

Probability of Detection (POD)-based metric for evaluation of Classifiers used in Driving Behavior Prediction

Daniel Adofo Ameyaw¹, Qi Deng², and Dirk Söffker³

^{1,2,3} *Chair of Dynamics and Control
University of Duisburg-Essen
Lotharstraße 1-21, 47057 Duisburg
Germany
daniel.adofo-ameyaw@uni-due.de
qi.deng@uni-due.de
soeffker@uni-due.de*

ABSTRACT

Classifiers are functional tools/algorithms that implement classifications and are widely used in science and technology for state of health estimation, diagnosis systems, and situation/intention recognition of human operators. Certification of these classifiers plays a crucial role in their selection for a specific task. Current certification approaches utilize the Receiver Operator Curve (ROC) as a standard tool that provides graphically the performance of classifiers. Beside the ratio of Detection Rate and False Alarm Rate (combined as ROC), other properties related to process parameters are not considered. In this research, a new evaluation method based on the Probability of Detection (POD) reliability measure is developed discussing the effect of further process parameters on the classification results. Probability of Detection (POD) serves as a performance measure for quantifying the reliability of conventional Nondestructive Testing (NDT) procedures and Structural Health Monitoring (SHM) systems. The approach considers statistical variability of sensor-based measurements. In this publication for the first time the signal-response and the binary (hit/miss) approaches are implemented in combination with a process parameter. As example, the prediction time of driving behavior classification is used as process parameter. The signal-response approach is applied to compare Fuzzy Logic-Hidden Markov Model (FL-HMM), Artificial Neural Network (ANN), and Support Vector Machine (SVM) with respect to the reliability of the prediction time for driver behavior. The hit/miss approach is also applied on FL-HMM as example for predicting an upcoming driving maneuver. To account for data uncertainty and variability, confidence bounds are established. A

typical and useful criteria for detection at a 90 % probability of detection level with 95 % confidence level is successfully implemented as a new reliability measure and certification standard for classifiers. In this article a new approach is established permitting a new evaluation measure to Machine Learning approaches used as classifiers. The proposed approach introduces a POD-based measure for comparison of binary classifiers.

1. INTRODUCTION

Probability of detection has been implemented in the field of NDT for decades and lately SHM systems. The POD is a probabilistic method to quantify the reliability of an NDT/SHM procedure taking into account statistical variability of sensor and measurements properties (Department of Defense, 2009). The time and cost involved in POD has given rise to Model-assisted POD (MAPOD) to improve the effectiveness of POD models with little or no specimen testing by utilizing model generated data (Knopp et. al., 2007). However numerical efforts and computational time difficulties have to be solved for convenient application in practice. The POD evaluation uses the so-called POD curve. The POD curve is constructed by plotting the accrual of flaws detected against the varying parameter or produce a response over a specified threshold (Georgiou, 2007). The POD approach is implemented in predicting a driver's intention.

Predicting drivers intention is useful for ensuring driving safety in autonomous and/or assisted driving. An important tool used in these predictions are classifiers. These predictions are used in Driver Assistance Systems to assist drivers. A key idea is to establish models by learning from the given driving behaviors and subsequently predict the decisions and behaviors. Research in this field is concerned with new methods to realize and improve driving behavior prediction how-

Daniel A. Ameyaw et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ever reliability of the proposed approaches are usually not given much attention. The Receiver Operator Curve (ROC) is often used by many authors to assess the performance of classifiers. Beside the ratio of Detection Rate and False Alarm Rate, other properties related to process parameters are not considered. This limitation in verifying the effect of process parameters on the classifier performance is addressed in this paper. The POD reliability metric overcomes this limitation. In this work, three driving maneuvers are considered: lane changing to right S_1 , lane keeping S_2 , and lane changing to left S_3 . Input variables affecting driver's decisions are measured through a professional driving simulator SCANerTM studio. The related inputs and outputs (lables) (Deng & Söffker, 2018) are used.

The article is organized as follows: in section 2 the classifiers used are briefly introduced, followed by the newly developed POD reliability measure and its application to driving maneuver prediction in section 3. Comparison is made between different classifiers in section 4 and intention recognition in section 5 using the proposed approaches and subsequently the conclusion.

2. BINARY CLASSIFIERS

In this work, three classifiers (FL-HMM, SVM, and ANN) are used and compared with respect to their ability for prediction using the new measure introduced. A brief introduction and review of these classifiers are summarized.

2.1. Fuzzy Logic-based Hidden Markov Models

The authors in (Deng & Söffker, 2018) developed a new approach, Fuzzy Logic-based Hidden Markov Models (FL-HMM). The FL approach will be used for distinction of driving scenes into very safe, safe, and dangerous driving scenarios. Afterwards, a corresponding standard HMM will be trained for comparison of each driving scenario. Three different driving behaviors: left/right lane change and lane keeping, are modeled as hidden states for these HMMs. An HMM (Rabiner, 1989) describes the relationship between two stochastic processes: one consists of a set of unobserved (hidden) states $S = \{S_1, S_2, \dots, S_N\}$, with N as the number of hidden state which cannot be measured directly. In this research $N = 3$. The other stochastic process is denoted by a set of M observable symbols $V = \{V_1, V_2, \dots, V_M\}$. The hidden state and observation symbol at time t are defined as Q_t and O_t respectively. Thus a hidden state sequence is $Q = \{Q_1, Q_2, \dots, Q_T\}$ and an observation sequence is $O = \{O_1, O_2, \dots, O_T\}$, where T is the length of the sequence. In a given observation sequence O and its corresponding hidden state sequence Q , the HMM parameters can be computed and adjusted to best fit both sequences using Baum-Welch algorithm. Based on the saved HMM, the most probable sequence of driving behaviors, which has the highest probability, are calculated using Viterbi algorithm. That means, in each step, using HMM

the probabilities of each hidden states $\{P_{S_1}, P_{S_2}, P_{S_3}\}$ can be calculated separately.

2.2. Support Vector Machine

Support Vector Machine (SVM) developed by Vapnik (1979) is a widely applied classification technique (Cortes & Vapnik, 1989). In this contribution, a two-class classifier as a supervised machine learning method is used to distinguish different classes of driving behaviors. In SVM the different driving behaviors are classified by transforming the observation variables into an observation vector and thus generating a distribution in a high dimensional space. Each observation vector is assigned to corresponding classes based on training data. The driving behaviors are separated using a hyperplane. The observation vectors of the different classes are on the different sides of the hyperplane. The process of SVM learning involves finding an optimal hyperplane between observation vectors of different classes to generate a maximal geometric margin. However, the SVM was originally designed only for two classes. The driving behaviors prediction model is a multiclass problem. For this reason often a multiclass model is used. Most popular solutions are shown in (Schölkopf & Smola, 2002) like one-against-all and one-against-one. Several binary classifiers of SVM are required to analyze multiclass problems. In this study default Matlab code and the one-against-one approach is used to establish the model.

2.3. Artificial neural network

Artificial Neural Network (ANN) also denoted as Neural Network (NN), is a computational model used in machine learning. It imitates a biological neural network. Typically, ANN contains many layers. The first and last layer represent input and output respectively. It is a computational model similar to the animal's central nervous system and applied in fields of human behavior problems. For example, Neural Network (NN) models have been used in predicting the acceleration distribution for car following on highways (Mannering & Bhat, 2014), lane changing prediction (Xiong, 2014), among others. The default Matlab NN hyperparameter is used.

3. POD ASSESSMENT OF BINARY CLASSIFIERS

Probability of detection is an established certification tool used to access the reliability of NDT/SHM measurement procedures. Data used in producing POD curves are categorized by the main POD controlling factors/variables. These factors/variables are either discrete or continuous and can be classified as

1. Hit/miss: produce binary statement or qualitative information about the existence of a target and
2. Signal-response: systems which also provide some quantitative measure of target.

Both approaches are adapted and implemented in comparing different classifiers and predicting an impending maneuver.

3.1. Signal-response approach to POD

The signal response approach is used when there exist a linear relationship between a monotonous increasing function and a monotonous increasing parameter. In the derivation of the signal-response POD curve, a regression analysis of the data gathered has to be realized (Fig. 4) (Department of Defense, 2009), (Annis, 2017), (Gandossi & Annis, 2010). The regression equation for a line of best fit to a given data set is given by

$$y = b + mx, \quad (1)$$

where m is the slope and b the intercept. The Wald method is used to construct the confidence bounds. Here the 95 % Wald confidence bounds on y is constructed by

$$y_{(a=0.95)} = y + 1.645\tau_y, \quad (2)$$

where 1.645 is the z -score of 0.95 for a one-tailed standard normal distribution and τ_y the standard deviation of the regression line. The Delta method is a statistical technique used to transition from regression line to POD curve (Department of Defense, 2009). The confidence bounds are computed using the covariance matrix for the mean and standard deviation POD parameters μ and σ respectively. To estimate the entries, the covariance matrix for parameters and distribution around the regression line needs to be determined. This is done using the Fisher's information matrix I . The information matrix is derived by computing the maximum likelihood function f of the standardized deviation z of the regression line values. The entries of the information matrix are calculated by the partial differential of the logarithm of the function f using the parameters of $\Theta(m, b, \tau)$ of the regression line.

From

$$z_i = \frac{(y_i - (b + mx_i))}{\tau} \quad (3)$$

and

$$f_i = \prod_{i=1}^n \frac{1}{2\pi} e^{-\frac{1}{2}(z_i)^2} \quad (4)$$

the information matrix I can be computed as

$$I_{ij} = -E\left(\frac{\partial^2}{\partial\Theta_i\partial\Theta_j} \log(f)\right). \quad (5)$$

The inverse of the information matrix yields ϕ as

$$\phi = I^{-1} = \begin{bmatrix} \sigma_b^2 & \sigma_b\sigma_m & \sigma_b\sigma_\tau \\ \sigma_m\sigma_b & \sigma_m^2 & \sigma_m\sigma_\tau \\ \sigma_\tau\sigma_b & \sigma_\tau\sigma_m & \sigma_\tau^2 \end{bmatrix}. \quad (6)$$

The mean μ and standard deviation σ of the POD curve are

calculated by $\mu = \frac{c-b}{m}$, where c is the decision threshold and $\sigma = \frac{\tau}{m}$. The cumulative distribution Φ is calculated as

$$\Phi(\mu, \sigma) = \frac{1}{2} \left[1 + \operatorname{erf} \frac{x-\mu}{\sqrt{2}\sigma} \right]. \quad (7)$$

The POD function is derived as

$$POD(a) = \Phi \left[\frac{a-\mu}{\sigma} \right]. \quad (8)$$

Using this formula, the POD-curve can be set up for varying parameters. For this example, the varying parameter is the time t . The intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$ are statistically estimated from the observations.

3.2. Hit/Miss approach to POD

An efficient implementation of the binary data is to posit an underlying mathematical relation between POD and parameter and consequently model the probability distribution (Department of Defense, 2009). The use of ordinary linear regression is inaccurate since the data are not continuous but discrete and bounded. Generalized Linear Models (GLM) overcome this challenge by linking the binary response to the explanatory variables through the probability of either outcome, which does vary continuously from 0 to 1 (Nelder & Wedderburn, 1972) (Department of Defense, 2009). The GLM attains this through

1. A random component specifying the conditional distribution of the response variables, Y_i (for the i -th of n independently sample observations)
2. A linear predictor that is a function of regressors
3. A smooth and invertible linearizing link function $g(y)$, which transforms the expectation of the response variables $P_i \equiv E(Y_i)$ to the linear predictor.

The transformed probability can then be modeled as an ordinary polynomial function, linear in the explanatory variables. The POD can be generated from the GLM as explained in the case of linear regression. The commonly used GLM in POD are the log, logit, Probit, loglog and weibull. Depending on the data distribution a model may be appropriate compared to the other. One criteria used is to select GLM with least deviance.

4. COMPARISON OF FL-HMM, SVM, AND ANN BY THE SIGNAL-RESPONSE APPROACH

A driving simulator *SCANeRTM* studio is applied to perform driving simulation. The simulator is equipped with five monitors, base-fixed driver seat, steering wheel, and pedals. The three rear mirrors, which are essential to decide the lane change are displayed on the corresponding positions of the monitors. The driving setting uses a highway scenario with four lanes of two directions and simulated traffic environment.

To evaluate the predicted performance, a method stimulated by (Zhao et al., 2017a) (Zhao et al., 2017b) is utilized in this paper. In (Zhao et al., 2017a) the authors discussed the impact of surrounding cyclists on the ego driver's driving behavior. The process parameter discussed is the Detection Rate (DR) as a function of the distance to upcoming collision. This is used for evaluating the classifiers. It is noteworthy that the DR values are calculated every two meters from the start point respectively. Here, the start point is defined as the first data point the scenario begins. The DR values and their corresponding distances from the start point to the points where the DR values reaches 100 % can be calculated. The shorter the distance reaches the higher DR implies better performance. In (Zhao et. al., 2017b) the authors analyzed human driver behaviors in interaction with roundabouts based on SVM using steering angle and steering angle velocity. To evaluate the model performance, the exact known driving route for each driver is divided into 10 parts evenly by 11 points, which are defined as recognition sites. For each recognition site a corresponding DR value is calculated. Stimulated by these two references a new approach is developed and used. Similarly, each lane change behavior is defined as a separate event. From 6 seconds before to the time of actual lane change a DR value will be calculated for performance evaluation. The time interval is divided into 120 time points, i.e. every 0.05 s. These time points are defined as "recognition time points". The DR value is calculated based on True Positive (TP) as well as False Negative (FN) numbers (Mukhopadhyay et. al., 2014). As explanation a multiclass confusion matrix is illustrated (Fig. 1) to describe the parameters for S_1 (lane changing to the right), where TP (True positive) is defined as the number of samples when the estimated maneuver is S_1 (positive) and the actual one is also positive.

Multiclass Confusion Matrix		Predicted		
		S_1	S_2	S_3
Actual	S_1	TP	FN	FN
	S_2	FP	TN	TN
	S_3	FP	TN	TN

Figure 1. Explanation of multiclass confusion matrix (S_1)

Contrastively FN denotes the number of events when the estimated maneuver is not S_1 (negative) and the actual value is positive. Likewise computations can be made for S_2 and S_3 .

The value of DR can be calculated by

$$DR = \frac{TP}{TP + FN}. \quad (9)$$

The DR values are calculated by the three classifiers (FL-HMM, SVM, and ANN) for each recognition time step. Based on the computed values, the signal response method is utilized in this section to compare these 3 binary classifiers as a new reliability standard. The aim is to produce a POD vs parameter (here: time) that is representative of probability distribution of lane change scenarios (Fig. 2). Four models comprising combinations of logarithmic and cartesian scales (Fig. 3) are established to ascertain model with

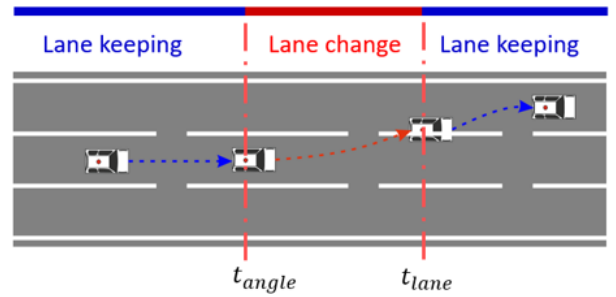


Figure 2. Lane change scenarios

- 1 Linearity of the parameters: $E(y_i|X) = x_i\beta$, where x_i is the i -th row of X and
- 2 Uniform variance: $var(y_i|X) = \sigma^2, i = 1, 2, 3, \dots, n$.

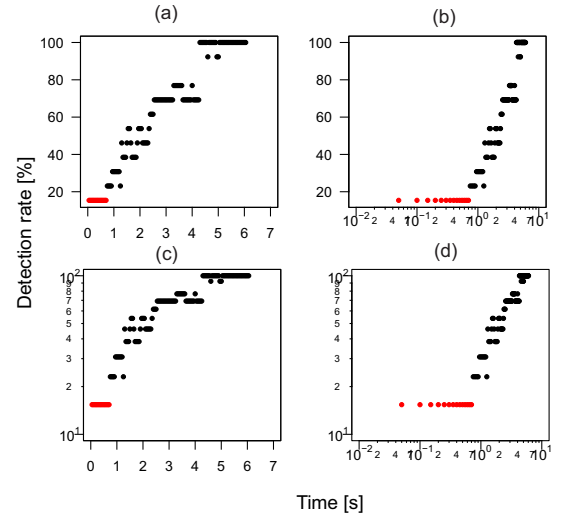


Figure 3. Four possible models

The model that fits linearity and uniform variance criteria best is Fig. 3b and hence selected. Regression analysis is implemented on the selected model and the decision threshold (response value below which the signal is considered as noise) is

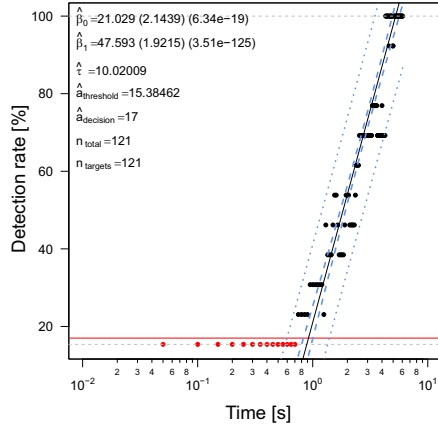


Figure 4. Regression analysis

constructed as illustrated in Fig. 4. The POD curve for each classifier and the actual maneuver is constructed for both right (Fig. 5) and left lane change (Fig. 6).

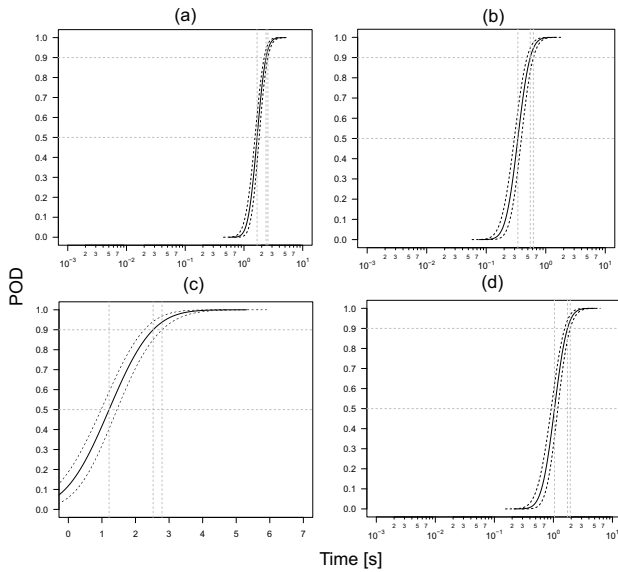


Figure 5. a: ANN b: FL-HMM c: Actual d: SVM POD for right lane change

The POD distribution is different for all classifiers. The 90/95 reliability values for the right and left lane change POD estimations are illustrated in Table 1. The least time values represent best results. This is because the algorithm is able to predict the lane change within the shortest possible time.

For lane change to right, the 90/95 POD value for the actual lane change occurs at 2.79 s but the FL-HMM is able to predict at 0.6199 s, ANN at 2.556 s and SVM at 1.93 s. For lane change to left, the 90/95 POD value for the actual lane change occurs at 3.239 s but the FL-HMM is able to predict at 1.302 s, ANN at 3.379 s and SVM at 2.888 s. It becomes evident from the analysis that FL-HMM has best prediction results

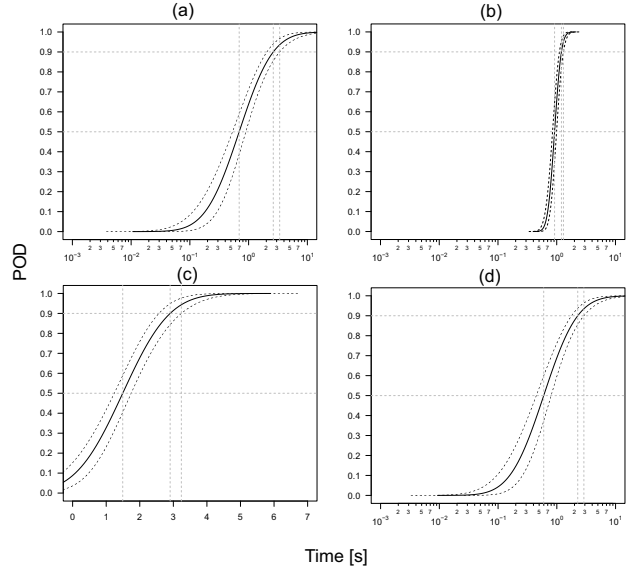


Figure 6. a: ANN b: FL-HMM c: Actual d: SVM POD for left lane change

Table 1. Lane change POD

Classifier	Right lane POD [s]	Left lane POD [s]
Actual	2.790	3.239
FL-HMM	0.620	1.302
ANN	2.556	3.379
SVM	1.930	2.888

while ANN produce the worst results in both the estimation for left and right lane change.

The introduced approach permits a new POD-based comparison method for binary classifiers based on their reliability of prediction. The approach also considers a process parameter in the evaluation procedure.

5. HIT/MISS APPROACH TO FL-HMM INTENSION RECOGNITION

This section briefly details a new reliability evaluation to FL-HMM algorithm. The FL-HMM algorithm allocates probabilities to indicate if the impending maneuver is a right, left or lane keeping. The probabilities are assigned for each time step 6 s before the actual maneuver occurs on the 6.05 s mark. Each step size is 0.05 s culminating in a total of 121 step sizes. Each maneuver has different probability distribution. To standardize all maneuvers the hit/miss approach is utilized. The sum of assigned probabilities ($S_1 + S_2 + S_3$) at every time step equals 100 %. It can be concluded safely that a maneuver with a 50 % or more assigned probability refers to that maneuver. To implement the proposed approach, a new criteria is defined. For probability $P \geq 0.5$ is assigned a hit value {1} whilst $P < 0.5$ is assigned a miss {0}.

In this studies, the Weibull (cloglog) is implemented to map $-\infty < x < \infty$ to $0 < y < 1$. It is selected for this specific analysis because it has the least data deviance in comparison to the other known POD link functions. The Weibull function is expressed as

$$f(X) = g(y) = \log(-\log(1 - p)),$$

where $f(X)$ is an algebraic function with linearized parameters and p the probability. The probability of detection as a function of time for the Weibull model is

$$POD(t) = 1 - \exp(-\exp(f(X))).$$

Using the Weibull, surface contours are constructed to ascertain the most plausible GLM (Fig. 7).

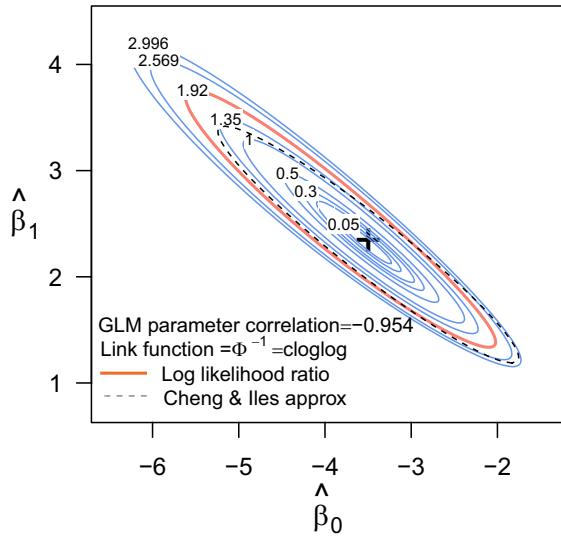


Figure 7. Log likelihood surface contour

The likelihood ratio test is used to assess goodness of fit. The log likelihood ratio contour encloses all β_0, β_1 pairs that are plausibly supported by the data. The confidence bounds are constructed on the surface contours using the Cheng & Iles approximation (Cheng & Iles, 1983). Using the maximum likelihood estimation (MLE) method, the GLM values for the intercept and slope are $\beta_0 = -3.4978$ and $\beta_1 = 2.3033$ at angle of -0.9802 rad. With these values, a GLM model and confidence bounds fitting the data is constructed (Fig. 8).

From the fitted GLM model, the POD curve is generated (Fig. 9). Confidence bounds are plotted accompanying POD curves to highlight specific points. The high safety standards of aerospace industries requires use of the 90/95 certification criteria, which is adapted in this work. From the POD, the 90/95 reliability value is 2.305 s. This implies with the actual maneuver occurring at 6.05 s, the FL-HMM algorithm is able to predict 3.745 s (6.05-2.305) before with a probability of 90 % at a reliability level of 95 %. This introduces a new

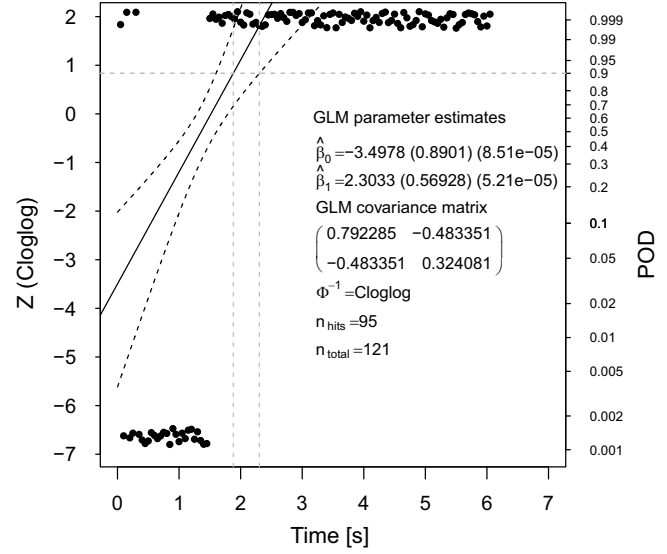


Figure 8. GLM and confidence bounds

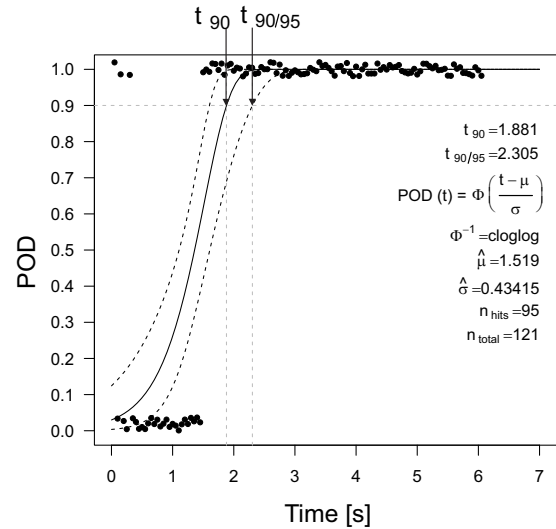


Figure 9. POD model

evaluation metric and insight into the driver maneuver prediction capabilities of the FL-HMM classifier and classifiers in general.

6. CONCLUSION

In this contribution a new approach and insight into the certification of classifiers is presented. This is needed because often additional process parameters affect the classification results but are not considered within known measures like the ROC curve. The approach is derived from the POD evaluation metric and allows comparison of different binary classifiers. The proposed approach is demonstrated on experimental data from real human driving behaviors (taken from driving simulator). The proposed signal-response analysis is

used to compare different classifiers and the results indicate FL-HMM has better estimation capabilities of driving scenarios compared to ANN and SVM for this example task. The hit/miss method is implemented on FL-HMM maneuver prediction and aids in estimating an impending maneuver with the 90/95 certification criteria. The novel approach introduced serves as a new reliability measure for classifiers.

ACKNOWLEDGEMENT

This work is partly supported by Deutscher Akademischer Austauschdienst (DAAD) through a Ph.D. fellowship for the first author performing his Ph.D. at the Chair of Dynamics and Control, U DuE, Germany.

REFERENCE

Deng, Q., & Söffker, D. (2018). *Improved driving behaviors prediction based on Fuzzy Logic-Hidden Markov Model (FL-HMM)*. IEEE Intelligent Vehicles Symposium (IV), pp. 2003-2008.

Rabiner, L. R. (1989). *A tutorial on Hidden Markov Models and selected applications in speech recognition*. Proceedings of the IEEE, vol. 77, pp. 257-286.

Cortes, C., & Vapnik, V. (1995). *Support-vector networks*. Machine Learning, pp. 273-297.

Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. The MIT Press, London, England, Chapter 1, pp. 15-20.

Mannering, F. L., & Bhat, C. R. (2014). *Analytic methods in accident research: Methodological frontier and future directions*. Analytic Methods in Accident Research, vol. 1, pp. 1-22.

Xiong, G., Li, Y., Wang, S., Li, X., & Liu, P. (2014). *HMM and HSS Based Social Behavior of Intelligent Vehicles for Freeway Entrance Ramp*. International Journal of Control and Automation, vol. 7, no. 10, pp. 79-90.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2014). *A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I*. IEEE Transactions on Evolutionary Computation, vol. 18, no. 1, pp. 4-19.

Zhao, M., Käthner, D., Söffker, D., Jipp, M., & Lemmer, K. (2017). *Impact of Surrounding Cyclists on Car Driver Behavior Recognition at Roundabouts*. IEEE Intelligent Vehicles Symposium, pp. 1-6.

Zhao, M., Kathner, D., Jipp, M., Söffker, D., & Lemmer, K. (2017). *Modeling Driver Behavior at Roundabouts: Results from a Field Study*. IEEE Intelligent Vehicles Symposium, pp. 908-913.

Department of Defense (2009). *Department of Defense Handbook, Nondestructive Evaluation System Reliability Assessment, MIL-HDBK-1823A*. Department of Defense.

Georgiou, G. A. (2007). *PoD curves, their derivation, applications and limitations*. Insight-Non-Destructive Testing and Condition Monitoring, 49(7), 409-414.

Annis, C. (2017). *Statistical best-practices for building Probability of Detection (POD) models*. R package mh1823. <http://StatisticalEngineering.com/mh1823/>

Gandossi, L. & Annis, C. (2010). *Probability of detection curves: Statistical best-practices*. ENIQ report, 41. Office for official Publications of the European Communities, Luxembourg.

Nelder, J. A., & Wedderburn, R. W. (1972). *Generalized linear models*. Journal of the Royal Statistical Society: Series A (General), 135(3), 370-384.

Knopp, J. S., Aldrin, J. C., Lindgren, E., & Annis, C. (2007). *Investigation of a model-assisted approach to probability of detection evaluation*. AIP Conference Proceedings 894(1), 1775-1782.

BIOGRAPHIES



Daniel A. Ameyaw received the M.Sc. degree in mechanical engineering from Kwame Nkrumah University of Science and Technology, Ghana, in 2014. He is currently pursuing a Ph.D in mechanical engineering at University of Duisburg-Essen, Germany. His research interests include structural analysis, vibration analysis, reliability, filtering, and classification applied to elastic mechanical systems.



ods.

Qi Deng received the M.Sc. degree in mechanical engineering from University of Duisburg-Essen, Duisburg, Germany, in 2014, where she is currently pursuing the Ph.D. degree in mechanical engineering. Her current research interests include modeling and prediction of driving behaviors and applications of machine learning methods.



Dirk Söffker received the Dr.-Ing. degree in mechanical engineering and the Habilitation degree in automatic control/safety engineering from University of Wuppertal, Germany, in 1995 and 2001, respectively. Since 2001, he leads the Chair of Dynamics and Control, University of Duisburg-Essen, Germany. His current research interests include diagnostics and prognostics, modern methods of control theory, human interaction with safe human interaction with technical systems, safety and reliability control engineering of technical systems, and cognitive technical systems.