features improved the performance of the classification when additional aggregations, such as minimum, maximum, and mean values of features are considered. The best performing approach includes aggregated features about the head angles, the mouth openness and the eye openness and is achieved with the Random Forest configuration (Table XI), which has a moderate processing time compared to random forest configurations with 32 individual trees. The corresponding confusion matrix is shown in Fig. 3.

The results are thus based on commodity hardware, in this case smartphones' sensor data. In order to further improve the algorithms or extend the feature set, it would be possible to add further data sources via sensor fusion. There are also recent works on this subject (e.g., [20]) that describe how they combine data from a Mixed Reality (MR) headset (Microsoft HoloLens), a smart-watch, a smartphone, and a data logger that logs data from the vehicle's on-board-diagnostics (OBD) interface. Apart from the fact that it is not allowed in road traffic in many countries, a Mixed Reality headset could provide precise data on head rotations, thereby complementing and making more robust the existing distraction and drowsiness detection based on video data explained in this work. Smart Watches, in turn, can provide data such as the pulse, but also arm movement data to detect distractions where the driver may be looking at the road but is mentally focused on something else, e.g., while changing the radio station, digging around in the pocket looking for something, or typing a text message.

Hence, this would contribute again to distraction detection, but also to drowsiness detection, e.g., as a reduced arm movement, which is typically necessary to keep the vehicle on track, can indicate fatigue [20].

While the data sources mentioned before primarily provide data about the driver and his status, data from the vehicle, on the other hand, can provide information on how the vehicle is used. In [21] the term Quantified Vehicle, which captures sensor data about itself (vehicle usage data) and its environment is introduced. Vehicle usage data could give insights on how risky the vehicle is used, e.g., whether it is often at risk of slipping off the road due to enormous lateral accelerations in curves, or how often and for how long wheels spin. Such events are considered to be part of an aggressive driving style, and this is, as mentioned in Section I, one of the causes of many deaths in road traffic. However, unfortunately such data is currently only available in test vehicles that read the data directly from the vehicle's bus system (e.g., CAN bus).

Services on the market so far mainly use the above mentioned OBD interface, which was actually not developed for this purpose. Theoretically, this interface should provide a lot of interesting data, but in practice with several tested vehicles of different manufacturers, there were only 10-15 relevant signals like vehicle speed, revolutions per minute (RMP), or oil temperature each, which allows driving style analyses, but only to a very limited extent. Recently, however, more and more vehicle manufacturers are discovering their interest in retaining customers with services, which increasingly creates the technical possibilities for third party service developers to gain access to their customers' vehicle data. Even marketplaces have already been created (e.g. caruso-dataplace.com), which offer data from several manufacturers in a uniform data format. This in turn shows that sensor fusion between smartphone sensor data and vehicle data can become a relevant topic, and that there is a clear trend towards turning vehicle and driver data into a business model, as described for example in [22]. In Table XII, we summarize a preliminary list of interesting events we think of extending our solution to in the future.

TABLE XI BEST PERFORMING CONFIGURATION

# of Trees	Max Depth	Min Leaf Samples	Macro F1 Score
8	32	1	0,75287793

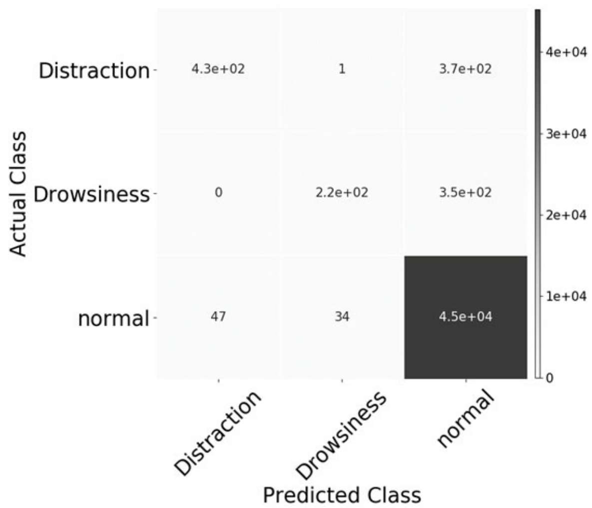

Fig. 3. Confusion matrix for the best performing algorithm

TABLE XII. PARAMETERS FOR THE PROPOSED MODEL EXTENSION

Interesting Events w.r.t risky driving	Data Source
Head movements related to distraction or drowsiness	Smart Glass / Mixed Reality Headset: head rotation sensor
Arm movements that are not intended for driving	Smart Watch: acceleration sensor
Individual arm movement on the steering wheel to keep the vehicle on track, and unusual deviations therefrom	Smart Watch: acceleration sensor
Individual driving style like typical rpm usage and gear change behavior per road segment, to detect unusual behavior	Vehicle OBD interface data (limited set of vehicle sensor data)
Wheel speed per wheel to detect wheel spinning or locking tires	Vehicle CAN bus data (access to more vehicle sensors): wheel speed
Harsh braking, harsh accelerating, harsh cornering	Vehicle CAN bus data (access to more vehicle sensors): vehicle acceleration sensor
Distances to other vehicles and vulnerable road users	Vehicle CAN bus data (access to more vehicle sensors): RADAR/LiDAR/video sensor

VIII. CONCLUSION

The goal of our paper is to present an approach based on ML to extend the existing system with a component that enables learning and adapts to the individual needs of the driver. The long-term vision is to learn from the collected data and make the previously rigid thresholds more flexible. This will be achieved by using ML techniques to establish correlations between the observed behavior and the underlying data and to gain new insights. These findings should be evaluated and can be integrated into the threshold-based system afterwards in order to increase the performance regarding the correct recognition of dangerous situations for the individual driver. In order to meet this requirement, extensive experiments were carried out with the help of the data to determine how an implementation can be done.

The experiments show how important the handling of time series data is with regards to performance and accuracy of the models. Especially when dealing with variable sized time windows problems occur which can skew the resulting classifications. As [15] notes that dynamic time warping, as a representative of a time series analysis methods, can automatically cope with time deformations it seems interesting to test the data analysis step with time series analysis methods. So, we would like to mention that the computation time in the cloud is not critical for the proposed system since it is post processing of the data.

In the future the user confirmed dangerous event information is implemented as additional input besides the captured data by the threshold system. As a starting point it is planned to have questionnaires about the confirmation of dangerous events. That data can afterwards be used to have a more precise classification. Let us assume the cases where the driver denies the proposed dangerous state by the threshold-based system all have certain unknown criteria. These unknown criteria can be linked by the ML component to those cases where the threshold-based system does not perform as well as intended and propose different dangerous states accordingly. This way the ML component of the system can improve on user feedback and give improving recommendations to the threshold-based system regarding its thresholds.

We believe that the research we present is useful for both science and practice to further facilitate the development of smartphone-based vehicle information systems (see e.g., [23] for more details on vehicle IS). Our research indicates that ML technologies such as Neuronal Networks and Random Forests as promising approaches for learning and adapting threshold-based reasoning about individual drivers' states.

ACKNOWLEDGMENT

The research is funded by the Russian Science Foundation (project # 18-71-10065). The architecture model (Section III) has been developed in scope of the Russian State Research #0073-2019-0005. Sections I and VII of this work are partially funded by the InSecTT project (https://www.insectt.eu/). InSecTT has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey. In Austria the project was also funded by the program "ICT of the Future" and the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK).

REFERENCES

[1] National Highway Traffic Safety Administration, "Early Estimate of Motor Vehicle Traffic Fatalities in 2019,"

[2] H. O. World, "Global status report on road safety 2018. Geneva: World Health Organization; 2018," 2018.

[3] A. Bolovinou, A. Amditis, F. Bellotti, and M. Tarkiainen, "Driving style recognition for co-operative driving: A survey," in *The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications*, 2014, pp. 73–78.

[4] N. P. Chandrasiri, K. Nawa, and A. Ishii, "Driving skill classification in curve driving scenes using machine learning," *J. Mod. Transport.*, vol. 24, no. 3, pp. 196–206, 2016.

[5] W. Wang, J. Xi, H. Chen, and R. Yang, "Modeling and Recognizing Driver Behavior Based on Driving Data: A Survey," *Mathematical Problems in Engineering*, vol. 2014, p. 245641, 2014.

[6] A. Hanelt, E. Piccinini, R. W. Gregory, B. Hildebrandt, and L. M. and Kolbe, "Digital Transformation of Primarily Physical Industries - Exploring the Impact of Digital Trends on Business Models of Automobile Manufacturers," in *Wirtschaftsinformatik Proceedings*, 2015. [Online]. Available: https://aisel.aisnet.org/wi2015/88

[7] F. Lindow and A. Kashevnik, "Driver Behavior Identification Based on Smartphone Sensor Data and Machine Learning Methods," *25th Conference of Open Innovations Association (FRUCT)*, 2019, pp. 196–203.

[8] A. Kashevnik, I. Lashkov, A. Ponomarev, N. Teslya, and A. Gurtov, "Cloud-Based Driver Monitoring System Using a Smartphone," *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6701–6715, 2020.

[9] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and Mobile Application for Driver Behavior Analysis and Accident Prevention," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[10] D. M. Vlachogiannis, E. I. Vlahogianni, and J. Golias, "A Reinforcement Learning Model for Personalized Driving Policies Identification," *International Journal of Transportation Science and Technology*, 2020.

[11] F. K. Adamidis, E. G. Mantouka, and E. I. Vlahogianni, "Effects of controlling aggressive driving behavior on network-wide traffic flow and emissions," *International Journal of Transportation Science and Technology*, 2020.

[12] Y. Ma, Z. Zhang, S. Chen, Y. Yu, and K. Tang, "A comparative study of aggressive driving behavior recognition algorithms based on vehicle motion data," *IEEE Access*, vol. 7, pp. 8028–8038, 2018.

[13] Z. E. A. Elassad, H. Mousannif, H. Al Moatassime, and A. Karkouch, "The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review," *Engineering Applications of Artificial Intelligence*, vol. 87, 2020.

[14] C. Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accident Analysis & Prevention*, vol. 126, pp. 95–104, 2019.

[15] G. Singh, D. Bansal, and S. Sofat, "A smartphone based technique to monitor driving behavior using DTW and crowdsensing," *Pervasive and Mobile Computing*, vol. 40, pp. 56–70, 2017.

[16] D. Yi, J. Su, C. Liu, M. Quddus, and W.-H. Chen, "A machine learning based personalized system for driving state recognition," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 241–261, 2019.

[17] J. Yu, Z. Chen, Y. Zhu, Y. Chen, L. Kong, and M. Li, "Fine-Grained Abnormal Driving Behaviors Detection and Identification with Smartphones," *IEEE Trans. on Mobile Comput.*, vol. 16, no. 8, pp. 2198–2212, 2017.

[18] A. Kashevnik, K. Karelskaya, and M. Repp, "Dangerous Situations Determination by Smartphone in Vehicle Cabin: Classification and Algorithms," in *2019 24th Conference of Open Innovations Association (FRUCT)*, 2019, pp. 130–139.

[19] H. Blockeel and J. Vanschoren, “Experiment Databases: Towards an Improved Experimental Methodology in Machine Learning,” in *Lecture notes in computer science, Knowledge Discovery in Databases: PKDD 2007*, J. N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenič, and A. Skowron, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 6–17.

[20] Lechner, G., Fellmann, M., Festl, A., Kaiser, C., Kalayci, T. E., Spitzer, M., & Stocker, A. "A Lightweight Framework for Multi-device Integration and Multi-sensor Fusion to Explore Driver Distraction." International Conference on Advanced Information Systems Engineering. Springer, Cham, 2019, pp. 80-95.

[21] A. Stocker, C. Kaiser, and M. Fellmann, “Quantified Vehicles,” *Business & information systems engineering*, vol. 59, no. 2, pp. 125–130, 2017.

[22] C. Kaiser, A. Stocker, G. Viscusi, A. Festl, P. Moertl, and M. Glitzner, Eds., ”Quantified cars: An exploration of the position of ICT start-ups vs. car manufacturers towards digital car services and sustainable business models”, 2017.

[23] C. Kaiser, A. Stocker, A. Festl, G. Lechner and M. Fellmann, “A Research Agenda for Vehicle Information Systems”, European Conference on Information Systems, 2018.