Towards Neuro-adaptive Modelling Environments: Report from a Study on Prediction of Business Process Model Comprehension Performance using Biometric Data

John Krogstie* and Kshitij Sharma*

Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Abstract

Visual conceptual modelling is a central activity in information systems analysis and design. Comprehension of visual models such as process models are also central in the use as part of many information systems such as company quality systems, and it is both in development and use of IT important to be able to support model comprehension tasks in an efficient manner. Model comprehension can be looked upon as a learning process, which in general is studied in more detail than modelling also using biometrics. In multi-modal analytics, one collects biometric data from different sensors, including EEG, eye-tracking, wristbands and facial expression (through cameras). This paper presents an approach of using ensemble (machine) learning as a top-down approach for getting performance prediction from the combined dataset collected from this kind of sensors when working with business process representations such as BPMN models. We present in this paper the results on the use of ensemble learning for comprehension performance prediction of comprehension of business process models and textual process descriptions with 57 subjects. High accuracy is witnessed. Although best prediction is found when we use data from all sensors, equally satisfactory results can be found in using only data from eye-tracking and cameras for detecting facial expression in combination. We also witness early prediction of results. This opens possibilities for using such techniques for studying other modelling activities with high ecological validity using techniques from multi-modal analytics and for developing neuro-adaptive modelling environments.

Keywords

Business Process Modelling, Multi-modal biometrics, Ensemble learning, Neuro-adaptive systems

1. Introduction

Business process descriptions are found at various levels in all organizations and are often represented as models, written using modelling languages such as BPMN, although [40] shows that different forms of representations are better for different task types. Whereas for some task types though, one finds better result in comprehension when knowledge is represented as visual process models. Comprehension of visual models are also central in many IT-solutions, as exemplified in [22] where company procedures are structured in more than 2000 process models, being the centre of the company quality system. Since people change jobs internally quite often in the company, it is important that they can quickly familiarize themselves with the new work processes to ensure safe, compliant operations, and thus can learn to do the new work process quickly. There is also a close relationship between business process models and operations research (OR) in that business processes provide the framework and context for decision-making [9], while operations research provides the analytical and quantitative tools for maximizing these processes and improving decision-making [19]. The combination of these technologies enables organizations to analyse,

Companion Proceedings of the 17th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling Forum, M4S, FACETE, AEM, Tools and Demos co-located with PoEM 2024, Stockholm, Sweden, December 3-5, 2024

* *Corresponding author.*

john.krogstie@ntnu.no (J. Krogstie); Kshitij.Sharma@ntnu.no (K. Sharma)

0000-0003-4830-1876 (J. Krogstie); 0000-0003-3364-637X (K. Sharma)

optimize, and enhance their operations in a comprehensive manner. Providing a clear and comprehensible visual representation of processes, Business Process Models (BPMs) allow us to understand the flow of activities, identify redundancies, and spot bottlenecks more easily because they provide a clear visual representation [38]. Using this visualization, stakeholders will if being able to comprehend these be able to see the current state of the process on a visual basis, which is the first step towards optimization. As a result of mapping out the entire process, BPMs are able to identify any steps that result in delays, inefficiencies, or additional costs by identifying the faulty steps [41]. As more and more of business processes are automated and potentially taken over fully or partly by AI and other digital systems, it is important to make representations of the processes so that they are understandable by humans, ensuring a more human-centric approach with potential up-skilling and reskilling, being a main focus of Industry 5.0.

Multi-modal data have been used to capture, analyse, and predict learners' behaviour [2], task-performance [10] and learning gains [7] in the area known as multi-modal learning analytics (MMLA). Most of the research carried out for instance to provide feedback to learners is based on measurements rooted in unimodal data streams such as emotions [3], content-based data (using log data, [42]), and attentional data (using gaze data, [43]).

Although techniques such as eye-tracking are used quite a bit for understanding process model comprehension [5, 49], other techniques and sensors for capturing other biometric data than from eye-tracking are only gradually being taken into use including wristband collecting electrodermal activity (EDA) [1, 50].

Section 2 provides some additional background on MMLA and the use of such techniques in connection to modelling. Section 3 presents the experiment. Section 4 presents the data handling for performance prediction. Section 5 discuss the results and look upon the possible usage in neuro-adaptive modelling systems and section 6 revisit our research question and concludes the paper with further work to be done within the approach.

2. Background and related work

MMLA combines several theoretical frameworks [16]. Two of main ones are Cognitive Load Theory (CLT) [36] and the Affective Learning Framework (ALF) [8]. One can look at these in concert with traditional methods to understand comprehension and other processes in for instance process modelling [21, 27]. CLT considers the cognitive processes of learning and retention to explain the learner's behaviour. The working principle of CLT is that humans have a limited information processing capacity when they engage with a cognitive task. To model the learner behaviour using the principles of CLT, one can use EEG and Eye-tracking (ET) data streams. Combining EEG with ET can provide us a broad view of "how learners process the given information to create knowledge and understanding?"

The second theoretical framework supporting our work is the Affective Learning Framework (ALF). ALF is mostly concerned with "how learners feel while they are processing the information presented?" ALF is also concerned with "how learning experiences are internalized?" These learning experiences, once internalized, later guide the learners' attitudes and behaviour in future and in turn, affect the learning outcome in the future.

To incorporate ALF in the methodology, we use facial and physiological (Heart rate variability (HRV), electrodermal activity (EDA), blood volume pulse (BVP), temperature (TMP)) data. Using facial data, one can capture several emotions. Using physiological data (i.e., HRV, BVP, EDA, TMP) one can also detect the emotional states of the learners and their stress and arousal levels. Recently, with the advance in wearable technologies, researchers have been able to compute physiological arousal and stress in various settings [11, 15].

The number of authors publishing what we call neuro-conceptualization (use of neuroscience techniques in conceptual modelling) research is quite small. The thematic focus so far has been on model comprehension, including how e.g. ambiguity influence comprehension [14]. The main technique used here is eye-tracking. Early use of eye-tracking for researching model comprehension is found in [34, 47], but these works focus on capturing area of interest (i.e. what the modeler is looking at). Later, eye-tracking is also used for capturing other characteristics such as cognitive load in process model comprehension [5, 14, 49]. Most of the work is based on understanding operational process models (in functional process models in BPMN [6, 51] or EPC [52]) and declarative process models [1]) There are also examples of analysis of comprehension of UML class diagrams [4], decision models [12], goal models [17] and combinations of rule and process models [46].

The sensor-toolset has lately been extended from eye-tracking by some to include wristbands for capturing for instance EDA [1, 50]. Some use of EEG is also reported [17]. [31] list specifically as one of limitation of current work on process model comprehension that one has not looked at emotional aspects. In [23] an example of the combined use of all four modalities (EEG, ET, facial and physiological) are found in connection to process model comprehension for the first time. This data has also been used for causality analysis between measurements based on data from all four modalities in [24]. In this paper we have looked upon how the data from the same experiment can be used to predict the model comprehension performance based on the biometric data. This information can be used as a basis for the development of neuro-adaptive systems.

Our research group has experience with using similar data following up individual [25] and pair [18,39] programming tasks, which can inspire also mechanism for supporting modelling, depending on what is found to be the most important measurement to predict performance. For the programming tasks, in particular (excessive) cognitive load was important to detect and help rectify e.g. through different feedback mechanisms that should not provide excessive cognitive load in itself. Also, engagement and stress were found important [18]. In pair programming task, joint attention was also important to help to enforce (based on eye-tracking data).

3. Description of experiment

We have in the work presented in this paper collected data from an experiment investigating the comprehension of process models and text-explanations of the same settings. The main research question pursued in this paper is:

RQ: To what extent can business process model comprehension performance be predicted by analysing biometric data collected during the modelling task.

Process model comprehension performance is assessed by the number of correct answers to various comprehension questions answered based on the model presented.

The sensors used were like those used in [28] on studying programming tasks, with the addition of a wristband for capturing physiological data. We briefly describe these below:

1. EEG data: The EEG signals were recorded with a 20-channel ENOBIO device following the international 10–20 system. The raw EEG signal data were recorded at a 500 Hz using a portable EEG cap and divided into the following band powers: delta (below 4 Hz), theta – θ (4–7 Hz), alpha – α (8–12 Hz), and beta – β (18–30 Hz). The Fz electrode in the middle was used as a signal reference electrode, two channels were used for EOG correction, one channel for electric reference, and three Channels Accelerometer with sampling rate at 100 Hz.
2. Gaze data: To record gaze, we used a Tobii X3-120 eye-tracking device at a 120 Hz sampling rate and using a 5-point calibration. The device is non-invasive and mounted at the bottom of a computer screen. The screen resolution was 1920 x 1080 and model interpreters were 50–70 cm away from the screen. All sat on a non-wheeled chair.
3. Facial expression data: To capture face expressions we used LogiTech web camera, pointed straight at the subject from the screen, capturing video at 30 frames-per-second (FPS). The web camera focus zoomed at 150% onto the faces of the modellers. During the tasks, they exhibited a minimal body and gesture interaction; hence, the video recordings hold high quality data from modellers' facial expressions. The video resolution was 640 x 480.
4. The Wristband was positioned at non-dominant hand was an Empatica E4. The active hand was used to move and click on a mouse to respond to comprehension questions. We extracted

the following features: mean, median, variance, skewness, maximum, minimum of (1) Blood volume pressure, (2) Electrodermal activity (EDA), (3) heart rate and (4) Temperature.

The experiment tasks redo parts of the experiment of [40] which focused on the comprehension of different models and texts for different task types. A focus in [40] was to investigate different business process models task types as for appropriateness of textual or BPMN – models. This is not the main focus in the current experiment, due to the experiment setup, we chose a somewhat limited coverage of the tasks used in [40], primarily including comprehension tasks for search and recognition, and inference.

The experimental tasks refer to two separate cases of typical business processes: a Goods Receipt Handling Process (GHP) and a Procure-to-Pay Process (PPP). A BPMN-model of the GHP process is seen in Figure 1. Following is a textual version of this model

"A truck driver registers at the goods receiving department with a delivery note to a goods receipt officer. The officer identifies the delivery type. In this case, it is a delivery related to a purchase order. In case of deliveries without a purchase order, a booking clerk has to be contacted. The booking clerk shall look up the procurement rules before authorizing the delivery to be accepted or not. When the decision has been made, the booking clerk notifies the goods reception officer to execute the acceptance or rejection and records the receiving transaction in a log. Following the assignment of a delivery ramp to the truck driver, the goods are inspected after offloading them. Since the goods inspection proceeds without complaints, the goods are placed into stock. In case of inspection complaints, the goods would have been rejected"

The models were the same as used in [40] reimplemented using the modelling tool Signavio. In [40] they first created a BPMN diagram for each case and then constructed a corresponding text according to the formally defined transformation rules of [26]. We reused also the texts from [40]. The PPP case can be regarded as being more complex than the GHP case. It had more activities, organizational units, and business objects. For the GHP case, the total number of elements (pools, lanes, events, gateways, and activities) in the BPMN model is 24. The text has 136 words. For the PPP case, the BPMN model contains 31 elements, whereas the PPP text is expressed in 255 words. Subjects were randomly assigned and counterbalanced on if they first answered questions related to the text of one of the cases and a BPMN-model of the other.

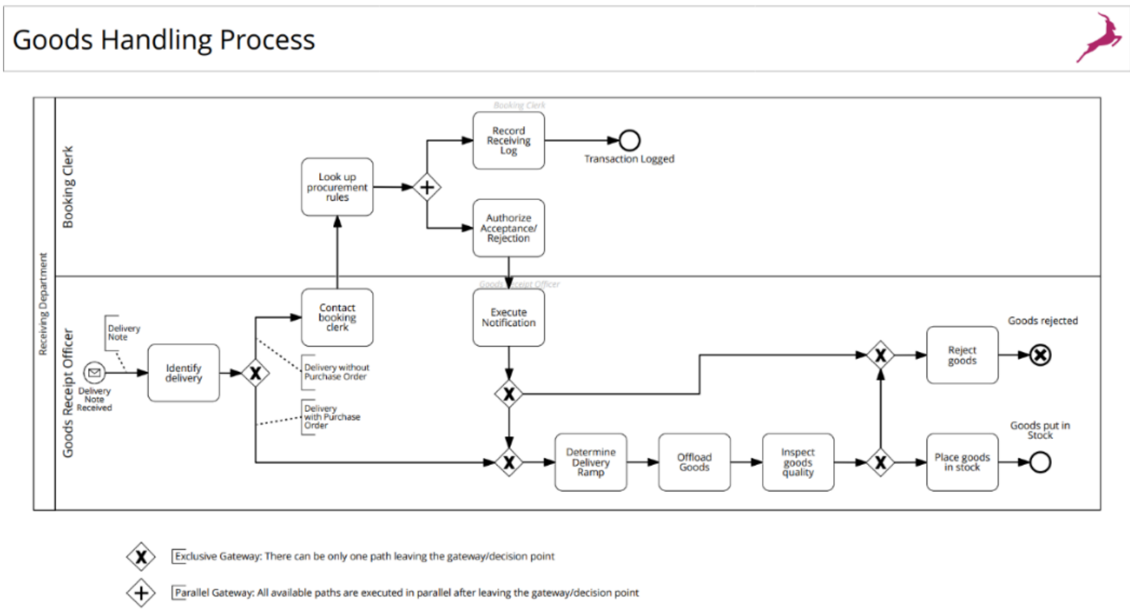

Figure 1: Model of Goods handling process used in experiment

For each case, the participants had to work on two of the four task types in [40] namely Search and recognition tasks, and inference tasks. A total of 44 comprehension questions were given on the two models. After each of the two tests, a NASA TLX [39] form was filled with self-reporting of the subjective experience of the difficulty of the previous task. Before the tasks the informants also gave an indication of their prior knowledge of the domain of the case and text-based, model-based, BPMN-based representation formats. In addition, they reported knowledge of English. The full test-setup is available on github[†], which can be consulted for a complete overview of the experiment.

After getting permission from the national ethical board, the experiment took place. 68 persons (mainly students and employees at the university) took part in the experiment done in the Autumn and Winter of 2023-2024. Data from 57 experiments could be used after preprocessing described below.

4. Ensemble learning from the biometric sensors

The following data was collected and processes from the sensors:

- EEG – per channel: After obtaining data from the electrodes, we normalize it between 0–1 and computed the first 10 Auto-correlation Coefficients. Auto-correlation coefficients depend on the sampling frequency; however, we used the same sampling frequency EEG for all our participants. Therefore, the dependency remains consistent for all the participants. To identify which frequency bands are more important, we computed the Fourier transform of the electrode signals and take the first 10 Coefficients (first 10 dominant frequencies). An Independent Component Analysis (ICA) was used to remove the noise from the jaw movements. We also applied an EOG filter (in-build function in the ENOBIO software for neural data processing) to remove the noise from the blinks and the eye-brow movements, and an additional filter to remove the noise from the tongue movements. A 60 Hz line filter was also used to remove any noise coming from the interference within the EEG wires.
- Video-Face: to extract features from the videos, we extract the facial landmarks and then take the pair-wise distances between the points from the regions.
- Gaze (through eye-tracking): Eye movement data provide the mean, variance, minimum, maximum, and median of several parameters, such as pupil diameters, fixation details, saccade details, blink details, and event statistics. Tobii's default algorithm (i.e., in-build function in the Tobii software for gaze data processing) was used to identify fixations and saccades (for details please see [35]). A filter (i.e., in-build function in the Tobii software) was used to remove the raw gaze points that were classified as blinks.
- Wrist band: Four data streams, HRV, EDA, Skin Temperature, and blood volume pulse (BVP), were analysed using a simple smoothing function to remove any unwanted spikes in the time series that would we based on noise. We used a moving window of 100 samples and overlapping windows of 50 samples between two consecutive windows for the analysis. HRV, BVP, and skin temperature are some of the physiological response data that are susceptible to a wide range of subjective and contextual biases. Time of day, physical health condition, gender, age, overnight sleep, and a variety of other factors can contribute to these biases. To remove the subjective and contextual biases from the data, the first 30 seconds of all four streams of data were used to normalize the remainder of the time series.

4.1. Basic predictions

We divided the dataset into 80% training (44 participants, 1936 questions) and 20% testing (13 participants, 572 questions) sets. We further applied a 10-fold cross-validation on the training set to remove any subjective and selection bias. For prediction, we used an ensemble predictor comprising of eight predictors as illustrated in Figure 2. We combined Support Vector Machines (with linear, radial, and polynomial kernels), Gaussian process models (with linear, radial, and polynomial

---

[†] https://github.com/johnkrog/neurocon2023

kernels), random forest and XGBoost. The final prediction was obtained as a weighted average by using the cross-validation NRMSE (Normalized Root Mean Square Error) as weights. We observe that the final prediction provided an NRMSE of 4.87% on the testing set. This error corresponds to two out of 44 questions in the experiment. In other words, by using the multi-modal data, we can obtain a prediction that is predicting the correctness of answers with a potential error of two questions.

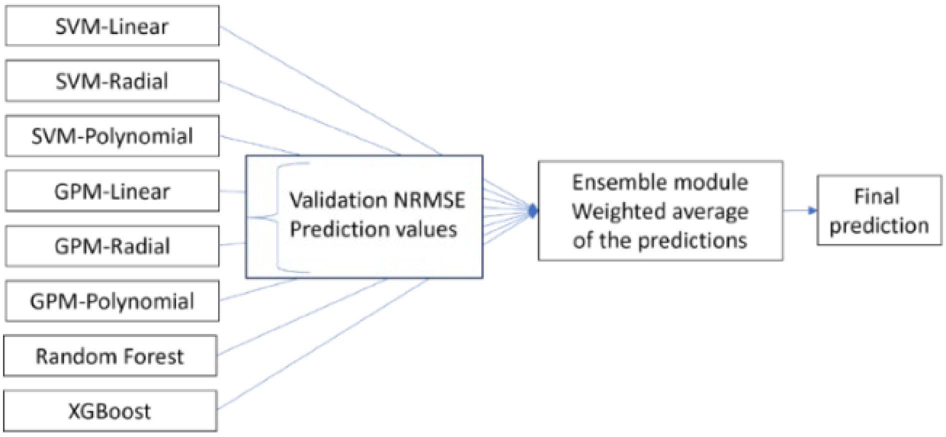

Figure 2: Structure of the ensemble learning applied

We use NRMSE because it is scale-independent since it is normalized by the range of the observed values. As a result, the metric is unaffected by the scale of the data, making it possible to compare datasets with different units or scales. In NRMSE, the error is measured in the same units as the target variable. Since the error is expressed in the original units of the data, stakeholders can easily interpret and communicate the model's accuracy. As NRMSE is normalized, it allows comparisons between models and datasets. Consequently, it is particularly useful when comparing models on datasets of different magnitudes and units. NRMSE is less sensitive to outliers than other metrics like Mean Absolute Error (MAE) or Mean Squared Error (MSE). By penalizing larger errors more heavily, the square can provide a more balanced evaluation of model performance. As a result of its mathematical properties, such as its derivation from the root mean square error, NRMSE is suitable for optimization and statistical analysis. For fine-tuning models and algorithms, it is widely used in optimization algorithms. As a relative measure of error, NRMSE compares the observed values to the range of the error. It is especially useful when comparing models in different contexts or determining the absolute magnitude of the errors.

We use Ensemble learning for several reasons. The accuracy of ensemble methods is often higher than that of individual models. As a result of combining the strengths of multiple models, ensemble learning can compensate for the weaknesses of individual models, which can result in more accurate and robust predictions. Ensembles generalize well and reduce overfitting. During overfitting, a model learns the training data too well, including its noise and outliers. By combining predictions from diverse models, ensembles can provide a more balanced and generalized result. Noise and outliers are less likely to affect ensemble methods. A more robust and stable overall prediction can be achieved by combining the predictions of different models, as they may be affected differently by outliers. Models and algorithms such as decision trees, neural networks, and support vector machines can be learned using ensemble learning. Diverse ensembles can be created based on this versatility, making them adaptable to several types of data and problems. Data patterns may be captured differently by different models. An ensemble method can handle a wide range of data learning patterns by combining them. Compared to complex individual models, ensemble models can be easier to interpret. When simpler models, such as decision trees, are combined, the overall model can be more understandable and easier to interpret. When training data varies slightly, ensembles tend to perform more consistently and have fewer drastic performance changes. In real-world applications, where data distributions may change over time, stability is valuable. In ensemble learning, more diverse models can be added to the ensemble and the learning can be scaled easily. Due to its scalability, it can be applied to large datasets and complex problems. The application of ensemble methods to classification, regression, and clustering has been proven successful. Additionally, they have been used in a variety of fields, including finance, healthcare, and natural language processing.

We used the ensemble learning setup to predict the task-based performance using different combinations of the modalities (sensors) as illustrated in Figure 3. We use the 1) unimodal data-based features, 2) pairs of modalities, 3) triads of modalities. We observe that the model with features from "All" the modalities is not the best predicting set of features. Although, there is no statistical difference between the top 6 (based on the mean NRMSE) we can observe that the eye-tracking based features have the lowest NRMSE, the combination of eye-tracking and facial data has the second lowest NRMSE and the model with features from "All" the modalities has the third lowest NRMSE. Further we observe, while obtaining predictions using features from different combinations of modalities, that there are five distinct groups of combinations where there is no within-group statistical difference in the NRMSE.

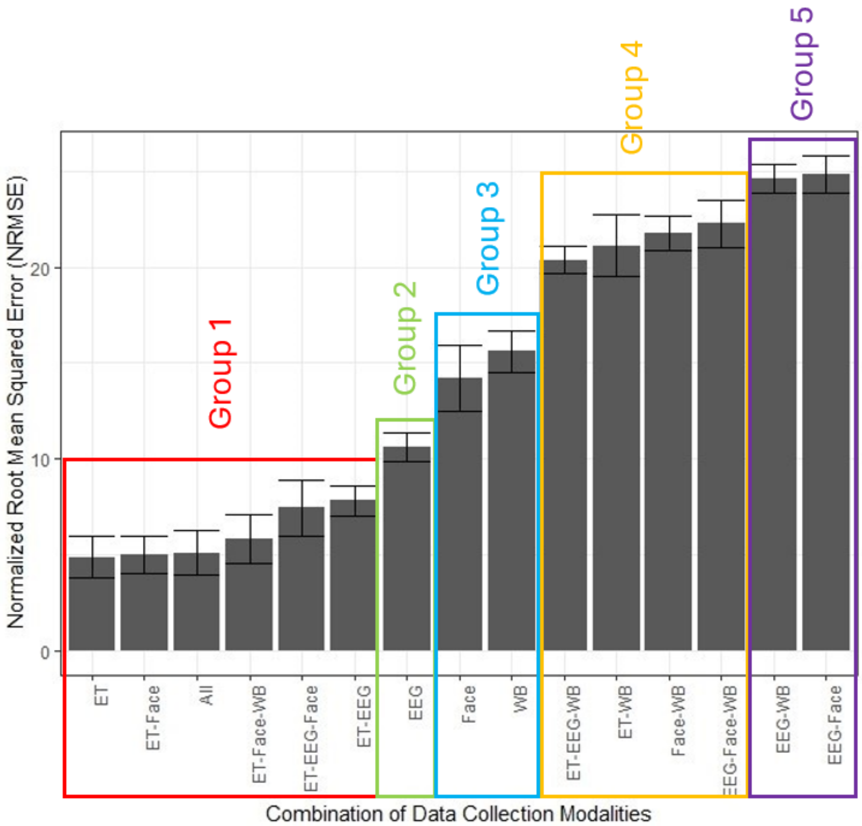

Figure 3: Results with different combinations of modalities and the five distinct groups of combinations where there is no within-group statistical difference in the NRMSE.

1. Group 1: eye-tracking only; eye-tracking and face data; all the modalities; eye-tracking, face data and wristband data; eye-tracking, EEG data and face data; eye-tracking and EEG data.
2. Group 2: EEG data only.
3. Group 3: face data only; wristband data only.
4. Group 4: eye-tracking, EEG data and wristband data; eye-tracking and wristband data; face data and wristband data; EEG data, face data and wrist-band data.
5. Group 5: EEG data and wristband data; EEG data and face data.

The different sensors collect a large number of features. Investigating what features are most involved in the prediction, we have done similar analysis after removing one and one feature to see how much it influences the prediction, and find the following as the most important context independent variables:

- ET:
  - Cognitive load based on pupil dilation
  - Familiarity based on saccade velocity skewness
  - Average attention based on fixation duration
- Face camera
  - Confusion based on relevant AUs (Action Units)
  - Boredom based on relevant AUs (Action Units)
- EEG
  - Convergent thinking based on upper beta band activity (beta band 13-30 Hz)
- Wristband
  - Stress based on heart rate measures
  - Engagement based on EDA

These are features that are candidates also for supporting neuro-adaptive systems, given they are easily captured under modelling activities, which in particular applies to the use of ET and facial camera, and to some extent also wristband-based measures.

4.2. Temporal Predictions

We use the same prediction algorithms and the ensemble learning setup for early prediction. Here we want to find the minimum possible length of the data, in terms of time, that provides the closest NRMSE as compared to the NRMSE when we used the data from all sensors.

In other words, we predicted the dependent variable using partial data. For this, we took 87.5% of the data from each participant and used the methods described above to predict the dependent variable. Then, we keep removing 12.5% (one-eighth) data based on the time up to 12.5% of the data. For each partial dataset we evaluate the prediction performance, and the set of most important features as described in Table 1 and illustrated in Figure 4. This set is chosen based on the features' importance computed from the random forest classifier and has a value of >75 (out of 100).

Table 1
Statistical significance on decrease in performance prediction

Data Length in time	Test NRMSE	Statistical difference from the full-length data prediction	
		T-Value	P-value
100 %	5.09	--	--
87.5%	5.83	-0.74	0.20
75%	6.13	-1.04	0.21
62.5%	6.37	-1.28	0.14
50%	6.94	-1.87	0.06
37.5%	8.60	-3.50	0.0004
25%	9.88	-4.71	0.0001
12%	17.04	-11.94	0.0001

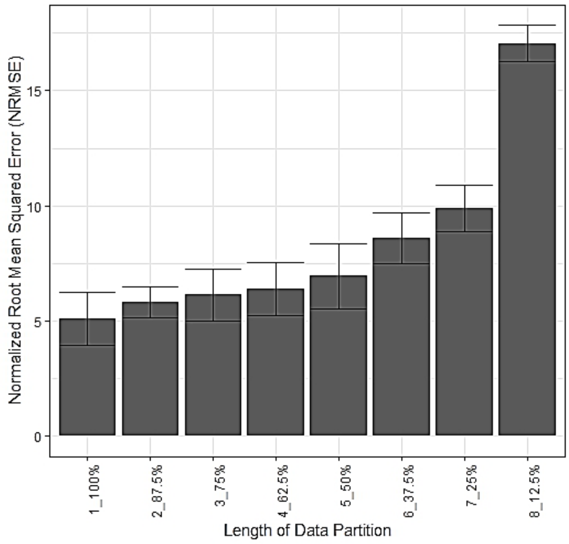

Figure 4: Results with using different parts of the full dataset

We observe that from the 100% data to 50% data, in time, there is no statistical difference in the NRMSE for the different data lengths. Once we reduce the data length from 50% to 37.5%, we observe a significant increase (as compared to 100% data length) in NRMSE for predicting the task-based performance. Further reducing the data by one-eighth length, we see further increment in the NRMSE values, as compared to the 100% data length.

4.3. Predictions based on the task difficulty

We observe from Fig. 5 and Fig 6. that there is no significant difference between the prediction performance if we divide the data into the easiest and the more difficult task. We observe that the five groups of modality-based combinations in the terms of precision are consistent across easy and difficult task. We also observe that the 50% data produces similar predictions for both the difficulty types.

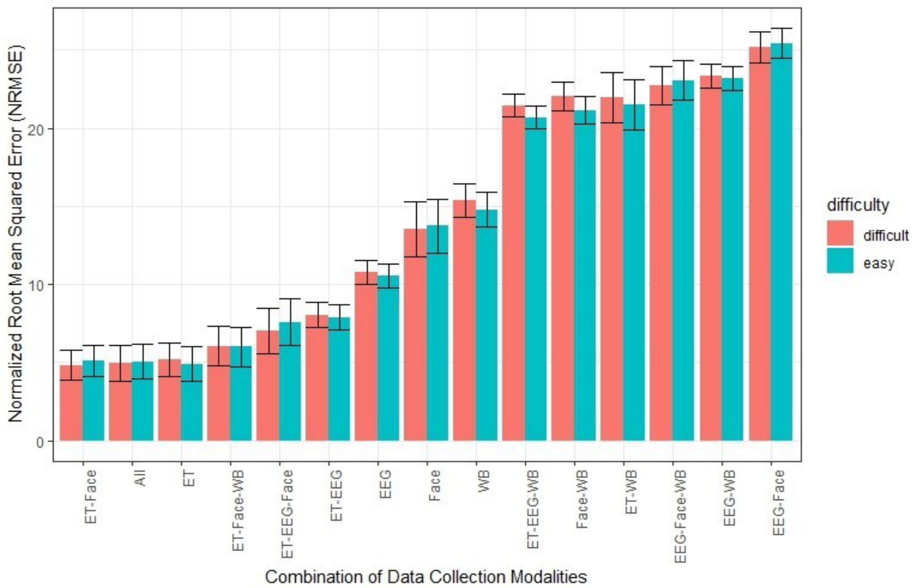

Figure 5: Comparing predictions on the two different tasks

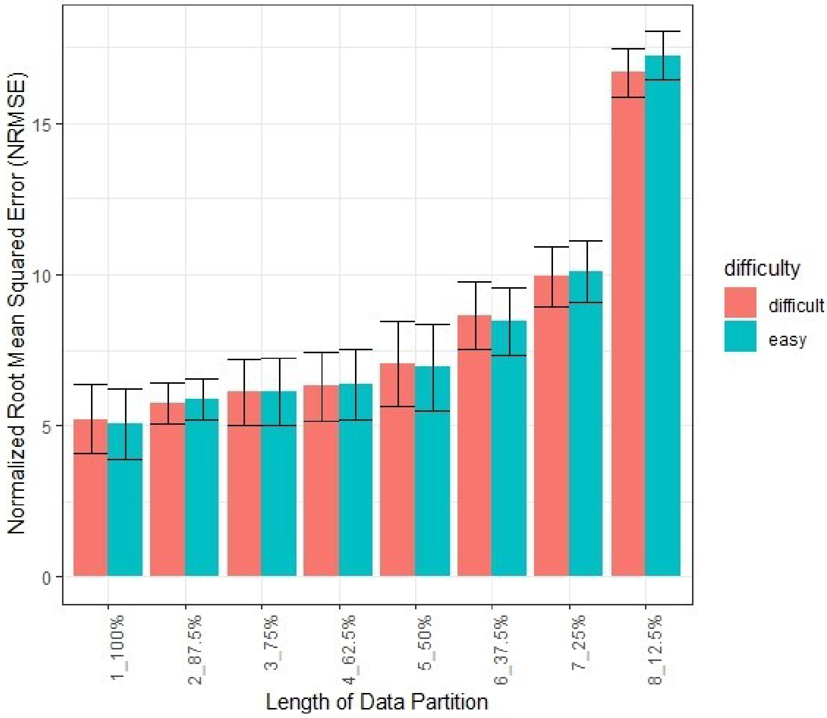

Figure 6: Comparing predictions of different task difficulty using limited part of the data

5. Towards Neuro-adaptive modelling systems

Once we know that there is a high accuracy and precision in predicting model comprehension, then we can design tools to support comprehension with the use of multimodal data. Predicting BPM comprehension brings forth enhanced process clarity by improving understanding and enabling consistency in the processes [13, 30, 31]. By predicting which parts of the process model are likely to be misunderstood, businesses can use comprehension metrics to pinpoint specific areas of confusion. This allows for targeted revisions to improve clarity [20]. Clear and straightforward documentation can be created based on these predictions, helping users to better understand the steps and their purposes. Furthermore, ensuring that all stakeholders interpret the process model consistently minimizes variations in how tasks are performed, leading to uniform outcomes improving coordination and collaboration. This, in turn, creates a foundation for developing standardized training materials that reinforce the correct understanding of processes. Predicting BPM comprehension can also aid in training and development with targeted training and resource allocation [48].

Better BPM understanding also can also support compliance and risk management [45]. Clearer understanding of processes help ensure that all regulatory requirements are understood and adhered to, reducing the risk of non-compliance [37]. It also makes it easier to prepare for audits by ensuring that processes are well-documented and understood.

Similar to work on programming environments, cognitive load appears as an important measure, and one have illustrated in [25] how to capture cognitive load while working in a programming environment, both support situations with low cognitive load (providing help) and high cognitive load. Haugen [18] propose several types of interventions for (too) high cognitive load. In designing interventions that aim to modify cognitive load, introducing visual elements during modelling may indeed reach the goal of decreasing cognitive load, but still not helping comprehension. The intervention could also increase the cognitive load.

- Content based help: Provide relevant solution, e.g. by also looking at what part of the model that is in focus, provide information on the modelling language, for instance a description of the syntax and semantics of a concept, or with other examples of using the same construct. Another possibility is to show a textual version of the part of model in focus that can be automatically generated. On the other hand, this may provide the modeller with too much new information and increase cognitive load too much. As a countermeasure, this content-based help could be tailored to the participant.

- Hints are a type of intervention that are meant to compensate for a non-sufficient level of prior knowledge. They can however have the opposite effect on people with a sufficient base of prior knowledge, and act as distractors, decreasing cognitive load, but at a cost of diverting attention to objects not relating to higher performance.
- In the case of having access to expert gaze patterns this can be used to provide intervention for novice learners. E.g. for areas attended to longer, and with higher cognitive load one could hint to where a novice should aim their gaze.

Haugen also propose intervention for stress and engagement.

- Stress: As stress is a measure of affective state more so than measurable cognitive state, interventions catering to this measure may take the form of encouraging actions outside of the problem solving. The purposes of these interventions would therefore be to encourage the modeller to step away from the problem solving, taking a break.
- Engagement: Concepts that require a higher level of understanding might limit engagement, regardless of effort, and can be a demotivating factor for novices. Similarly for high performer, they might express lower engagement if the concept they are tackling is too basic. A form of intervention that may tackle both of these issues, is an adaptive interventions that scale the complexity of the model e.g. using filtering techniques. Interventions that help to understand progress towards the goal of modelling or visualizes their achievement may help modify their engagement.

6. Conclusion

We have in this paper provided a some of the results from an experiment collecting biometric data on model comprehension tasks, in particular highlighting the use of ensemble learning techniques to support the prediction of performance on model comprehension tasks based on biometric data.

Revisiting our **RQ:** *To what extent can business process model comprehension performance be predicted by analysing bio-metric data collected during the modelling task*,

we show that it is possible to predict the comprehension of business process models with an accuracy that is lower than two wrong questions for each participant.

Interestingly, we find that more data and data sources are not necessarily better with this relatively simple pipeline, thus one might get just as satisfactory results using ET and face data as with the use of EEG in combination with these data sources. Since especially the collection of EEG-data struggle with ecological validity given that it limits movement, it is interesting to see how this makes it possible to also study activities that necessitates more movement with biometric sensors. The prediction results from this study are similar across text and models of different complexity. Since EEG can detect other aspects of the modelling situation though [23, 24], one should still use this when possible and useful in studying the modelling process.

The presented work has several limitations. It has been done in a laboratory setting with quite extensive instrumentation (EEG etc), which limits the ecological validity to some extent. It is also done with students and employees mostly from a computer science department which although had quite different self-reported skill in BPMN is not a general sample of the population likely to need to understand process models or other visual representations. It has also only looked at relatively simple model comprehension tasks, whereas neuro-adaptive modelling environments should also support other modelling tasks, both alone and in groups of modelers and model interpreters.

In future work we will also investigate other modelling task than comprehension using MMLA-techniques. This includes validation tasks, model integration and modelling, both individually and in groups for developing joint models. Model comprehension is a central element also in these tasks; thus, first tool-support will aim at supporting this process. We will also experiment with these kinds of tools outside a lab environment [29], at the first stage experiment with ET, and at a later stage experiment with less intrusive EEG [33] and newer wristband technologies. Using facial cameras in the wild depends on how the use of this will be limited by the AI Act, although the facial landmarks that is collected is not identifying the people. Neither the other biometric data can be used for identifying the person undergoing experiments or usage of the neuro-adaptive tool.

References

[1] Abbad-Andaloussi, A., Burattin , A., Slaats, T., Kindler, E., Weber, B.: Complexity in declarative process models: Metrics and multi-modal assessment of cognitive load. Expert Systems with Applications 233 (2023)

[2] Andrade, A., Danish, J. A., Maltese, A. V.: A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. Journal of Learning Analytics, 4(3), 18-46. (2017)

[3] Baker, R., D'Mello, S. K., Rodrigo, M. M. T., Graesser, A. C.: Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies, 68(4), 223 (2010)

[4] Boot, W.R., Dunn, C.L., Fulmer, B.P., Gerard, G.J., Grabski, S.V.: An eye tracking experiment investigating synonymy in conceptual model validation, International Journal of Accounting Information Systems, Volume 47, 2022,–241 (2010)

[5] Batista Duarte, R.,, Silva da Silveira, D., de Albuquerque Brito, V. Lopes, C.S.: A systematic literature review on the usage of eye-tracking in understanding process models, Business Process Management Journal, Vol. 27 No. 1, pp. 346 (2021)

[6] Boutin, K.D., Davis, C., Hevner, A., Léger, P.M.: Don't overthink it: The paradoxical nature of expertise for the detection of errors in conceptual business process models Frontiers in Neuroscience Volume 16 – (2022)

[7] 7Blikstein, P., Worsley, M.: Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics, 3(2), 220-238 (2016)

[8] D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction 22, 145–157 (2012)

[9] De Maio, C., Fenza, G., Loia, V., Orciuoli, F., Herrera-Viedma, E.: A framework for context-aware heterogeneous group decision making in business processes. *Knowledge-Based Systems*, *102*, 39-50. (2016)

[10] Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., Specht, M.: Learning pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data. In Proceedings of the seventh international learning analytics & knowledge conference (pp. 188-197). ACM (2017)

[11] Di Lascio, E., Gashi, S., Santini, S.: Unobtrusive Assessment of Students' Emotional Engagement During Lectures Using Electrodermal Activity Sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), 1-21 (2018)

[12] Djurica, D., Kummer, T. F., Mendling, J., Figl, K.: Investigating the impact of representation features on decision model comprehension. European Journal of Information Systems, 1–25. (2023)

[13] Figl, K.: Comprehension of procedural visual business process models: a literature review. Business & Information Systems Engineering, 59, 41-67. (2017)

[14] Fransceschetti, M., Abbad-Andaloussi, A., Schreiber, C., Lopez, H. A., Weber, B.: Exploring the Cognitive Effects of Ambiguity in Process Models BPM, Krakow September (2024)
[15] Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., Kostakos, V.: Fitbit for learning: Towards capturing the learning experience using wearable sensing. International Journal of Human-Computer Studies, 136, 102384 (2020)
[16] Giannakos, M., Cukurova, M.: The Role of Learning Theory in Multimodal Learning Analytics. British Journal of Educational Technology (2023)
[17] Gralha, C., Goulão, M., Araujo, J.: Are there gender differences when interacting with social goal models? Empir Software Eng 25, 5416–5453 (2020).
[18] Haugen, Ø. R.: Predicting students' performance with the use of eye tracking and facial recognition to inform design principles for IDE intervention during collaborative debugging. Master Thesis, NTNU, (2023)
[19] Jensen, P. A., Bard, J. F.: Operations research models and methods. John Wiley & Sons (2002)
[20] Jie, L. Y., Wong, D. H. T., Zain, Z. M., Sjarif, N. N. A., Ibrahim, R., Maarop, N.: Metrics and benchmarks for empirical and comprehension focused visualization research in the sales domain. Indonesian Journal of Electrical Engineering and Computer Science, 12(3), 1340-1348 (2018)
[21] Krogstie, J.: Quality in Business Process Modeling, Springer (2016)
[22] Krogstie, J., Heggset, M., Wesenberg, H.: Business Process Modeling of a Quality System in a Petroleum Industry Company Business Process Management Cases Springer (2017)
[23] Krogstie, J., Sharma, K.: Enhancing Our Understanding of Business Process Model Comprehension Using Biometric Data. In: van der Aa, H., Bork, D., Schmidt, R., Sturm, A. (eds) Enterprise, Business-Process and Information Systems Modeling. BPMDS EMMSAD 2024 2024. Lecture Notes in Business Information Processing, vol 511. Springer, Cham. (2024)
[24] Krogstie, J., Sharma, K.: Enhancing Our Understanding of Business Process Model Comprehension Using Biometric Data. In NeuroIS Retreat 2024, Received best, most innovative paper award. Springer, CHAM (2024)
[25] Larsen, E. S., Romskaug, T.: Real time stress-aware feedback system for programming. Master thesis NTNU (2022)
[26] Leopold, H., Mendling, J., Polyvyanyy, A.: Supporting process model validation through natural language generation. IEEE Transactions on Software Engineering 40 (8): 818–840. (2014)
[27] Malinova, M., Mendling, J.: Cognitive diagram understanding and task performance in system analysis and design. Management Information Systems Quarterly 46 (2022)
[28] Mangaroska, K., Sharma, K., Gašević, D., Giannakos, M.: Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity Journal of Computer Assisted Learning (2022)
[29] Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., Li, X., Dix, S., Jaggard, H., Wotherspoon, R.: Lessons learnt from a multimodal learning analytics deployment in-the-wild. ACM Transactions on Computer-Human Interaction 31, 1 (2023)
[30] Maslov, I., Poelmans, S.: Facilitating the comprehension of business process models for unexperienced modelers using token-based animations. Information & Management, 61(5), 103967. (2024)
[31] Mendling, J., Malinova, M.: Experimental evidence on the cognitive effectiveness of diagrams Procedia Computer Science 197, 10-15 (2022)
[32] Mendling, J., Recker, J., Reijers, H. A., Leopold, H.: An empirical review of the connection between model viewer characteristics and the comprehension of conceptual process models. Information Systems Frontiers, 21, 1111-1135 (2019)
[33] Niforatos, E., He, T., Vourvopoulos, A., Giannakos, M.: Democratizing EEG: Embedding Electroencephalography in a Head-Mounted Display for Ubiquitous Brain-Computer Interfacing. Accepted for publication in IJHCI (2024)
[34] Nordbotten, J.C., Crosby, M.E.: The effect of graphic style on data model interpretation. Information Systems Journal, 9: 139-155 (1999)

[35] Olsen, A.: The Tobii I-VT Fixation Filter. Algorithm description. Tobii Technology (2012)
[36] Paas, F., Van Merriënboer, J.J.: Instructional control of cognitive load in the training of complex cognitive tasks. Edu. Psych. review 6, 351–371 (1994)
[37] Reichert, M., Weber, B.: Business process compliance. Enabling Flexibility in Process-Aware Information Systems: Challenges, Methods, Technologies, 297-320 (2012)
[38] Reijers, H. A., Mansar, S. L: Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. *Omega*, *33*(4), 283-306. (2005)
[39] Rimolsrønning, A., Plassen, O.: Eye Tracking Studio – Designing and Evaluating a Feedback System Utilizing Eye-Tracking to Improve Remote Collaboration Between Pair Programmers. Master Thesis, NTNU (2022)
[40] Ritchi, H., Jans, M., Mendling, J., Reijers, H. A.: The Influence of Business Process Representation on Performance of Different Task Types. Journal of Information Systems 1 March; 34 (1): 167–194 (2020)
[41] Rudden, J.: Making the case for BPM: A benefits checklist. BPTrends (2007)
[42] Rus, V., D'Mello, S., Hu, X., Graesser, A.: Recent advances in conversational intelligent tutoring systems. AI magazine, 34(3), 42-54 (2013)
[43] Sharma, K., Caballero, D., Verma, H., Jermann, P., Dillenbourg, P.: Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. International Society of the Learning Sciences. Inc. [ISLS] (2015)
[44] Task Load Index – TLX: https://humansystems.arc.nasa.gov/groups/TLX/ Last accessed 31/7-2024
[45] Van Eijk, K.: Comprehension of Business Process Models: An evaluation of metrics for understandability. Business (2015)
[46] Wang, W., Chen, T., Indulska, M., Sadiq, S., Weber, B.: Business process and rule integration approaches—An empirical analysis of model understanding, Information Systems, Volume 104, (2022)
[47] Weber, B., Pingera, J., Neurater, M., Zugal, S., Martini, M., Furtner, M., Sachse, P., Schnizer, D.: Fixation Patterns During Process Model Creation: Initial Steps Toward Neuro-Adaptive Process Modeling Environments, 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA (2016)
[48] Weinzierl, S., Zilker, S., Dunzer, S., Matzner, M.: Machine learning in business process management: A systematic literature review. Expert Systems with Applications, 124181. (2024)
[49] Winter, M., Neumann, H., Pryss, R., Probst, T., Reichert, M.: Defining gaze patterns for process model literacy – Exploring visual routines in process models with diverse mappings Expert Systems with Applications, Volume 213 (2023)
[50] Winter, M., Bredemeyer, C., Reichert, M., Neumann, H., Pryss, R.: A Comparative Cross-Sectional Study on Process Model Comprehension driven by Eye Tracking and Electrodermal Activity Research Square preprint https://doi.org/10.21203/rs.3.rs-3705553/v1 (2023)
[51] Winter, M., Pryss, R.: The effects of modular process models on gaze patterns - A follow-up investigation about modularization in process model literacy, Expert Systems with Applications, Volume 237, Part A, (2024)
[52] Zimoch, M., Mohring, T., Pryss, R., Probst, T., Schlee, W., Reichert, M.: Using Insights from Cognitive Neuroscience to Investigate the Effects of Event-Driven Process Chains on Process Model Comprehension. In: Teniente, E., Weidlich, M. (eds) Business Process Management Workshops. BPM 2017. Lecture Notes in Business Information Processing, vol 308. Springer (2018)