# Abschlussbericht

## VIVID - Videobasierte Vigilanz Detektion

### Im Rahmen der BMBF-Bekanntmachung

IKT 2020 - Forschung für Innovationen

| | |
|---|---|
| Projektlaufzeit: | 01.07.2015 - 30.09.2018 |
| Zuwendungsempfänger: | SASS acoustic research & design GmbH |
| Förderkennzeichen: | 01IS15024 **B** |
| Vorhabensbezeichnung: | "Mathematische Modellierung" |
| Berichtszeitraum: | 01.01.2018 - 30.06.2018 |

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS15024 B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

| Version | Datum | Autor |
|---|---|---|
| 1.0 | 28.03.2019 | Rudolf Bisping |

# Revisionen

| *Datum* | *Version* | *Beschreibung* | *Autor* |
|---|---|---|---|
| **22.02.2019** | **Version 0.1** | **Erste Version** | **Rudolf Bisping** |
| **25.02.2019** | **Version 0.2** | **Zweite Version** | **Rudolf Bisping** |
| **28.03.2019** | **Version 1.0** | **Endfassung** | **Rudolf Bisping** |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Inhaltsverzeichnis

# 1 Einleitung

Im Abschlussbericht wird der Stand des SASS VIVID Projekts zum 30.09.2018 beschrieben.

## 1.1 Aufgabenstellung

Ziel der generischen Modellierung von SASS ist es den physiologischen Vigilanzstatus von Personen mit Hilfe von videobasierten Messungen vorherzusagen. Typisch für die generische Modellierung ist es, die Daten des jeweiligen Nutzers mit denjenigen einer Stichprobe von Probanden zu vergleichen. Das Modell verarbeitet die einlaufenden Messdaten und verdichtet sie zu einem Vigilanzwert.

Als primäres Ziel für das Projektende ist geplant die Modellierung des generische Modells (AP 3.1) auf der Basis neuer Daten zu vervollständigen. Diese Daten stehen in Q3 2018 zur Verfügung.

Ziel des individualspezifischen Modells (AP 3.2) ist es, den Vigilanzstatus einzelner Personen vorherzusagen. Der Unterschied zum generischen Modell besteht darin, dass beim individualspezifischen Modell die Vorhersage ausschließlich auf den Daten des Nutzers selbst beruht, die ständig aktualisiert werden. Durch den erhöhten zeitlichen Aufwand für das generische Modell muss die ursprüngliche Zielsetzung eingeschränkt werden. Die Aufgabenstellung zum Projektende besteht darin, den bestmöglichen Algorithmus zur Realisierung des individualspezifischen Modells zu finden.

### Gliederung

In Kapitel 2 wird der abschließende Projektstand in Relation zur Planung zusammengefasst.

In Kapitel 3 werden die in der Einleitung beschriebenen Arbeiten dargestellt. Kapitel 4 zieht ein Fazit des Projektstands. Kapitel 5 beschreibt die Fortschreibung des Verwertungsplans.

# 2 Zusammenfassung des Projektstandes

## 2.1 Vergleich des Stands des Vorhabens mit der ursprünglichen Planung

Die Datenbasis für das generische Modell hat sich im Verlaufe des VIVID Projekts sukzessive erweitert. In Abhängigkeit von diesen Erweiterungen war eine Vielzahl von Neumodellierungen erforderlich, die über den ursprüngliche geplanten Zeitrahmen weit hinausgingen und erst jetzt, zum Projektende, ihren Abschluss finden.

Beim individualspezifischen Modell können nicht alle geplanten Ziele auf Grund der zeitlichen Ausweitung des generischen Modells erreicht werden, u.a die Programmierung eines C-Scoring Codes als einem wesentlichen Teilziel. Dennoch kann ein Vergleich durchgeführt werden, mit dem der am besten für die individualspezifische Modellierung geeignete Algorithmus gefunden wird.

## 2.2 Durchführungsrelevante Ergebnisse Dritter

Es wurden keine Ergebnisse Dritter bekannt, die für das vorliegende Projekt relevant sind.

## 2.3 Notwendige Änderungen der Zielsetzung

Für AP3.1 erfolgt trotz der veränderten Zeitplanung keine neue Zielsetzung. Die neue Zielsetzung für AP3.2 sieht vor, grundlegende Analysen der Thematik vorzunehmen, aber keine technische Online Lösung anzustreben.

## 2.4 Arbeitstreffen und Kommunikation im Projekt

Die SASS GmbH hat an geplanten Treffen und Telefonkonferenzen teilgenommen.

# 3 Durchgeführte Arbeiten in AP 3 (Mustererkennung)

## 3.1 Generisches Modell

### 3.1.1 Modellstruktur

Die aktuelle von Psyrecon gemessene Datenbasis umfasst N=321 Labor- und Fahrversuche, deren Dauer zwischen 1.5 und 4.5 Stunden liegt. 60% der Daten werden auf der Straße gemessen. Die übrigen Daten werden im Labor erhoben. Der Altersrange von Probanden beiderlei Geschlechts liegt zwischen 18-65 Jahren.

In jeder experimentellen Sitzung werden zwei verschiedene Typen von Variablen gemessen:

1.) Physiologische Daten (Electroenzephalogramm – EEG, Okulomotorische Aktivität – EOG, Elektrodermale Aktivität – EDA, Herz Aktivität – EKG)
2.) Video Kopf- und Augenbewegungen

Im Modell sind die physiologischen Variablen als Zielgrößen definiert. Die Videovariablen sind die Prädiktoren.

Abbildung 1 zeigt die Datenstruktur des Modells. Vor der eigentlichen Modellierung werden die Daten vorverarbeitet und resultieren in zwei getrennten Datensätzen für die Zielvariablen und die Prädiktoren. In der nachfolgenden Modellierung mittels Support Vector Machines (SVM[1], [1]) werden sie mathematisch miteinander kombiniert. Die Vorhersage der Vigilanz erfolgt ausschließlich auf der Basis der Videovariablen.

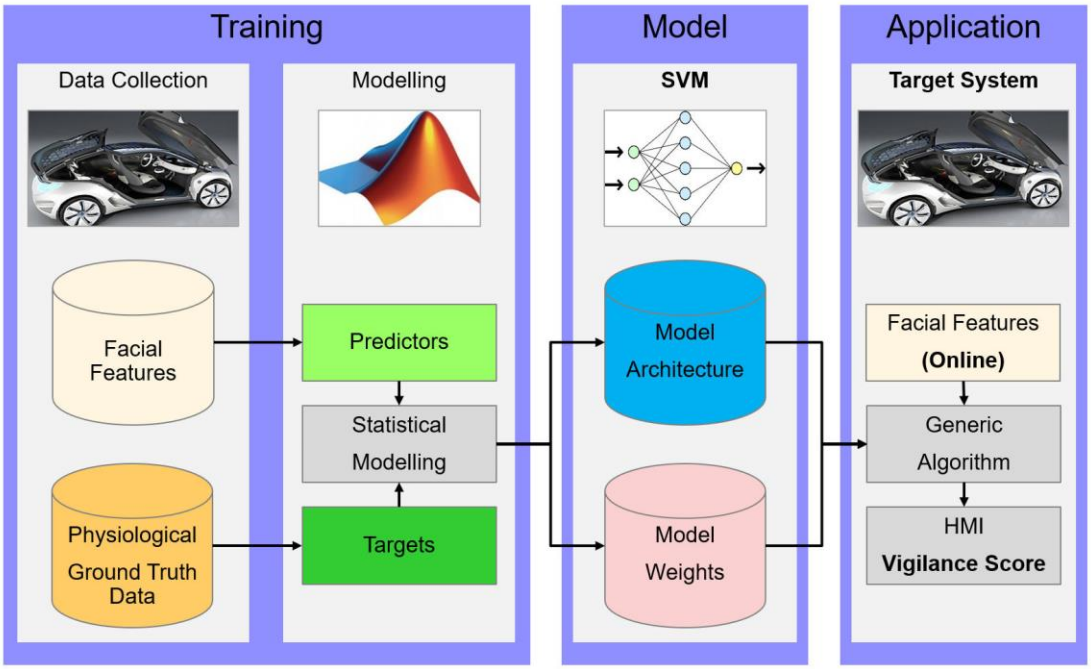
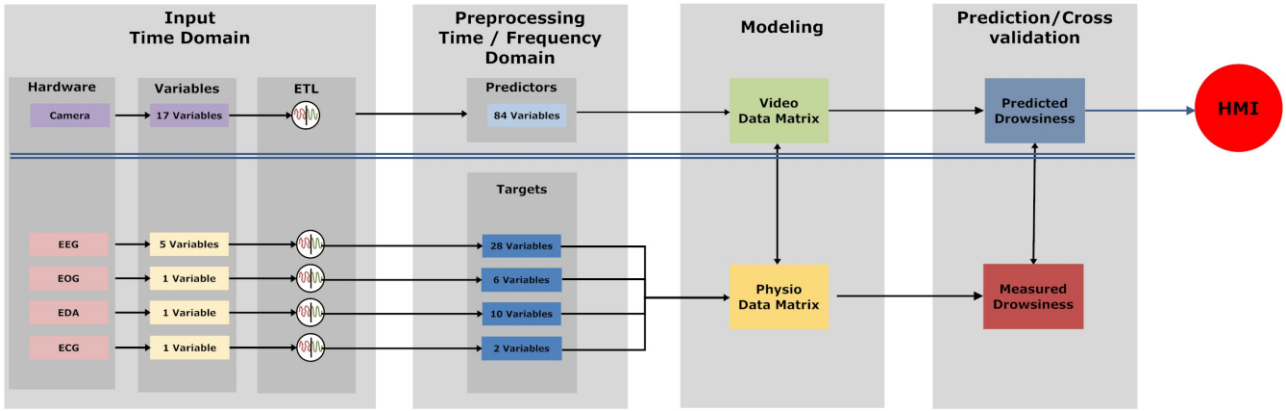



Abbildung 1: Struktur des Modells

[1] Der SVM Algorithmus ist eine Methode des sogenannten überwachten Maschinenlernens. Dessen Kennzeichen ist es, die Modellparameter dadurch zu finden, dass die vorhergesagten Zielvariablen so eng wie möglich den gemessenen Zielvariablen angenähert werden.

Abbildung 2 zeigt den Datenfluss im Modell.




.

Abbildung 2: Datenfluss im Modell

Im Zeitbereich werden 17 verschiedene Augen- und Kopfbewegungsvariablen erzeugt. Die Psyrecon Messtechnik liefert 5 EEG Kanäle, 1 EOG, 1 EDA und 1 EKG Messung. Alle Variablen werden einer Rauschminderungsprozedur durch einen ETL-Prozess (Extract Transfer Load) unterworfen. Mit Ausnahme der im Preprocessing generierten EKG und einiger der EDA Variablen werden alle übrigen Variablen in den Frequenzbereich transformiert. Als Ergebnis liegen anschließen 84 Video Prädiktorvariablen und 46 physiologische Zielvariablen vor, die in die Modellierung eingespeist werden. Die Modellierung berechnet für jede einzelne Zielvariable ein eigenes Modell. Die resultierenden 46 Einzelmodelle werden anschließend zu einem einzelnen vorhergesagten physiologischen Vigilanzscore und einem einzelnen gemessenen physiologischen Vigilanzscore verdichtet [2]. Zu beachten ist, dass das Modell mit einem ausgewählten Satz von Daten trainiert wird, die aber beim Test des Modells nicht verwendet werden dürfen, da dies zu einer artifiziellen Schätzung der Vorhersagekraft des Modells führen könnte. Für den Test werden deshalb dem Modell unbekannte Daten verwendet. Erst wenn es gelingt diese modellfremden Daten vorherzusagen, kann von einem validen Modell gesprochen werden. Im Validierungsschritt werden die gemessenen und die vorhergesagten Scores korrelationsstatistisch bzw. mittels binärer Klassifikation miteinander verglichen. Die Höhe der resultierenden Validierungsindizes ist ein Maß für die Vorhersagekraft des Modells.

### 3.1.2 Validierung des Modells

Abbildung 3 zeigt die Vorhersage von 88 Testdatensätzen (mehrstündige Fahrversuche) auf der Basis von 233 Trainingsdatensätzen (mehrstündige Labor- und Fahrversuche). Jeder einzelne Punkt im Scatterdiagramm repräsentiert eine Messdauer von 10 Sekunden. Die blauen Punkte beziehen sich auf die Trainingsdaten (reference data), die roten auf die Testdaten (new data). Beide (Pearson) Korrelationskoeffizienten sind vergleichsweise hoch. Bemerkenswert ist, dass die Korrelation zwischen gemessenen und vorhergesagten Werten bei den Testdaten etwas höher ist als diejenige der Trainingsdaten. Üblicherweise liegt die Korrelation bei Testdaten deutlich unter derjenigen der Trainingsdaten.

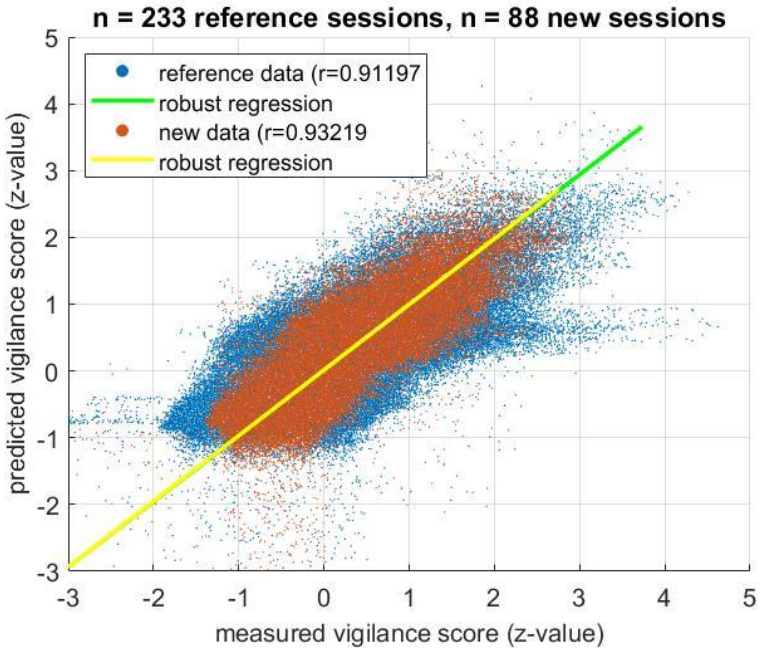



Abbildung 3: Korrelation zwischen gemessenen und vorhergesagten Trainings- und Testdaten

Das Ergebnis zeigt, dass die Vorhersagekraft des Modells sehr hoch ist. Abbildung 4 veranschaulicht exemplarisch die Testdatenkorrelation von r = 0.932 an Hand der Kovariation der zeitlichen Verläufe von gemessenen und vorhergesagten Vigilanzscores.

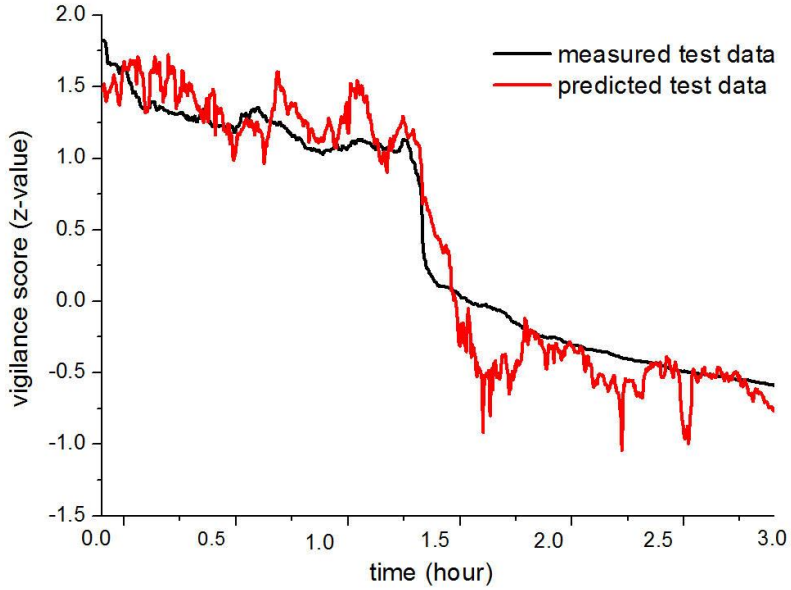



Abbildung 4: Kovariation des zeitlichen Verlaufs von gemessenen und vorhergesagten Vigilanzscores (ausgewählte Testsession)

Die Validitätsmessung mittels Pearson Korrelation geht davon aus, dass die gemessenen und die vorhergesagten Werte mindestens die numerische Qualität einer Intervallskala haben. Reduziert man das Skalenniveau auf eine einfache ja/nein Entscheidung wie es z.B. bei klinischen Diagnoseverfahren der Fall ist (z.B

Diabetes: ja/nein) lassen sich andere Validtätskriterien wie z.B der Sensitivitäts- und der Spezifitätsindex berechnen. Eine hohe Sensitivität bedeutet im vorliegenden Fall, dass ein gemessener Vigilanzstatus, d.h. eine Vigilanz die real existiert, auch mit hoher Wahrscheinlichkeit vorhergesagt wird. Der gemessene Vigilanzstatus fungiert in diesem Fall als sog. Goldstandard. Eine hohe Spezifität bedeutet demgegenüber, dass ein Vigilanzstatus, der in der Messung nur selten auftritt auch entsprechend selten vorhergesagt wird. Die beiden Validitätskriterien verhalten sich somit komplementär zueinander.

Um die Validität der vorhergesagten Vigilanz mittels dieser beiden Kriterien beurteilen zu können, müssen deren Werte diskretisiert werden, d.h. aus einem Kontinuum möglicher Werte muss eine ja/nein Entscheidung gemacht werden. Dazu ist ein Cut Off Kriterium notwendig, dass entscheidet ob eine vorhergesagte Vigilanz als „ja“ oder „nein“ klassifiziert wird. Da der Vigilanzscore z-normiert ist bietet sich als ein einfaches Kriterium ein z-score = 0 an (siehe Abbildung 4). Alle vorhergesagten Werte, die im Bereich z >=0 liegen werden als „ja“ klassifiziert, alle Werte die in einem Bereich z<0 liegen als „nein“. Jetzt lassen sich alle vorhergesagten ja/nein Entscheidungen in Beziehung zum Kontinuum der gemessenen z-scores setzen, d.h. bei jedem gemessenen z-score wird geprüft, ob der zughörige vorhergesagte z-score „ja“ oder „nein“ ist.

Abbildung 5 zeigt das Ergebnis dieser Zuordnung für alle 88 Testdatendsätze.

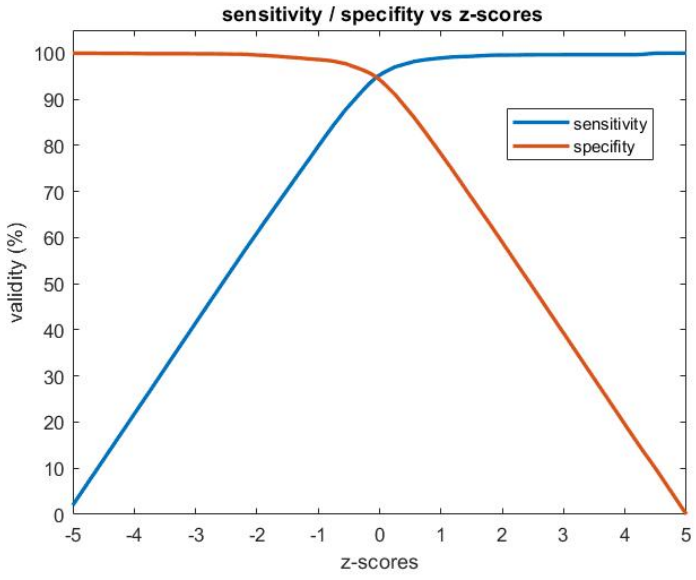



Abbildung 5: Sensitivität und Spezifität der vorhergesagten (binären) Vigilanzscores der N= 88 Testdatensätze in Relation zu ihren gemessenen (kontinuierlichen) Vigilanzscores

Die so gewonnenen Zuordnungen von vorhergesagten und gemessenen Sensitivitäts- und Spezifitätswerten lassen sich in eine ROC-Kurve (Receiver Operation Characteristic) transformieren. An Hand dieser Kurve kann dann der Validitätsindex AUC als Fläche unter der ROC-Kurve berechnet werden [3].

Abbildung 6 zeigt die ROC-Kurve für die Testdaten.

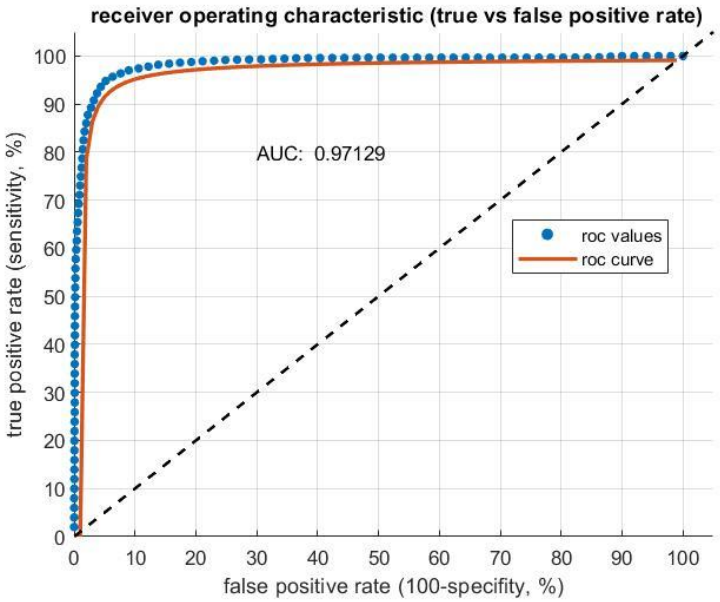



Abbildung 6: ROC.Kurve der N = 88 Testdatensätze

Der AUC beträgt 0.97128. Ein AUC-Wert zwischen 0.9 und 1.0 gilt in der Statistik als exzellent. Der hohe Grad an Vorhersagekraft des Verfahrens, das sich schon in der Korrelation zwischen gemessenen und vorhergesagten Testscores gezeigt hat, wird durch die binäre Validierung bestätigt.

### 3.1.3 Eine statistische Definition der physiologischen Vigilanz

Für die praktische Anwendung des Verfahrens, z.B. als Assistenzsystem im Fahrzeug, ist es wichtig Grenzen zu definieren, innerhalb bzw. außerhalb derer ein definierter Vigilanzstatus angenommen werden kann. Eine theroriegeleitete oder empirisch durch Hinzuziehung weiterer Vigilanzkriterien (z.B. Performance Maße) untermauerte Festlegung dieser Grenzen ist momentan noch nicht möglich. Die Tatsache, dass die Vigilanzscores z-normiert sind, erlaubt jedoch bereits jetzt eine tentative Definition statistischer Grenzen im Sinne einer Arbeitshypothese, sofern diese Grenzen zu interpretierbaren Ergebnissen führen.

Abbildung 7 zeigt den zeitlichen Verlauf von drei Kategorien von gemessener Vigilanz für alle bisherigen Datensätze. Die Verläufe repräsentieren die relative Häufigkeit von z-scores innerhalb definierter Grenzen alle 10 Minuten über einen zeitlichen Verlauf von 4.8 Stunden.

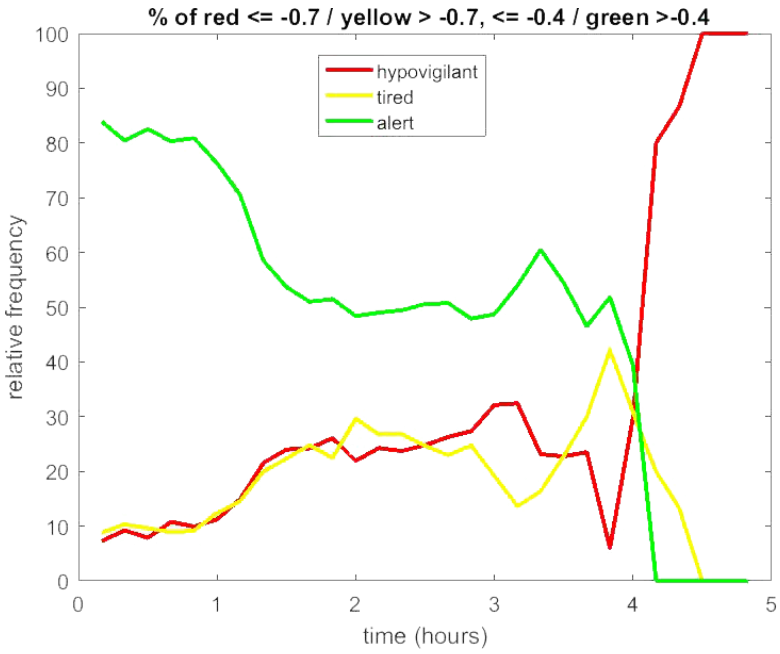



Abbildung 7: Zeitlicher Verlauf der relativen Häufigkeiten der z-scores von drei verschiedene Vigilanzkategorien

Die Daten werden nach drei verschiedenen Grenzkriterien gruppiert:

Vigilanz z-score <=-0.7 gilt als „hypovigilant“
Vigilanz z-score >-0.7 und <= -0.4 gilt als „müde“
Vigilanz z-score > -0.4 gilt als „wach“

Neben dieser Klassifikation sind auch andere Einteilungen der z-scores denkbar. Die Abbildung zeigt jedoch, dass diese Art der Einteilung plausibel ist. Die grüne Kurve („wach“) startet bei einem hohen Wert von 85 % der Fälle und fällt dann 4.1 Stunden auf einen Wert von 0 %. Dieser Verlauf für die Kategorie „wach“ ist leicht vorstellbar wenn ein Fahrer ununterbrochen über einen Zeitraum von mehr als vier Stunden fährt. Demgegenüber startet die Kategorie „müde“ gemeinsam mit der Kategorie „hypovigilant“ bei einem Wert von ca. 10% relativer Häufigkeit. Nach ca. 3.8 Stunden fällt die relative Häufigkeit der Kategorie „müde“ rapide ab, während die relative Häufigkeit der Kategorie „hypovigilant“ steil ansteigt, um dann die beiden anderen Kategorien völlig zu überschatten. Für weitere Experimente im Bereich der physiologischen Vigilanz kann aus diesem Befund arbeitshypothetisch geschlossen werden, dass ein hypovigilanter Zustand nach ca. 3.8 Stunden bei einer moderaten kognitiven Beanspruchung, entsprechend einer realen Autofahrt oder einer vergleichbaren Laborsimulation, erreicht wird. Mit anderen Worten: Experiment, die zum Ziel haben den Zustand der physiologischen Hypovigilanz unter praxisnahen kognitiven Belastungsbedingungen zu erforschen, sollten eine Länge von 4-5 Stunden nicht unterschreiten.

## 3.2 Individualspezifisches Modell

Um herauszufinden welches mathematische Modell am besten für die individualspezifische Modellierung geeignet ist, werden fünf verschiedene Algorithmen des überwachten Maschinenlernens miteinander verglichen. Der Vergleich erfolgt an Hand des Datensatzes von N = 87 Versuchsfahrten, die eine Teilmenge des gesamten Datenpools sind. Die Modellierung von Trainings- und Testdaten, die nach Zufall ausgesucht werden, erfolgt jeweils fünf mal hintereinander.

Folgende Algorthmen werden getestet[2]:

1.) ANN (Artificial Neural Network)
2.) CHAID (Qhi-square Automatic Interaction Detectors)
3.) C&RT (Classification & Regression Trees)
4.) MR (Multiple Regression)
5.) SVM (Support Vector Machine)

Abbildung 8 zeigt die Mittelwerte und Standardabweichung der Trainings- und Testkorrelationen der fünf Algorithmen.

Die Abbildung zeigt, dass der SVM-Algorithmus die höchste Vorhersagekraft in Relation zu den übrigen Algorithmen aufweist. Für die geplante spätere Realisierung des individualspezifischen Modells wird deshalb der SVM-Algorithmus verwendet.

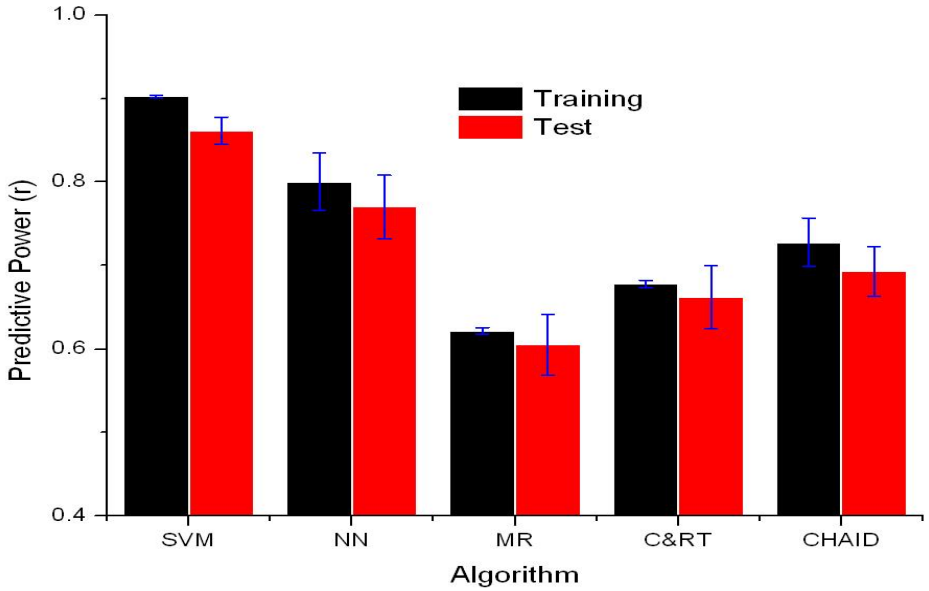



Abbildung 8: Prädiktionskraft von fünf verschiedenen Algorithmen des überwachten Maschinenlernens

# 4 Zusammenfassung und Fazit

Die aktuelle Datenbasis des generischen Modells umfasst N=321 Labor- und Fahrversuche, deren Dauer zwischen 1.5 und 4.5 Stunden liegt. 60% der Daten werden auf der Straße gemessen. Die übrigen Daten werden im Labor erhoben. Der Altersrange von Probanden beiderlei Geschlechts liegt zwischen 18-65 Jahren.

In jeder experimentellen Sitzung werden zwei verschiedene Typen von Variablen gemessen:

1.) Physiologische Daten (Electroenzephalogramm – EEG, Okulomotorische Aktivität – EOG, Elektrodermale Aktivität – EDA, Herz Aktivität – EKG)
2.) Video Kopf- und Augenbewegungen

Im Modell sind die physiologischen Variablen als Zielgrößen definiert. Die Videovariablen sind die Prädiktoren. Vor der eigentlichen Modellierung werden die Daten vorverarbeitet und resultieren in zwei getrennten Datensätzen für die Zielvariablen und die Prädiktoren. In der nachfolgenden Modellierung mittels

[2] Andere Algorithmen wie Bayes Netze, Kohonen Netze, Logistic Regression, etc., werden ebenfalls getestet. Es zeigt sich jedoch, dass diese Algothmen entweder in Vortests sehr schlecht abschneiden (und deshalb verworfen werden) oder ein dichotomes Datenniveau verlangen (z.B. Bayes Netze).

Support Vector Machines (SVM) werden sie mathematisch miteinander kombiniert. Die Vorhersage der Vigilanz erfolgt ausschließlich auf der Basis der Videovariablen.

Die Video Rohdaten sind 17 verschiedene Augen- und Kopfbewegungsvariablen. Die Psyrecon Rohdaten sind 5 EEG Kanäle, 1 EOG, 1 EDA und 1 EKG Messung. Der größte Teil der Rohdaten werden im Zuge des Preprocessings in den Frequenzbereich transformiert. Als Ergebnis liegen 84 Video Prädiktorvariablen und 46 physiologische Zielvariablen vor, die in die Modellierung eingespeist werden. Die Modellierung berechnet für jede einzelne Zielvariable ein eigenes Modell. Die resultierenden 46 Einzelmodelle werden anschließend zu einem einzelnen vorhergesagten physiologischen Vigilanzscore und einem einzelnen gemessenen physiologischen Vigilanzscore verdichtet. Zu beachten ist, dass das Modell mit einem ausgewählten Satz von Daten trainiert wird, die aber beim Test des Modells nicht verwendet werden dürfen, da dies zu einer artifiziellen Schätzung der Vorhersagekraft des Modells führen könnte. Für den Test werden deshalb dem Modell unbekannte Daten verwendet. Erst wenn es gelingt diese modellfremden Daten vorherzusagen, kann von einem validen Modell gesprochen werden. Im Validierungsschritt werden die gemessenen und die vorhergesagten Scores korrelationsstatistisch bzw. mittels binärer Klassifikation miteinander verglichen. Die Höhe der resultierenden Validierungsindizes ist ein Maß für die Vorhersagekraft des Modells.

Für die korrelationsstatistische Validierung des Modells werden 88 modellfremde Testdatensätzen (mehrstündige Fahrversuche) auf der Basis von 233 Trainingsdatensätzen (mehrstündige Labor- und Fahrversuche) vorhergesagt. Die Korrelationen zwischen gemessenen und vorhergesagten Vigilanzscores sind sowohl für die Trainings als auch die Testdaten vergleichsweise hoch ($r = 0.911$, $r = 0.932$). Bemerkenswert ist, dass die Korrelation zwischen gemessenen und vorhergesagten Werten bei den Testdaten etwas höher ist als diejenige der Trainingsdaten. Üblicherweise liegt die Korrelation bei Testdaten deutlich unter derjenigen der Trainingsdaten. Die Validierung des Modells mittels der Receiver Operating Characteristik (ROC) bestätigt die hohe Vorhersagekraft des Modells. Sie ergibt einen Validitätsindex von AUC = 0.971, der in der Statistik als exzellent gilt.

Für die praktische Anwendung des Verfahrens, z.B. als Assistenzsystem im Fahrzeug, ist es wichtig Grenzen zu definieren, innerhalb bzw. außerhalb derer ein definierter Vigilanzstatus angenommen werden kann. Eine theroriegeleitete oder empirisch durch Hinzuziehung weiterer Vigilanzkriterien (z.B. Performance Maße) untermauerte Festlegung dieser Grenzen ist momentan noch nicht möglich. Die Tatsache, dass die Vigilanzscores z-normiert sind, erlaubt jedoch eine tentative Definition statistischer Grenzen im Sinne einer Arbeitshypothese. Die Daten werden nach drei verschiedenen Grenzkriterien gruppiert:

Vigilanz z-score <=-0.7 gilt als „hypovigilant“
Vigilanz z-score >-0.7 und <= -0.4 gilt als „müde“
Vigilanz z-score > -0.4 gilt als „wach“

Der mehrstündige zeitliche Verlauf der relativen Häufigkeiten dieser drei Gruppen zeigen, dass die Kategorie „wach“ bei einem Wert von ca. 85 % startet, um nach 4.1 Stunden auf einen Wert von 0 % abzusinken. Dieser Verlauf für die Kategorie „wach“ ist leicht vorstellbar wenn ein Fahrer ununterbrochen über einen Zeitraum von mehr als vier Stunden fährt. Demgegenüber startet die Kategorie „müde“ gemeinsam mit der Kategorie „hypovigilant“ bei einem Wert von ca. 10% relativer Häufigkeit. Nach ca. 3.8 Stunden fällt die relative Häufigkeit der Kategorie „müde“ rapide ab, während die relative Häufigkeit der Kategorie „hypovigilant“ steil ansteigt, um dann die beiden anderen Kategorien völlig zu überschatten. Für weitere Experimente im Bereich der physiologischen Vigilanz kann aus diesem Befund arbeitshypothetisch geschlossen werden, dass ein hypovigilanter Zustand nach ca. 3.8 Stunden bei einer moderaten kognitiven Beanspruchung, entsprechend einer realen Autofahrt oder einer vergleichbaren Laborsimulation, erreicht wird. Mit anderen Worten: Experiment, die zum Ziel haben den Zustand der physiologischen Hypovigilanz unter praxisnahen kognitiven Belastungsbedingungen zu erforschen, sollten eine Länge von 4-5 Stunden nicht unterschreiten.

Um beim individualspezifischen Modell herauszufinden welches mathematische Modell am besten geeignet ist, werden fünf verschiedene Algorithmen des überwachten Maschinenlernens miteinander verglichen. Der Vergleich erfolgt an Hand des Datensatzes von N = 87 Versuchsfahrten, die eine Teilmenge des gesamten Datenpools sind.

Folgende Algorthmen werden getestet[3]:

1.) ANN (Artificial Neural Network)
2.) CHAID (Qhi-square Automatic Interaction Detectors)
3.) C&RT (Classification & Regression Trees)
4.) MR (Multiple Regression)
5.) SVM (Support Vector Machine)

Die Ergebnisse zeigen, dass der SVM-Algorithmus die höchste Vorhersagekraft in Relation zu den übrigen Algorithmen aufweist. Für die geplante spätere Realisierung des individualspezifischen Modells wird deshalb der SVM-Algorithmus verwendet.

# 5 Literatur

[1]Vapnik, V. (1998). Statistical Learning Theory, New York: Wiley.

[2] Hotelling, H. (1936). "Relations between two sets of variates". Biometrika, **28** (3–4), 321–377.

[3] Fawcett, T (2004). ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. In: Pattern Recognition Letters. 31, Nr. 8, S. 1–38.

[3] Andere Algorithmen wie Bayes Netze, Kohonen Netze, Logistic Regression, etc., werden ebenfalls getestet. Es zeigt sich jedoch, dass diese Algothmen entweder in Vortests sehr schlecht abschneiden (und deshalb verworfen werden) oder ein dichotomes Datenniveau verlangen (z.B. Bayes Netze).