Summary of "*Utilizing the omnipresent: Incorporating digital documents into predictive process monitoring using deep neural networks.*" - Sergej Levich, Bernhard Lutz, Dirk Neumann

Section 1: Introduction

Challenges presented in the paper:
Today, **Predictive Process Monitoring (PPM)** studies are only focused on the architecture of the underlying predictive model, and most models rely on structured data logs used with deep learning architectures similar to **Long Short-Term Memory (LSTM)** and neural networks.

Objectives:
Excel at operational business processes by reducing the number of labour-intensive and manual activities by implementing PPM, it focuses on process outcome, probabilities of event, and completion of a process at hand.
In this paper, the aim is to use external information found in digital documents like scanned letters, emails, PDF or image files, in order to serve as input data to PPM models. However, the digital documents often have unstructured data which, if processed manually it lengthens the process.
Significant performance improvements have been shown through the example of a German Insurance company, including better prediction of the types of damages and the probabilities of events.

Section 2: Related studies and Overview of PPM

Previous studies have predicted the next event, process time, delays, and outcomes. The PPM models generated **Hidden Markov Models (HMM)**, which are part of a **Markov Chain** (⇒ probabilities of sequences of random variables that are *observable*). The HMM allows us to evaluate both *observable* and *hidden* events.

Recent studies have proposed approaches by different researchers:
Evermann, Mehdiyev, Tax: LSTM models are recurrent neural networks designed to predict a sequence of future events and the remaining time of process completion.
Taymouri: Combined LSTM with generative adversarial networks to predict events and the next one.
Mehdiyev: To reduce event log entries, stacked auto-encoders and a hashing technique could be used.
Heinrich: Advocates for the use of deep learning models with corresponding data.
The authors of the paper: Proposed a **Gated Convolutional Neural Network (GCNN)** and a key–value-predict attention network. This architecture was found to be useful for explaining decisions.

The first group extracted additional information from event attributes found in log data, also known as categorical variables. They encoded Hidden Markov Models (HMM) to determine the positive or negative outcomes of the sequence. By incorporating context information into Long Short-Term Memory (LSTM) models, the quality of predictions improved significantly. The authors also designed a dynamic belief network to specify the relationship between context and control flow.

The second group extracted data from external sources, including both structured and unstructured attributes such as emails, comments, and Internet of Things data to determine context. The results showed an increase in the prediction accuracy of the PPM and less error rate of time completion.

Section 3: Method

Visual extraction, such as images in PDF files, uses the **VGG-16 architecture** to classify them by layers.
Textual extraction of human language is done with the **BERT model (Bidirectional Encoder Representations from Transformers)**, which operates on Next Sentence Prediction and Missing Token Prediction.
Hybrid representation uses both visual and textual features through the **LayoutXLM**, which is designed for document processing tasks.
The **integration module** combines the previous features to apply at each event in the process sequence.
The **predictive model** is based on 2 LSTM layers that generate the predictions of the process, the next event, and the time it takes.

Section 4: Evaluation

The dataset is based on log files and documents from an insurance claims management process of a German company. Their goal is to predict the next event, the time until the next event, and the damage type, in order to improve workflows through better routing and identify fraud through a set of processes.

To measure accuracy, the authors applied a series of tests under the **ten-fold cross-validation** on all methods (see section 3).

Results: Visual and textual features both perform well and provide accurate results, even though textual features yield stronger results since the BERT model outperformed LayoutXLM. Log data is effective at predicting the time between events.

Shapley Additive Explanations (SHAP) values quantify the contribution of each document in the process. It indicates the average change in the probability of the damage type.

Section 5 & 6: Discussion and Conclusion

Digital documents help BPM to yield data flow and offer richer experiences. The findings still offer some limitations so, progress is still yet to be made, despite the promising results shown in this paper.

Levich, S., Lutz, B., & Neumann, D. (2023). *Utilizing the omnipresent: Incorporating digital documents into predictive process monitoring using deep neural networks*. Decision Support Systems, 175, 114043. https://doi.org/10.1016/j.dss.2023.114043