Predicting Human Errors from Gaze and Cursor Movements

Rachid Riad Saboundji*, Róbert Adrian Rill*†
*Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
†Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania
Email: s.rachid.riad@gmail.com (R.R.S.), rillroberto88@yahoo.com (R.A.R.)

***Abstract*—Intelligent interfaces are increasingly integrated into diverse technological areas. In complex high-risk environments, where humans represent a crucial part of the system and their attention is often divided between simultaneous activities, imminent human errors may have serious consequences. Enhancing interfaces with predictive capabilities promotes the safe and reliable operation of such systems. In this work, we employ a data-driven approach to predict human errors in a special divided attention task involving timing constraints and requiring focused concentration and frequent shifts of attention. We performed a longitudinal study with 10 subjects, and constructed time series from the experimental data using gaze movement and mouse cursor motion features in order to classify successful and failed actions. We evaluate classical machine learning algorithms, compare them with a more traditional temporal modeling approach and a deep learning based LSTM model. Employing a leave-one-subject-out cross-validation procedure we achieve a classification accuracy of up to 86%, with LSTM presenting the highest performance. Furthermore, we investigate the trade-off between evaluation metrics and anticipation window, i.e. the time remaining until the correct action can still be performed. We conclude that prediction is feasible and accuracy and F1-score increases, despite the training dataset becoming greatly imbalanced. Investigating the anticipation window allows to understand how far in advance human errors need to be predicted in order to initiate preventive measures. Our efforts have implications for the design of predictive interfaces involving decision making under time pressure in dynamic divided attention environments.**

***Index Terms*—human error, time series, LSTM, gaze tracking, anticipation window, predictive interface.**

I. Introduction

Human error is one of the most prominent and important metrics influencing the performance and efficiency of human-machine collaboration [1]. It is well-known that humans have a limited amount of mental resources that can be divided among simultaneous tasks [2]. In complex dynamic environments and professions requiring situation awareness, such as automobile drivers, pilots, nuclear power plant operators, airspace controllers, medical providers or civil emergency operators, human errors can have serious consequences. Therefore, their prevention is of crucial importance for the reliable and safe operation of complex systems.

One of the most important factors affecting human performance, shown by a significant amount of research is cognitive load [3]. Higher cognitive load leads to an increase in the number of errors and as a result performance is impaired. Since eye movements can be used to infer users' cognitive states and intentions [4], and the notions of intention and error are inseparable [5], gaze motion features represent a useful source of information for predicting near-future errors. Eye movements can be recorded in a non-obtrusive way and analyzed in parallel without interfering with the execution of a task. Although gaze patterns are often predictable, for instance gaze centers on the object of interest before corresponding motor movements begin, gaze and cursor coordination can show complex and nuanced interaction patterns as well [6], [7]. For example, gaze can leave the target area moving on to the next task before the motor action is completed. Accordingly, it is not straightforward to anticipate and prevent user errors.

The technological and scientific advances in the last decades create an opportunity to realize complex systems that monitor humans and support decision-making. Designing predictive intelligent interfaces can facilitate the prevention of human errors and, therefore, enhance user experience and more importantly promote the beneficial interaction between humans and computers. In this work, we exploit a special divided attention task that requires continuous focused concentration and frequent shifts of attention. We define human errors as missed/failed user actions and attempt to differentiate them from successful actions using gaze and mouse cursor movement features. After performing a longitudinal study with 10 participants, we constructed time series from the experimental data and formulated the problem as classification. The definition is illustrated by Fig. 1, which also shows the properties of the time series we investigate: (i) history and (ii) anticipation window size, respectively. Since we seek to predict in advance whether the user will perform the correct action, a successful action is included into the training set only if it falls within the anticipation window, i.e. it has not been performed already at the current time. Therefore, although the classification problem becomes easier as the prediction time is closer to the deadline, the training data becomes increasingly imbalanced as less correct actions are included.

To summarize, the contributions of the present work can be outlined as follows.

- We employ a data-driven approach for human error prediction in a complex divided attention task, by classifying time series data constructed using gaze movement and

The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00001 and EFOP-3.6.3-VEKOP-16-2017-00002). R.R.S. gratefully acknowledges the support of the Tempus Public Foundation for sponsoring his PhD study.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

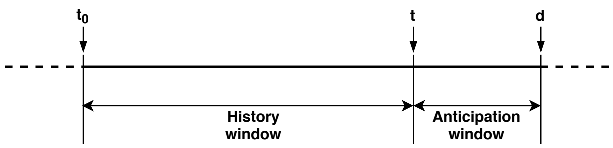


Fig. 1. Prediction problem definition and properties of the time series data. At the current time t we attempt to anticipate whether the user will perform the correct action before the deadline d. The size of the history window (t-t_0) represents the amount of temporal data used for prediction. The size of the anticipation window (d-t) denotes how far in advance the prediction is made.

mouse motion features.

- We evaluate classical machine learning algorithms and compare the results with the traditional dynamic time warping (DTW) [8], [9] and k-nearest neighbors (kNN) combination.
- We improve upon our preliminary results [10] with a simple deep learning based long short term memory (LSTM) [11], [12] model, which compares favorably against the other approaches and achieves a cross-validated accuracy of up to 86%.
- We investigate the trade-off between accuracy and anticipation window size, as this latter parameter is decreased towards zero and the training data becomes highly imbalanced. We observe that both accuracy and F1-score increase significantly, that is, human errors can be predicted in this case too in our task.

The rest of the paper is organized as follows. Section II reviews related works concerned with inferring user intentions, actions and errors, providing a background for the present study. Section III introduces the divided attention task, summarizes the experiments we performed with 10 participants, and describes our methodology for computing features and applying the classification algorithms. Section IV presents our qualitative findings. In Section V we discuss our results, their limitations and highlight future directions. Finally, Section VI concludes the work.

II. Related work

Related research includes a variety of studies concerned with inferring and predicting users' task-related intentions and goals in order to improve interface usability and enhance user experience in web-browsing tasks [7], [13], [14], gameplay environments [15], [16], human-robot collaboration tasks [17]–[20]. Enhancing situation awareness represents another means to provide user assistance in manufacturing environments [21], multiple screen surveillance [22], or driving [23], [24]. This latter is one of the most widely investigated tasks, and besides exploring the driver's level of inattention, involves predicting future intentions and actions as well to anticipate dangerous situations [25]–[27]. Anticipating user responses may have applications in collaborative contexts [28], daily activity logging [29], and mouse interaction tasks in general [30], [31].

The common aspect in most of the works listed is using eye movements and gaze tracking data as input to predictive models. Eye tracking measures can also be used to investigate cognitive load [32], for instance pupil diameter and microsaccade magnitude [33], or eye landmarks, pupil trajectories and eye region images [3]. Estimating cognitive load, and cognitive processes in general, helps in inferring users' understanding and awareness of the current situation, and in predicting the ability to respond effectively to challenges.

To sum up, a large variety of studies propose to assist human users in a diversity of tasks aiming from improving usability and user experience, through sharing autonomy with robots, enhancing situation awareness, inferring cognitive load, to predicting actions in advance. Fewer studies address the problem from a human error perspective, and attempt to anticipate near future errors. It is important to note that human errors can also be considered actions, or lack thereof. In any case, they represent deviation from expectation or intention and may have serious consequences in safety-critical systems.

Cognitive architectures have been successfully applied to predict errors in different tasks [34], [35]. However, they use simulations to estimate aggregate probability, i.e. not when errors are likely to occur during the execution of the task. Also, such computational models are sensitive to parameter changes, which may limit their generalization. Formal models can be used as well to predict human error (see, e.g., [36] and the references therein), but they suffer from the state space explosion problem.

Similar to ours, some works have used a data driven approach. Ratwani and Trafton investigated the prediction of postcompletion errors from reaction time and eye movement measures [37], [38]. However, this type of error occurs after the main goal has already been accomplished. Accordingly, in high risk environments failing to complete the main goal may have serious effects. Damacharla et al. [39] proposed machine learning models to predict errors in human-machine teaming systems, but evaluation was performed on synthetic data only.

In the present work we employ a data-driven approach to classify omission errors and successful user actions in a novel divided attention task that has not been previously investigated to the best of our knowledge. This task is highly dynamic, requiring frequent shifts of attention and fast mouse cursor movements. As opposed to many other works concerned with action prediction that simulated one scenario with a fixed target, there are multiple targets here to be clicked repeatedly, the user needs to handle simultaneous tasks over 2-3 minutes and select the appropriate sequence of the targets to maximize performance. The omission errors in our task result in performance decrements, and may be equivalent to having serious consequences in real-world environments where users are under pressure and time constraints, and their attention is divided between simultaneous activities. To predict the errors, we evaluate several classical machine learning algorithms, a traditional temporal modeling algorithm and a deep learning based LSTM model, with this latter method achieving the best performance.

III. Methods

A. Task, participants and experiments

To analyze human performance, we have designed and implemented a simplified version of *Train of Thought*, one of the most popular Lumosity games. Lumosity[1] is an online training platform comprised of a set of computerized games designed by scientists, each aiming to train one of five core cognitive abilities [40]: attention, processing speed, memory, flexibility and problem solving. Train of Thought is an attention game that tests users' visual divided attention and working memory by requiring them to simultaneously concentrate on multiple moving objects over 2-3 minutes and direct them to their correct destinations through mouse clicks. In our custom version of this divided attention task we have simplified the graphical design, made efforts to control the complexity of the game, and kept one variable to manipulate difficulty (namely the speed of the moving objects) in order to keep players challenged as they are progressing.

Using our version of the divided attention task, we conducted a longitudinal study with 10 participants aged between 25 and 30 years, who had normal or corrected to normal vision and reported no attentional disorders nor color deficiency. The subjects were asked to play with the divided attention task over a several day period, resulting in 60 trials each. The difficulty of each gameplay was adjusted based on the score achieved in the previous trial. Data about experiments was logged for later analysis, including mouse and eye-gaze movements. For gaze tracking we used the Tobii EyeX Controller [41] device. The sampling frequency for all data was 60 Hz.

For a detailed description about the design process of the divided attention task and the experiments performed, the interested reader is referred to our previous work [42].

B. Prediction task, features and classification algorithms

Exploiting the data collected in our longitudinal study, we seek to anticipate omission errors and frame the prediction problem as binary classification. For an input sample we consider the time interval before a given deadline in our divided attention task. This is partitioned into history window and anticipation window, as shown in Fig. 1, and we attempt to predict whether the user will manage or fail to complete the successful click action before the deadline. The history window size represents the span of the time interval used for prediction. Anticipation window size denotes the time remaining until the last possible moment for performing the successful click action for the given deadline, and illustrates how far in advance we predict correct/failed click actions.

The historical time series data used for prediction is constructed using gaze movement and hand (i.e. mouse cursor) motion features. The prediction problem is not straightforward, as human gaze and mouse sequences might show complex and nuanced interaction patterns [6], [7]. A descriptive snapshot visualizing one gaze and mouse cursor trajectory is shown in Fig. 2: the mouse cursor moves towards the target more or less

[1]https://www.lumosity.com/

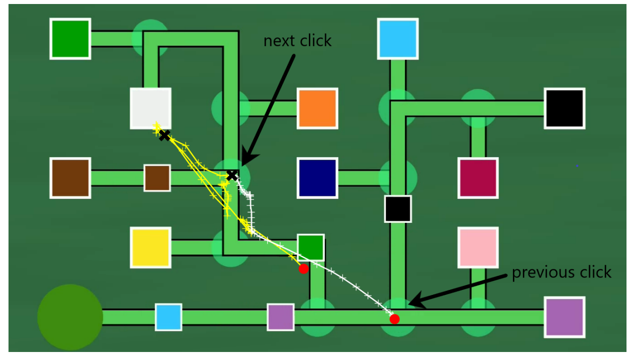


Fig. 2. Snapshot of the divided attention task used in the experiments, with 1 second long mouse and gaze trajectory overlaid. The small squares enter the screen from the big green circle, are moving continuously along the tracks and need to reach their color matching large square. Gaze and mouse trajectory are shown with yellow and white, respectively. The black X's show the current positions (the user is about to click and correct the direction of the switch node on the path of the green square), and the red dots are the positions 1 second ago (the user switched for the purple square). For more details about the divided attention task, see our previous work [42].

straightforwardly, while the gaze scans the scene and leaves the target before the click action.

The features used for classification are three types of screen distances that characterize the cursor and gaze movement:

1) gaze-target: distance of the gaze point (screen coordinates of gaze direction) from the target to be clicked,
2) mouse-target: distance of the cursor from the target,
3) gaze-mouse: distance between the gaze point and cursor.

We computed these distances for all frames in a time interval before deadlines from the experimental data, and use the resulting multivariate time series for classification.

During our evaluations, we experiment with the different combinations of the three features, and several values for the history and anticipation window size, respectively. In each case we train a new model and report the classification accuracy values, and employ a leave-one-subject-out cross-validation scheme where an algorithm is trained on the data of 9 subjects and evaluated on the test data of the remaining participant; this procedure is repeated for all 10 possible combinations and the weighted average performance is computed.

To allow for a consistent presentation of the results, we define default values for the time series properties. For the history window size 90 frames (1.5 seconds) is a reasonable choice. The default value for anticipation window was set to 40 frames (approx. 0.67 seconds), as this results in a balanced dataset of 2296 positive (failed action, i.e. omission error) and 2293 negative (correct action) samples in total. It is expected that decreasing the anticipation window makes the prediction problem easier, but in the same time the training dataset becomes increasingly imbalanced, as negative samples falling before the anticipation window have to be dropped. In other words, it makes sense to predict a correct click action only if it has not been performed already. Accordingly, the F1-scores are also reported in this case.

The classical machine learning algorithms evaluated are

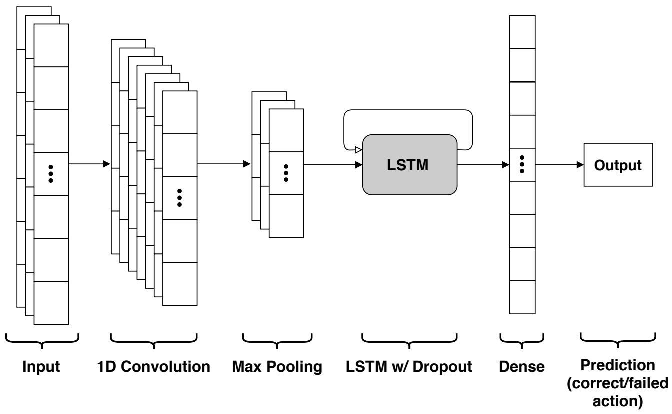


Fig. 3. LSTM network architecture: the input layer is followed by a 1D convolution layer (with 100 filters of size 20 and ReLU activation function), max pooling operation (with a pool size of 2), LSTM block (with 100 cells), dropout layer (with rate 0.5) and finally a dense layer with the sigmoid activation function.

the scikit-learn [43] implementation of logistic regression (LR), support vector machines (SVM) with RBF kernel and random forests (RF). We also experiment with the traditional temporal modeling algorithm DTW [8], [9]. In this case we used a kNN approach with the DTW kernel, and tuned the k hyperparameter to achieve the best possible accuracy. The multivariate time series vectors were flattened out in the above cases. Finally, the results are compared to those of a recurrent neural network, namely an LSTM [11], [12] model. Here, at each round of the cross-validation, the training data was split randomly into train and validation sets (a 90%–10% split was used), with the validation data being used to apply early stopping and save the best validation weights. The network architecture is shown in Fig. 3. The batch size during training was 64 and a maximum number of 200 training epochs was used applying early stopping with a patience of 20 epochs by monitoring the validation loss. The LSTM models were trained via the RMSprop optimizer with a learning rate of $1e$–4, using the Keras [44] library with TensorFlow [45] backend.

IV. Results

Table I shows the results for the different combinations of the three features. History and anticipation window size was set to 90 and 40 frames, respectively. Considering the gaze-mouse distance as the sole predictor gives accuracy values only slightly above chance level, meaning that the coordinated movement between subjects' gaze and the cursor does not differ significantly among correct and failed click actions. We can also see that adding the mouse-target distance feature to the other two predictors results in the largest increases in accuracy, when compared to the other similar cases. That is to say, mouse-target distance was found to bring higher accuracy in all cases than gaze-target distance. The highest performance was achieved with LSTM with all three features, followed by kNN with DTW kernel using the gaze-target and mouse-target features only.

Fig. 4 depicts the accuracy values for each algorithm separately for the subjects from our experiments. History and anticipation window size was set to 90 and 40 frames, respectively. The performance is mostly above 80%, except in the DTW approach. In general LSTM outperforms the other methods by a slight margin. More importantly, the variation of accuracy among participants is not considerably high in case of the same algorithm.

Fig. 5 illustrates the accuracy values when the history window size was varied, keeping the anticipation window fixed at 40 frames. The LR, SVM and RF methods show close values to each other. The kNN with DTW method shows a fast drop from 84% at 0.17 seconds to 75% at 2 seconds. LSTM outperforms the other algorithms and shows a roughly constant pattern. A general conclusion is that varying the history window size does not have a significant effect on accuracy.

The effects of increasing the anticipation window size are illustrated by Fig. 6. History window size was set to 90 frames. Clearly the accuracy decreases in all cases as we try to predict the action of the users more and more in advance. At 2 seconds the accuracy is only slightly above chance level, i.e. predicting the user action over 2 seconds in advance is equivalent with random guessing.

The impact of shrinking the anticipation window towards zero and obtaining an increasingly imbalanced training dataset are shown in Fig. 7. Accuracy is inversely proportional to anticipation window size (Fig. 7a), i.e. prediction performance increases as the deadline of performing the action is closer. This is confirmed by F1-score as well in Fig. 7b: the results outperform a simple majority classifier as the anticipation window is decreased all the way down to 0.05 seconds, where the weights of the positive samples are over 90%.

V. Discussion

The dataset used in this study highlights prediction performance in cases when there is only a limited amount of time left to perform the required click action. Classifying correct and failed user actions still remains challenging, since although the gaze and mouse often show predictable and coordinated movements, complex interaction patterns can also be observed [6], [7]. For instance the gaze might leave the target before the click action, or the user might click just a bit too late right after the last possible moment. These latter cases are considered failed user actions in our divided attention task.

TABLE I
Feature Combination Accuracy Values.

gaze–target	✓			✓	✓		✓
mouse–target		✓		✓		✓	✓
gaze–mouse			✓		✓	✓	✓
LR	77.21	80.98	54.43	81.91	77.84	80.95	**83.00**
SVM	80.26	83.61	57.90	83.63	81.94	83.05	**83.85**
RF	80.10	82.35	55.07	83.46	80.78	82.72	**84.07**
kNN w\ DTW	80.24	83.22	58.40	**85.12**	75.05	77.40	76.86
LSTM	81.48	84.40	57.38	85.40	83.02	84.48	<u>**85.88**</u>

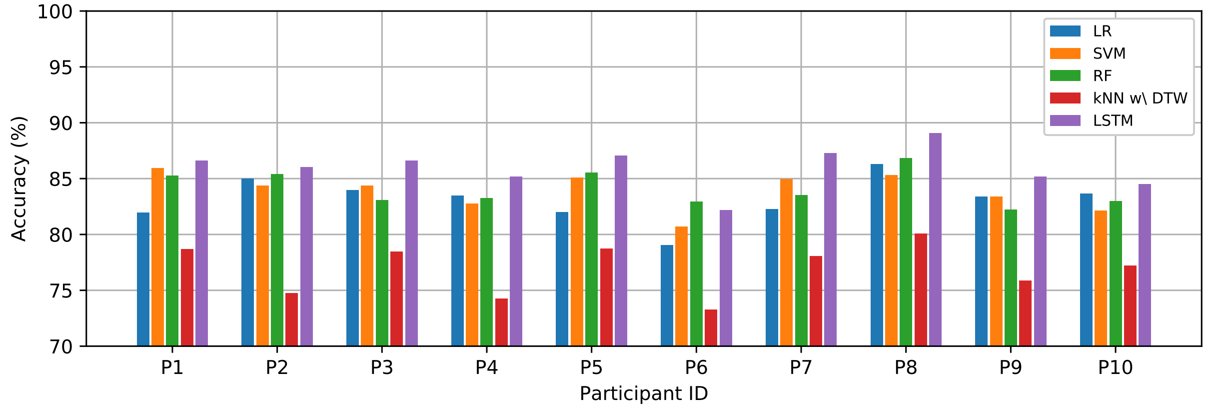


Fig. 4. Accuracy values separately for participants.

We have evaluated several classical machine learning algorithms, one traditional temporal modeling method and a deep learning based LSTM model for omission error prediction, and observed that considering longer time sequences for prediction does not increase accuracy (see Fig. 5). Furthermore, as expected, the performance drops as we attempt to classify the user actions more and more in advance (see Fig. 6). This observable trend may indicate that it is not effective to predict human errors more than 1.5 seconds in advance in dynamic environments where user interface operators are under time pressure and high cognitive load. However, it must be noted that selecting the appropriate anticipation window size is problem dependent and in highly dynamic environments where conditions change frequently, it may not be possible to predict human error many seconds in advance.

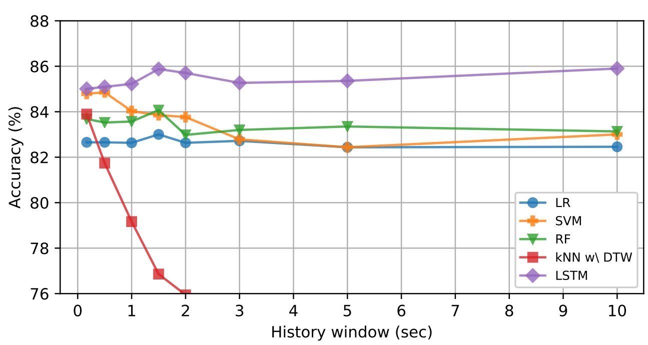


Fig. 5. Effect of history window size on accuracy.

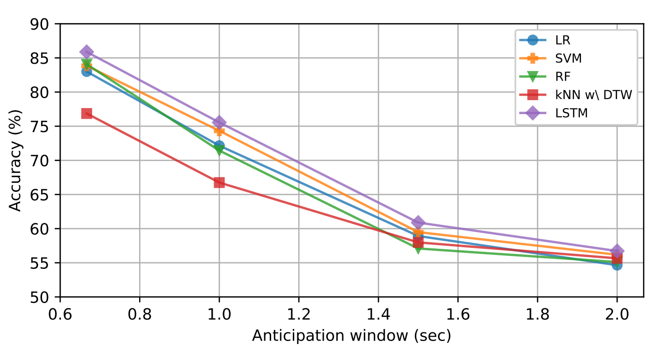
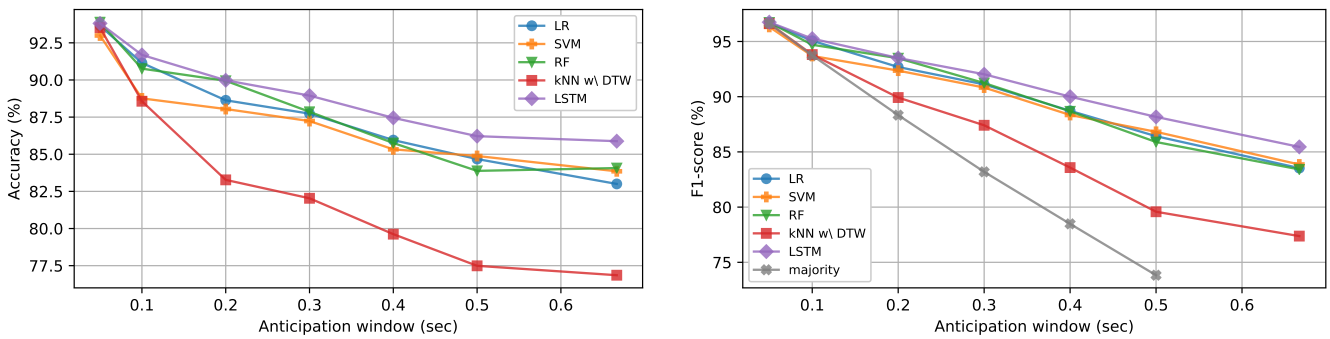


Fig. 6. Effect of increasing anticipation window size on accuracy.

Decreasing the anticipation window and using an imbalanced dataset for classification had an effect of increased prediction performance (Fig. 7). However, in realistic scenarios as the anticipation window gets smaller and the deadline for performing an action is approaching, the chance of successful and timely intervention, and the limit of human decision support as well get closer to zero. In other words, the effectiveness of prediction is not determined just by algorithm performance metrics, but by the anticipation window size too, since the execution time of the prediction itself and of the system intervention need to be considered in the same time. In specific safety-critical application scenarios it is crucial to optimize the trade-off between accuracy and anticipation window, and to understand how far in advance an error needs to be predicted in order to provide the needed assistance.

LSTM, capable of capturing long term dependencies [12], demonstrated the highest performance in all experiments, despite applying fine-tuning schemes in the case of the kNN algorithm with DTW kernel. Previous related works have also investigated LSTM models in various tasks and environments, such as inferring surgeons turn-taking intentions [20] or basic manipulation actions [19] in human-robot collaboration, predicting next mouse click interaction [31], anticipating driving maneuvers [26] or intent at intersection [27]. Similarly to our conclusions, they also found promising results with LSTM, superior against other algorithms, including DTW approaches.

Future works can further improve accuracy by considering other sensory input modalities, such as capturing electromyography and electroencephalography signals similarly to [20]. One novel idea considered by some of the related works is the adoption of multi-branch LSTM models that learn different information streams on separate branches and concatenate the hidden representations before the final output. For example Kwok et al. called such an approach dual-stream [31], or



(a) (b)

Fig. 7. Effect of decreasing anticipation window size on prediction performance.

Jain et al. sensory-fusion [26] LSTM. We also experimented with this idea where the input to one branch was the gaze and mouse interaction features investigated in this study, and using taskload features – derived from the number of tasks to be handled – as the input to another branch. Unfortunately this method did not improve the presented results. Nonetheless, we are confident that perceived cognitive load or mental workload measures, derived from electrophysiological or psychophysiological (e.g., changes in pupil diameter) signals or even from video data [3], represent a useful input stream for improving prediction performance. For the differences between taskload and mental workload, see, e.g. [46] and the references therein. Adopting a multi-branch LSTM model with mental workload as one of the input temporal streams might require further experiments with a larger sample size. Additionally, one could opt to use time-series forecasting instead of classification, or accumulate consecutive predictions in a sliding window fashion.

The features we investigated for human error prediction are general in the sense that they could be applied to other environments too. Gaze movement features alone can reach reasonable accuracy, and human motion can be recorded effectively and in a non-obtrusive way as well. Application areas and domains for the presented methods and their possible extensions include healthcare and robotic surgery, air traffic and flight control, vehicle interfaces, nuclear power plant operation, and critical decision support systems in general. As a concrete example, supervising operators can be assisted whose task is to monitor timed missions of multiple unmanned aerial vehicles [47].

VI. Conclusion

In this paper, we analyzed the classification of successful and failed user click actions in a highly dynamic divided attention task, requiring frequent shifts of attention and fast reactions under time constraints. We defined interaction features that characterize the gaze movements and mouse cursor motion of the subjects. Particularly, the distance over time of the cursor and gaze from the click target and from each other were considered. We constructed time-series from these, investigated different parameters for the sequences and evaluated several algorithms. A deep learning based LSTM model presented superior performance over classical approaches, achieving a cross-validated accuracy of up to 86%.

We investigated effects on prediction performance of the anticipation window, i.e. the time remaining before the last moment when an action can still be performed correctly. This allows to understand how early a predictor can reach the correct decision with a predefined accuracy. The further in advance prediction is performed, the greater the opportunity for preventive measures. We found that anticipating human errors in our dynamic task more 1.5 seconds in advance is not effective, but accuracy and F1-score increase as anticipation approaches zero, despite the training dataset becoming more and more imbalanced. Data-driven prediction models in general can provide meaningful information to intelligent systems, so that corrective actions can be taken to prevent human errors. Such human error prognostics may offer significant benefits to the sustainment of safety-critical systems. Enhancing user interfaces with predictive capabilities and even giving sufficient warning may allow preventive maintenance of human error generated impeding failures, and therefore efficient interaction with the system.

Acknowledgment

The authors acknowledge András Lőrincz, their PhD supervisor, for his help in starting this research project. The authors also thank their colleague, Kinga B. Faragó for her help in performing the data collection process.

Author contributions

Conceptualization: R.R.S., R.A.R.; Data curation: R.A.R.; Formal analysis: R.R.S., R.A.R.; Investigation: R.A.R.; Methodology: R.R.S., R.A.R.; Project administration: R.A.R.; Resources: R.R.S., R.A.R.; Software: R.R.S., R.A.R.; Supervision: R.A.R.; Validation: R.R.S., R.A.R.; Visualization: R.R.S., R.A.R.; Writing – original draft preparation: R.A.R.; Writing – review and editing: R.R.S., R.A.R.

REFERENCES

[1] P. Damacharla, A. Y. Javaid, J. J. Gallimore, and V. K. Devabhaktuni, "Common metrics to benchmark human-machine teams (HMT): A review," *IEEE Access*, vol. 6, pp. 38 637–38 655, 2018.

[2] D. Kahneman, *Attention and Effort*. Prentice-Hall, 1973.

[3] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman, "Cognitive load estimation in the wild," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 652:1–652:9. [Online]. Available: http://doi.acm.org/10.1145/3173574.3174226

[4] A. Borji, A. Lennartz, and M. Pomplun, "What do eyes reveal about the mind? Algorithmic inference of search targets from fixations," *Neurocomputing*, vol. 149, pp. 788–799, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214010078

[5] J. Reason, *Human Error*. Cambridge University Press, 1990.

[6] D. J. Liebling and S. T. Dumais, "Gaze and mouse coordination in everyday work," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ser. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1141–1150. [Online]. Available: http://doi.acm.org/10.1145/2638728.2641692

[7] S. Lee, J. Yoo, and G. Han, "Gaze-assisted user intention prediction for initial delay reduction in web video access," *Sensors*, vol. 15, no. 6, pp. 14 679–14 700, 2015.

[8] J. B. Kruskal and M. Liberman, "The symmetric time-warping problem: from continuous to discrete," in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, D. Sankoff and J. B. Kruskal, Eds. Addison-Wesley Publishing Co., Reading, Massachusetts, 1983, pp. 125–161.

[9] W. Meert and T. V. Craenendonck, "wannesm/dtaidistance v1.1.2," 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1314205

[10] R. R. Saboundji and R. A. Rill, "Predicting user actions under time constraints in a divided attention task," in *Pannonian Conference on Advances in Information Technology (PCIT 2019)*, Veszprém, Hungary, 2019, pp. 77–83. [Online]. Available: https://pcit2019.mik.uni-pannon.hu/images/program/PCIT_2019_Proceedings_NV.pdf

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[12] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[13] C. Kelton, J. Ryoo, A. Balasubramanian, and S. R. Das, "Improving user perceived page load times using gaze," in *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*. Boston, MA: USENIX Association, 2017, pp. 545–559. [Online]. Available: https://www.usenix.org/conference/nsdi17/technical-sessions/presentation/kelton

[14] N. Su, J. He, Y. Liu, M. Zhang, and S. Ma, "User intent, behaviour, and perceived satisfaction in product search," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser. WSDM '18. New York, NY, USA: ACM, 2018, pp. 547–555. [Online]. Available: http://doi.acm.org/10.1145/3159652.3159714

[15] J. Newn, "Enabling intent recognition through gaze awareness in user interfaces," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '18. New York, NY, USA: ACM, 2018, pp. DC15:1–DC15:4. [Online]. Available: http://doi.acm.org/10.1145/3170427.3173028

[16] R. Singh, T. Miller, J. Newn, L. Sonenberg, E. Velloso, and F. Vetere, "Combining planning with gaze for online human intention recognition," in *17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Stockholm, Sweden, 2018.

[17] H. Admoni and S. Srinivasa, "Predicting user intent through eye gaze for shared autonomy," *AAAI Fall Symposium Series*, 2016. [Online]. Available: https://www.aaai.org/ocs/index.php/FSS/FSS16/paper/view/14137

[18] R. M. Aronson, T. Santini, T. C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni, "Eye-hand behavior in human-robot shared manipulation," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: ACM, 2018, pp. 4–13. [Online]. Available: http://doi.acm.org/10.1145/3171221.3171287

[19] P. Schydlo, M. Rakovic, L. Jamone, and J. Santos-Victor, "Anticipation in human-robot cooperation: A recurrent neural network approach for multiple action sequences prediction," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–6, 2018.

[20] T. Zhou and J. P. Wachs, "Early prediction for physical human robot collaboration in the operating room," *Auton. Robots*, vol. 42, no. 5, pp. 977–995, Jun. 2018. [Online]. Available: https://doi.org/10.1007/s10514-017-9670-9

[21] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, S. Ladstätter, G. Paar, and R. Velik, "Towards real-time probabilistic evaluation of situation awareness from human gaze in human-robot interaction," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 247–248. [Online]. Available: http://doi.acm.org/10.1145/3029798.3038322

[22] S. Tremblay, D. Lafond, C. Chamberland, H. M. Hodgetts, and F. Vachon, "Gaze-aware cognitive assistant for multiscreen surveillance," in *Intelligent Human Systems Integration*, W. Karwowski and T. Ahram, Eds. Cham: Springer International Publishing, 2018, pp. 230–236.

[23] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, 2004.

[24] C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel, "Driver-activity recognition in the context of conditionally autonomous driving," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 1652–1657.

[25] A. Jain, H. S. Koppula, B. Raghavan, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3182–3190, exported from https://app.dimensions.ai on 2019/05/16. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1095746446andhttp://arxiv.org/pdf/1504.02789

[26] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3118–3125.

[27] A. Zyner, S. Worrall, J. R. Ward, and E. M. Nebot, "Long short term memory for driver intent prediction," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1484–1489.

[28] C.-M. Huang, S. Andrist, A. Sauppé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Front. Psychol.*, vol. 6, no. 1049, 2015.

[29] T. Kurashima, T. Althoff, and J. Leskovec, "Modeling interdependent and periodic real-world action sequences," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, pp. 803–812. [Online]. Available: https://doi.org/10.1145/3178876.3186161

[30] P. T. Pasqual and J. O. Wobbrock, "Mouse pointing endpoint prediction using kinematic template matching," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 743–752. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557406

[31] T. C. K. Kwok, E. Y. Fu, E. Y. Wu, M. X. Huang, G. Ngai, and H. V. Leong, "Every little movement has a meaning of its own: Using past mouse movements to predict the next interaction," in *23rd International Conference on Intelligent User Interfaces (IUI '18)*, Tokyo, Japan, 2018, pp. 397–401. [Online]. Available: https://doi.org/10.1145/3172944.3173002

[32] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring cognitive load using eye tracking technology in visual computing," in *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, ser. BELIV '16. New York, NY, USA: ACM, 2016, pp. 78–85. [Online]. Available: http://doi.acm.org/10.1145/2993901.2993908

[33] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze," *PLOS ONE*, vol. 13, no. 9, pp. 1–23, 09 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0203629

[34] M. Halbrügge, M. Quade, and K.-P. Engelbrecht, "A predictive model of human error based on user interface development models and a cognitive architecture," in *13th International Conference on Cognitive Modeling*, Groningen, Netherlands, 2015, pp. 238–243.

[35] F. P. Tamborello and J. G. Trafton, "Human error as an emergent property of action selection and task place-holding," *Human Factors*,

vol. 59, no. 3, pp. 377–392, 2017. [Online]. Available: https://doi.org/10.1177/0018720816672529

[36] M. L. Bolton and E. J. Bass, *Enhanced Operator Function Model (EOFM): A Task Analytic Modeling Formalism for Including Human Behavior in the Verification of Complex Systems*. Cham: Springer International Publishing, 2017, pp. 343–377. [Online]. Available: https://doi.org/10.1007/978-3-319-51838-1_13

[37] R. M. Ratwani and J. G. Trafton, "A generalized model for predicting postcompletion errors," *Topics in Cognitive Science*, vol. 2, no. 1, pp. 154–167, 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2009.01070.x

[38] ——, "A real-time eye tracking system for predicting and preventing postcompletion errors," *Human–Computer Interaction*, vol. 26, pp. 205–245, 09/2011 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07370024.2011.601692

[39] P. Damacharla, A. Y. Javaid, and V. K. Devabhaktuni, "Human error prediction using eye tracking to improvise team cohesion in human-machine teams," in *Advances in Human Error, Reliability, Resilience, and Performance*. Springer International Publishing, 2018, vol. 778, pp. 47–57. [Online]. Available: https://doi.org/10.1007\%2F978-3-319-94391-6_5

[40] J. L. Hardy, F. Farzin, and M. Scanlon, "The science behind Lumosity, Version 2," 2013, lumos Labs, Inc.

[41] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello, "Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research," *Behavior Research Methods*, vol. 49, no. 3, pp. 923–946, 2017. [Online]. Available: https://doi.org/10.3758/s13428-016-0762-9

[42] R. A. Rill, K. B. Faragó, and A. Lőrincz, "Strategic predictors of performance in a divided attention task," *PLOS ONE*, vol. 13, no. 4, pp. 1–27, 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0195131

[43] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[44] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[45] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," https://www.tensorflow.org/, 2015.

[46] R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Applied Ergonomics*, vol. 74, pp. 221–232, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687018303430

[47] M. Elfar, H. Zhu, A. Raghunathan, Y. Y. Tay, J. Wubbenhorst, M. L. Cummings, and M. Pajic, "Wip abstract: Platform for security-aware design of human-on-the-loop cyber-physical systems," in *2017 ACM/IEEE 8th International Conference on Cyber-Physical Systems (ICCPS)*, 2017, pp. 93–94.