoren.

Tab. 1: Vergleich des Aktiven Lernens mit dem traditionellen Vorgehen

	Aktives Lernen	Manuell annotiert	
		Zufällige Auswahl	Alle
Anzahl der annotierten Daten	35	65	300
Support Vektoren	25	29	46
Klassifikationsgenauigkeit	95%	95%	100%

Obwohl das Modell, welches auf den mit Aktivem Lernen selektierten Daten basiert, deutlich weniger annotierte Daten benötigt, ist eine Detektion und Differenzierung der beiden Pflanzenkrankheiten möglich. Im Gegensatz zu der zufälligen Selektion der Trainingsdaten konnte der Annotationsaufwand um nahezu die Hälfte reduziert werden. Beide Modelle liefern eine vergleichbare Prädiktionsgenauigkeit und Modellkomplexität.

Für die Detektion und Differenzierung der beiden Pflanzenkrankheiten über die Zeit wurde das auf den durch Aktives Lernen gewählten Trainingsdaten basierende Auswertemodell auf die

Zeitreihen der neun gemessenen Blätter (3 Cercospora Blattflecken, 3 Echter Mehltau, 3 Kontrolle) angewendet. Anschließend wurden für jedes Blatt die relativen Klassenhäufigkeiten bestimmt, sodass der Anteil an infizierter Blattfläche bestimmt werden kann. Die Häufigkeiten der beiden Krankheiten (Cercospora und Mehltau) sind in Abbildung 5 dargestellt. Darin sind die Kurvenverläufe der infizierten Blätter deutlich erkennbar, da der Anteil der infizierten Pixel an der Blattfläche bei den jeweils drei innokulierten Blättern über die Zeit zunimmt. Bei den infizierten Blättern ist erkennbar, dass die infizierte Blattfläche variiert. Jedoch lässt die Zunahme der infizierten Blattfläche eine Detektion der infizierten Blätter zu. Die Kontrollgruppe weist korrekterweise kaum infizierte Blattfläche auf. Somit lassen sich die innokulierten Blätter von den gesunden Blättern unterscheiden. Weiterhin können die unterschiedlichen Symptome voneinander unterschieden werden.

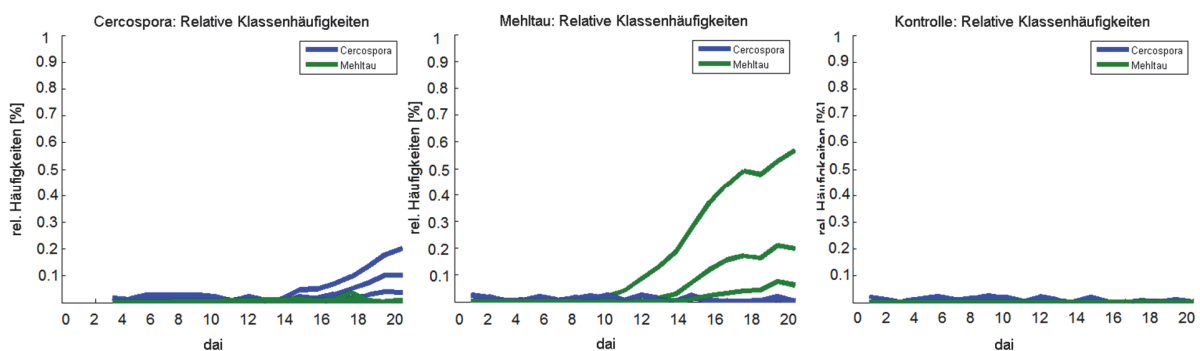


Abb. 5: Relative Klassenhäufigkeiten der mit Cercospora (links), Mehltau (mitte) innokulierten Blätter und der Kontrollgruppe (rechts) über die Zeit.

## 5 Fazit

Diese Arbeit hat gezeigt, dass durch die gezielte Auswahl der Trainingsdaten der Aufwand des Annotierens reduziert werden kann. Für die Annotation einer Instanz ist die kombinierte Visualisierung der räumlichen und spektralen Information erforderlich. Es wurde gezeigt, dass mit einer deutlich geringeren Anzahl an Trainingsdaten ein SVM-Modell gelernt werden kann, welches eine zu einem Referenzmodell vergleichbare Klassifikationsgüte erreicht. Trotz der geringeren Anzahl an Trainingsdaten konnte ein robustes Modell gelernt werden, welches erfolgreich auf die gesamte Zeitreihe der gemessenen hyperspektralen Aufnahmen angewendet wurde. Dadurch ist eine frühzeitige Detektion und Differenzierung der beiden Pflanzenkrankheiten möglich.

## 6 Danksagung

Die Autoren bedanken sich für die Förderung des Projekts CROP.SENSE.net im Kontext des Ziel 2 Programms NRW 2007-2013 „Regionale Wettbewerbsfähigkeit und Beschäftigung (EFRE)“ durch das Ministerium für Innovation, Wissenschaft und Forschung (MIWF) des Landes Nordrhein-Westfalens (NRW) und dem Fond für Regionale Entwicklung (EFRE) (005-1103-0018) der Europäischen Union während der Erstellung dieses Manuskripts.



## 7 Literaturverzeichnis

- BEHMANN, J., MAHLEIN, A.-K., RUMPF, T., RÖMER, C. & PLÜMER, L., 2014: A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, doi: 10.1007/s11119-014-9372-7 (online first).
- BISHOP, C.M., 2006: *Pattern Recognition and Machine Learning*. New York, USA, Springer.
- BURGES, C.J.C., 1998: A Tutorial on Support Vector Machines for Pattern Recognition. *Mining and Knowledge Discovery* **2**, S. 121-167.
- CHAPELLE, O., VAPNIK, V.; BOUSQUET O. & MUKHERJEE, S., 2002: Choosing Multiple Parameters for Support Vector Machines. *Machine Learning* **46**, S. 131-159
- FÜRNKRANZ, J., 2002: Round robin classification. *The Journal of Machine Learning Research* **2**, S. 721-747.
- LEWIS, D. D. & GALE, W. A., 1994: A sequential algorithm for training text classifiers. *Proceedings of the 17<sup>th</sup> annual international ACM SIGIR conference on Research and development in information retrieval*, Springer-Verlag New York Inc., S. 3-12.
- MAHLEIN, A.-K., STEINER, U., HILLNHÜTTER, C., DEHNE, H.-W. & OERKE, E.-C., 2012: Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant methods* **8** (1), S. 3.
- MEIER, U., 2001: *Entwicklungsstadien mono- und dikotyler Pflanzen - BBCH Monographie*. Biologische Bundesanstalt für Land und Forstwirtschaft, Germany.
- PLATT, J. C., 1999: *Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods*. *Advances in Large Margin Classifiers*, MIT Press
- SETTLES, B., 2009: *Active Learning Literature Survey*. Computer Science Technical Report 1648, University of Wisconsin-Madison.
- SEUNG, H.S., OPPER, M. & SOMPOLINSKY, H., 1992: Query by Committee. *Proceedings of the fifth annual workshop on computational learning theory*, S. 287-294.
- WUTTKE, S.; MIDDELMANN, W. & STILLA, U., 2014: *Bewertung von Strategien des Aktiven Lernens am Beispiel der Landbedeckungsklassifikation*. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, Band 23, Seyfert, E., Gülch, E., Heipke, C., Schiewe, J. & Sester, M. (Hrsg.), 34. Wissenschaftlich-Technische Jahrestagung der DGPF, 26.-28. März 2014 an der HafenCity Universität Hamburg, Tagungsband auf CD-ROM, Beitrag 197.
- XU, Z., AKELLA, R. & ZHANG, Y., 2007: Incorporating Diversity and Density in Active Learning for Relevance Feedback. *Proceedings of the European Conference on IR Research (ECIR)*, Springer-Verlag, S. 246-257.
- XU, Z., YU, K., TRESP, V., XU, X. & WANG, J., 2003: Representative sampling for text classification using support vector machines. *25<sup>th</sup> European Conf. on Information Retrieval Research*, Springer, S. 393-407.