Towards ML-supported Triage Prediction in Real-World Emergency Room Scenarios

Faraz Maschhur[1] **Klaus Netter**[2] **Sven Schmeier**[1] **Katrin Ostermann**[2]
Rimantas Palunis[3] **Tobias Strapatsas**[3,4] **Roland Roller**[1]
[1]German Research Center for Artificial Intelligence (DFKI)
[2]DNC Information Management GmbH
[3]Städtische Kliniken Mönchengladbach
[4]Klinik für Akut- und Notfallmedizin Asklepios Klinikum Harburg

Abstract

In emergency wards, patients are prioritized by clinical staff according to the urgency of their medical condition. This can be achieved by categorizing patients into different labels of urgency ranging from immediate to not urgent. However, in order to train machine learning models offering support in this regard, there is more than approaching this as a multi-class problem. This work explores the challenges and obstacles of automatic triage using anonymized real-world multi-modal ambulance data in Germany.

1 Introduction

The differentiation of treatment urgency is an important step in clinical emergency medicine. Various validated triage systems have been established for this purpose, and in Germany, their use is virtually mandatory. In practice, this means that within the first 10 minutes of a patient's arrival, assigning a treatment priority and thus setting a time target until contact with a medical professional is a required process.

According to the Manchester Triage System (MTS), the possible triage level ranges from immediate to non-urgent, which is mainly meant as guidance to lead employees in the emergency rooms (ER) in making their triage decisions. Although only five triage levels exist, the problem is not as straightforward as it seems. The triage model of MTS follows a decision tree, where on the first level, a so-called diagram or lead symptom (diagnosis) is determined, and on the second level, indications (discriminator) specific to the selected diagram are identified. The indications translate to predefined triage levels, where the most urgent triage level among them expresses the severity of the case. Since the diagrams and indications are not defined by sharp boundaries, it may well happen that a medical professional reaches the same indication through different diagrams. So, the same triage level can be decided on by choosing different, but equally valid indications.

The data used in this work is a mixture of multiple text fields describing the situation of the patient and some first diagnosis (in the form of text), and a large set of structured information, i.e. medical measurement of vital signs, age or sex. In particular, vital signs such as temperature, oxygen saturation, etc. are an essential part of the MTS model defining an indication.

In this work, we have built prototypical machine learning models using retrospective data for automatic triage in the emergency ward and examine the results and obstacles of our approaches. More specifically, we examine to which extent a transformer-based BERT model can address the problem of noisy, unbalanced, semi-structured multi-class real-world data. Different training strategies are explored, particularly to deal with the different interconnected classes as well as to deal with the varying label frequencies. Moreover, we investigate how we can extend a given BERT model, which is normally only suitable for text data, by additional structured information. Finally, we test an approach to build up a hybrid model, combining machine learning with a rule-based component.

2 Related Work

Various studies so far have looked at the possibilities of automatic triage but differ in terms of data, models/solutions, target, and results. Stewart et al. (2023) provide an overview of different triage use cases strongly related to NLP. However, many approaches target, for instance, text (Bergman et al., 2023) or a mix of structured and unstructured (text) data (Klug et al., 2020; Arnaud et al., 2023) to predict a binary label, such as mortality or hospitalization. Some others focus on a larger number of

Proceedings of the 23rd Workshop on Biomedical Language Processing, pages 559–569

triage labels used in emergency care, such as Levin et al. (2018); Sarbay et al. (2023). Depending on the given data, solutions such as gradient-boosting (Klug et al., 2020) or BERT-based approaches (Arnaud et al., 2023), also in a hybrid setup (Wang et al., 2023), are popular. Also, with the rise of large language models, LLM-based solutions have been tested (Frosolini et al., 2024; Levine et al., 2023; Sarbay et al., 2023). So far, however, there are no studies that have automatically determined treatment priorities based on data from emergency services, nor are there any that predict such a large number of different classes simultaneously as we are trying to do.

In this work, we deal with different types of data - (partly sequential) numerical, categorical, and text data. To handle such data types, many different approaches and architectures exist, to combine different 'modalities', such as for instance mapping all different information into one vector space (Rethmeier et al., 2020), the combination of transformers with linear layers and LSTMs (Yang and Wu, 2021; Deznabi et al., 2021), or LLMs with time series (Jin et al., 2023). However, in this work, we rely on a simple architecture based on transformers, integrating different data types and exploring how far we can get.

3 Data and Methods

This work is based on anonymized ambulance reports with triage assignments from a German emergency ward covering two years and including more than 18k cases. The data was recorded electronically and contains a wide variety of different information—overall, about 600 different features exist, ranging from binary to numeric and sequential (e.g., sequences of particular measurements during the ride in the ambulance). The data includes information such as age, sex, blood pressure, pain score, information about consciousness, burns, medications, or motoric skills. In addition to the structured information, the data also includes text fields, describing the emergency situations, an initial diagnosis, injuries, symptoms, as well as the original cause of alarm. An example of a patient case is provided in the Appendix.

The data represents real-world data and has been labeled with the triage categories, consisting of a diagram (diagnosis) and an indication (discriminator), by the emergency department staff in accordance with the MTS. As mentioned, the selected

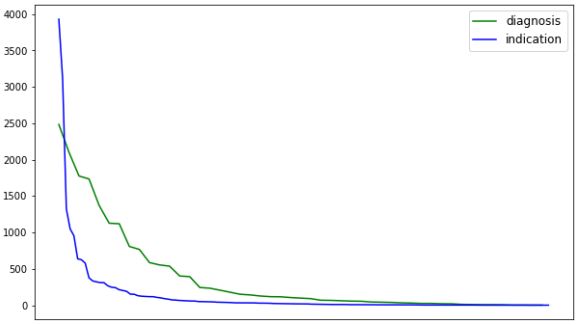

Figure 1: Distribution of diagrams and indications with the two most frequent diagrams *Discomfort in adults* (2480) and *Falls* (2104) and the two most frequent indications *Recent problem* (3931) and *Moderate pain* (3116). For more details see Table 13 and Table 14.

diagram limits the possible indications, and each indication directly translates to a triage level. Several different diagrams and indications may be valid, but only one of each is annotated—in the case of indications, it is the most urgent one. Even if several equally serious indications may apply, only one is labeled. According to MTS, 54 diagrams, 125 indications and 5 triage levels exist. However, due to the real-world context of the data, some labels do not occur in the dataset at all. Figure 1 provides an overview of the label distribution of indications and diagrams in the data. A more detailed overview is provided in the Appendix.

3.1 Data Challenges

Due to the nature of the data properties and the collection procedure, the dataset used in this work poses non-trivial challenges. Since the data has been gathered in the real-world, there is some amount of noise incorporated into the data. The text fields have been filled in by many different paramedics and the abbreviations are not standardized. In a few cases, patients' symptoms resolved between data collection and arrival at the hospital, resulting in a different label than suggested by the data collected. Additionally, the distribution of each of the three labels is unbalanced, with diagnosis and indication having many distinct labels resulting in a long tail problem, as shown in Figure 1 and Tables 13 & 14.

Some of the diagram or indication categories are similar to each other in how they are assessed but differently impact the triage process. Extensive experience guide medical professionals in choosing between these categories. For example, the two indications *Abnormal cardiac history* and *Cardiac*

pain, often only differ in how the medical professional assesses the state of the patient, while being associated with different triage levels.

Moreover, in many cases, the text features are not sufficient for the successful prediction of diagrams or indications. Non-text features like temperature, oxygen saturation and others can be crucial to identify certain diagrams or indications. For instance, the *Very Hot* indication is given at a body temperature above 41°C or *Hyperglycemia* is defined as a glucose level above 17 mmol/l. This limits the model's ability to learn from text features alone since these values are not necessarily included in the text features. Since only a single diagram and a single indication per data point were labeled, although several may apply, the data is less effective in training, as correlations between the diagram or indication classes are not learnable.

3.2 Models

For all our experiments, we rely on medBERT.de (Bressem et al., 2023) and examine different setups, with respect to how we train the model, as well as the input we consider. In Training we first differentiate between using the standard cross-entropy loss (normal), to establish a basic baseline, versus weighted cross-entropy loss (weighted). In addition, we examine models trained independently on each class (single) versus multi-task models (MT) trained on all three classes at the same time. In the MT setup, each class is trained together with the other target classes, and during training the focus (in terms of loss) slowly shifts towards the target class, as depicted in Figure 2.

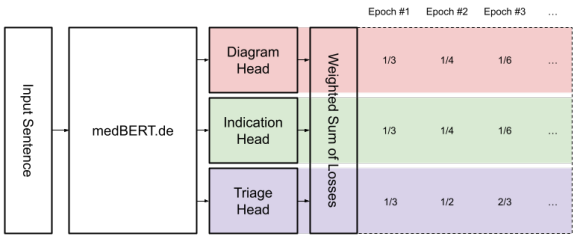

Figure 2: Training of a multi-task model with a focus on triage class - with each epoch the loss contribution optimizes towards triage class

In addition, we test different setups regarding input data: as stated before, not all relevant information for correct classification is given through text data. We therefore examine how additional structured information (pain score, temperature, blood sugar level, heart rate, diastolic/syst. blood pressure, age, sex) could be inserted into the BERT-based solution. In the first setup, we translate structured data into a single sentence using expert knowledge and add these sentences to the standard text data using [SEP] tokens. We refer to this approach as 'extra as text'. For instance, the pain score is translated into a sentence such as '*Pat. has [no/slight/moderate/severe/very severe/the worst imaginable] pain.*'[1] The mapping of numeric values into categories is done by medical guidelines.

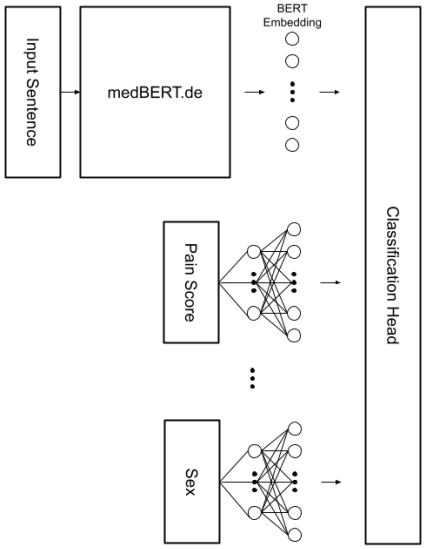

Figure 3: Overview of the architecture in the extra as feature approach - each extra feature is scaled-up through a two-layer MLP and is then inserted, together with the output of medBERT.de, into a classification head

Alternatively, in the second setup, we scale the features through two-layer MLPs and process them in a custom classification head together with the BERT embeddings of the standard text data, as depicted in Figure 3. We refer to this approach as 'extra as feature'. An advantage of this approach is that no bias is introduced through the manual translation into sentences. Since many labels do not occur frequently, the integration of a rule-based component that processes structured information seems helpful in certain scenarios. For this, we examine if an external, rule-based component using expert knowledge targeting vitals could be easily integrated into our system. We refer to this data as 'expert'. This data is also integrated into our model through the use of [SEP] tokens. Every model,

[1] As we work with German we use this translated pattern: '*Pat. hat [keine/leichte/mäßige/starke/sehr starke/stärkste vorstellbare] Schmerzen.*'. More examples can be found in the Appendix.

except the one using the standard cross-entropy loss (normal), incorporates class weights to address the dataset's unbalanced label distribution.

4 Experiments

4.1 Setup

For our experiment, we randomly split the data into training, development, and test sets (80/10/10%). Patient cases that were missing labels were removed, as well as labels that occurred only once. All models have been trained with early stopping and then applied to the test data and evaluated using precision, recall, and F1 (weighted & macro).

4.2 Results

Table 1 presents the weighted and macro F1 scores of different single-class and multi-task models. As the data contains a large number of different classes, which are unbalanced, it is not surprising that macro scores are generally much lower than weighted scores, particularly for indications, which include more than 120 labels. In the single model setup it is difficult to see any additional value of training the BERT model with weighted loss. What we can see, however, is that additional information (extra (text/feat) and expert) seems to have a positive impact on the model performance. In many cases, the impact is particularly visible in the case of macro F1. Most notable here is the inclusion of the simple, expert model.

Table 1: Performance according to F1 weighted (w) and macro (m) of (upper part) single models and (middle and lower part) multi-task models on triage data, including the prediction of the triage label (P) and the deduction of the triage label from the predicted indication (D).

	Diagram		Discrimin.		Triage (P)		Triage (D)	
	w	m	w	m	w	m	w	m
normal	0.592	0.384	0.33	0.102	0.54	0.34	0.542	0.334
weighted	0.607	0.401	0.272	0.12	0.539	0.356	0.528	0.33
extra (text)	0.607	0.414	0.279	0.132	0.542	0.349	0.533	0.338
extra (feat.)	0.587	0.415	0.232	0.133	0.536	0.325	0.513	0.305
expert	0.608	0.416	0.303	0.147	0.55	0.362	0.545	**0.379**
MT weighted	0.588	0.377	0.325	0.145	0.55	0.363	**0.564**	0.354
MT extra-text	0.6	0.411	0.323	0.136	**0.576**	0.379	0.549	0.368
MT extra-feat	**0.613**	0.41	0.316	0.113	0.558	0.392	0.544	0.314
MT expert	0.6	0.415	**0.331**	0.152	0.574	**0.415**	0.554	0.367
MT exp.&ext.-text	0.612	**0.428**	0.328	**0.157**	0.575	0.403	0.552	0.35
MT exp.&ext.-feat	0.599	0.393	0.27	0.121	0.556	0.389	0.548	0.375

Comparing the single and multi-task models, the table shows a clear tendency that multi-task models perform better than the single models. Again, this improvement can be particularly seen in the macro F1 evaluation. More notable (only included in the Appendix), is that our multi-task learning leads to improvements for the given target class. The multi-task models that combine the different expert and extra features generally appear to provide the best approach, especially the model including extra-text.

Table 1 depicts two approaches to predict the triage level: *Triage (P)* represents the direct prediction of triage labels and *Triage (D)* represents the deduction of the triage level from the predicted indication label. In the emergency ward, *Triage (D)* would be the regular way how to solve the problem. In many cases *Triage (P)* provides slightly better results, in terms of weighted and macro F1, compared to the deduction. However, while the direct approach sees triage labels as uncorrelated classes, in reality they are correlated. It certainly makes a difference, given a gold label *red* (immediate), if we predict *orange* (very urgent) or *green* (standard), as orange is closer to red and also more urgent. For this reason, we calculate the MRSE (mean root squared error) using the model *MT expert & extra-text* and for *Triage (P)* achieve a value of 0.588, and for *Triage (D)* a score of 0.525. This indicates that the deduction might be the better choice, as the deduction provides labels closer to the gold label.

Table 2: Top-3 performance according to F1 weighted (w) and macro (m) of a selection of models.

	Diagram		Discrimin.	
	w	m	w	m
normal	**0.86**	0.607	0.572	0.247
MT weighted	0.737	0.517	**0.633**	0.285
MT expert	0.859	0.612	0.628	**0.293**
MT exp. & ext.-text	0.843	0.589	0.624	0.28
MT exp. & ext.-feat	0.836	**0.631**	0.581	0.287

One of the challenges handling this data is that multiple diagram and indication labels can be valid, but only one is annotated. This can have an influence on the performance of our models in case we predict valid labels different to the annotation in the dataset. In order to examine this we evaluate our models by considering the top-3 predictions of diagrams and indications, as depicted in Table 2. The results show, in all cases, a very strong boost in performance, particularly for diagrams. In the case of indication, the weighted score achieves 0.633, while the macro score still remains below 0.3, which might be due to the long tail problem and the fact that many indications require additional structured information.

4.3 Analysis & Discussion

As the data includes a large variety of labels with a long tail problem - and many of the cases occur

only a few times - the task is very challenging. At the same time, many labels do not depend solely on the text features. Therefore, it is difficult to detect them unless the included text contains a clear hint. For instance, for the indication *Suspected Sepsis*, a patient needs to have at least two of the following symptoms: new onset of confusion, increased respiratory rate (above 22/min) or low blood pressure (below 100 mmHg systolic), where the last two symptoms depend on structured data. While our top-3 approach tries to overcome the multiple labels problem, the moderate results for top-3 indications show the limitations of pure text-based approaches for the triage classification problem. We assume that more structured data needs to be included in the model to better deal with labels that are less connected to text data. Moreover, it might be beneficial to include additional rule-based components/predictions, in order to deal with the long tail problem. Data-driven machine learning is popular, but if data is sparse or expensive to gather rule-based components might be a valid approach to overcome its problems.

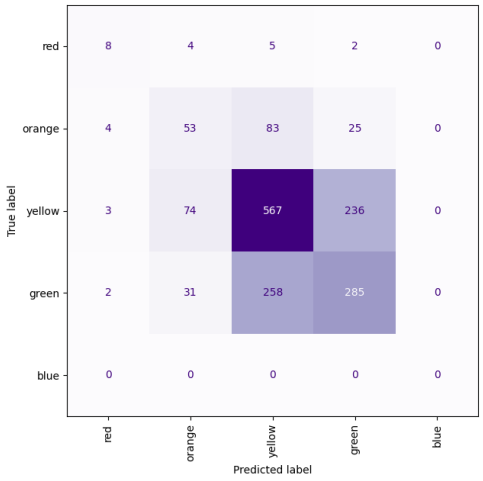

Figure 4: Confusion matrix of top-1 Triage (D) label prediction from red to blue (very urgent to not urgent).

Instead of tackling the problem with a pure text-based transformer approach, we achieve better results by integrating additional data. Text-based integration appears to be more promising than the feature-based approach. Unfortunately, the text-based approach is not scalable, as we need to deal with the model's limited input size. Therefore, the feature-based approach combining BERT embeddings with additional features might be the best approach for using more structured features.

In addition to the missing labels and the existing noise, many labels are generally difficult to predict because they are very abstract, such as *Recent Problem*. According to the definition 'A problem that occurred within the last week is referred to as a recent problem'. Although very general, it is, still one of the most frequent labels in our data, and similar others exist.

While unbalanced data is a problem for machine learning, in a real-world setup for triage prediction, it makes a difference if a patient is accidentally predicted with a triage label that is too high or too low. At the same time, particularly the very urgent classes are most important to predict correctly. Figure 4 depicts the confusion matrix for the top-1 triage label prediction. The figure shows, for instance, that various cases are assigned with a higher triage label and a similar number of cases with a lower triage label, which could risk a patient's life. Even more seriously, various of the patients labeled as red (immediate treatment) are labeled with a lower label. In order to introduce a (hybrid) machine learning system for automatic triage, this is the most important problem to address. Figure 5 (Appendix) shows an alternative confusion matrix when we apply the top-3 indication prediction, infer the triage level, and always choose the most urgent one. This might be a possibility to reduce triage predictions below the gold label. However this approach still offers space for improvements.

5 Conclusion

In this work, we presented a challenging real-world problem to support employees in an emergency ward. Although the data is multi-modal (numerical and text), we approached the problem with text-based transformer solutions. Considering the difficulties with noise, missing labels, the number of different labels, and the long tail problem, the results are promising. However, we foresee that we need to include additional information as extra features to further boost the performance and to provide models with a more substantial benefit in an emergency ward.

Limitations

The presented solution still has many limitations, as presented in the discussion. Naturally noise has some impact on the model's performance, but overall, we also need to investigate how to boost the performance further and particularly examine

how we perform in really urgent cases. While misclassification is negligible for uncritical cases, it is certainly not in very critical ones.

Ethical Statement

Experiments have been conducted on retrospective data. Therefore, our model does not directly impact patient treatments. In the foreseen application, the model is intended to be integrated into an assistance and decision-support system (when good enough), providing additional information for the human performing the actual triage. Where possible, the medical personnel will be provided with explanations and further details corroborating the suggested categorizations.

The project is based on a comprehensive protocol to ensure privacy and data protection. For the model's training and testing, the retrospective data has been completely anonymized and stripped of any personal, local, and temporal information that would allow reference to patients or medical personnel involved.

Acknowledgments

The project has received funding from the German Federal Ministry of Education and Research (BMBF) through the project KIBATIN (16SV9040).

Special thanks to Michael Deppe and Florian Oetke from DNC, who implemented the ICE-Analysis platform and application into which the categorizer is integrated and who were essential in providing us with the training and test material as well as valuable feedback to our approach.

References

Emilien Arnaud, Mahmoud Elbattah, Pedro A Moreno-Sánchez, Gilles Dequen, and Daniel Aiham Ghazali. 2023. Explainable nlp model for predicting patient admissions at emergency department using triage notes. In *2023 IEEE International Conference on Big Data (BigData)*, pages 4843–4847. IEEE.

Erik Bergman, Luise Dürlich, Veronica Arthurson, Anders Sundström, Maria Larsson, Shamima Bhuiyan, Andreas Jakobsson, and Gabriel Westman. 2023. Bert based natural language processing for triage of adverse drug reaction reports shows close to human-level performance. *PLOS Digital Health*, 2(12):e0000409.

Keno K. Bressem, Jens-Michalis Papaioannou, Paul Grundmann, Florian Borchert, Lisa C. Adams, Leonhard Liu, Felix Busch, Lina Xu, Jan P. Loyen, Stefan M. Niehues, Moritz Augustin, Lennart Grosser, Marcus R. Makowski, Hugo JWL. Aerts, and Alexander Löser. 2023. Medbert.de: A comprehensive german bert model for the medical domain. *arXiv preprint arXiv:2303.08179*.

Iman Deznabi, Mohit Iyyer, and Madalina Fiterau. 2021. Predicting in-hospital mortality by combining clinical notes with time-series data. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021*, pages 4026–4031.

Andrea Frosolini, Lisa Catarzi, Simone Benedetti, Linda Latini, Glauco Chisci, Leonardo Franz, Paolo Gennaro, and Guido Gabriele. 2024. The role of large language models (llms) in providing triage for maxillofacial trauma cases: a preliminary study. *Diagnostics*, 14(8):839.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. 2023. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*.

Maximiliano Klug, Yiftach Barash, Sigalit Bechler, Yehezkel S Resheff, Talia Tron, Avi Ironi, Shelly Soffer, Eyal Zimlichman, and Eyal Klang. 2020. A gradient boosting machine learning model for predicting early mortality in the emergency department triage: devising a nine-point triage score. *Journal of general internal medicine*, 35:220–227.

Scott Levin, Matthew Toerper, Eric Hamrock, Jeremiah S Hinson, Sean Barnes, Heather Gardner, Andrea Dugas, Bob Linton, Tom Kirsch, and Gabor Kelen. 2018. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Annals of emergency medicine*, 71(5):565–574.

David M Levine, Rudraksh Tuwani, Benjamin Kompa, Amita Varma, Samuel G Finlayson, Ateev Mehrotra, and Andrew Beam. 2023. The diagnostic and triage accuracy of the gpt-3 artificial intelligence model. medrxiv. *Published online February*, 1(2023):10–1101.

Nils Rethmeier, Necip Oguz Serbetci, Sebastian Möller, and Roland Roller. 2020. Efficare: better prognostic models via resource-efficient health embeddings. In *AMIA Annual Symposium Proceedings*, volume 2020, page 1060. American Medical Informatics Association.

İbrahim Sarbay, Göksu Bozdereli Berikol, and İbrahim Ulaş Özturan. 2023. Performance of emergency triage prediction of an open access natural language processing based chatbot application (chatgpt): A preliminary, scenario-based cross-sectional study. *Turkish Journal of Emergency Medicine*, 23(3):156–161.

Jonathon Stewart, Juan Lu, Adrian Goudie, Glenn Arendts, Shiv Akarsh Meka, Sam Freeman, Katie

Walker, Peter Sprivulis, Frank Sanfilippo, Mohammed Bennamoun, et al. 2023. Applications of natural language processing at emergency department triage: A narrative review. *Plos one*, 18(12):e0279953.

Bing Wang, Weizi Li, Anthony Bradlow, Eghosa Bazuaye, and Antoni TY Chan. 2023. Improving triaging from primary care into secondary care using heterogeneous data-driven hybrid machine learning. *Decision Support Systems*, 166:113899.

Bo Yang and Lijun Wu. 2021. How to leverage multimodal ehr data for better medical predictions? *arXiv preprint arXiv:2110.15763*.

A Appendix

Table 3 to Table 12 depict how the generation of the sentences is conducted in the case of *text as feature*. This categorization is in line with medical guidelines and how different indications are defined (e.g., hypertension).

Table 13 to Table 15 provide an overview of the frequency of the three different labels in our dataset.

Table 16 presents the detailed results of Table 1 above. Table 3 to Table 12 present the medical knowledge used to translate non-text features into text features for the extra as-text models.

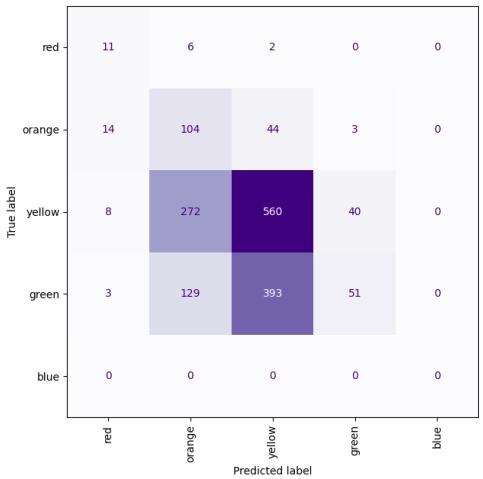

Figure 5: Confusion matrix of top-3 Triage (D) label prediction, where only the most urgent color among the top-3 predictions is counted, from red to blue (very urgent to not urgent).

Table 3: Translation of pain score into a template-based sentence: “Pat. hat ***pain_type*** Schmerzen.” (Pat. has ***pain_type*** pain.)

pain score	0-1	2-3	4-5	6-7	8-9	10
pain_type	“keine” “no”	“leichte” “light”	“mäßige” “moderate”	“starke” “strong”	“sehr starke” “very strong”	“stärkste vorstellbare” “strongest imaginable”

Table 4: Translation of sex value into a template-based sentence: “Pat. ist ***sex_type***.” (Pat. is ***sex_type***.)

sex value	0	1
sex_type	“männlich” “male”	“weiblich” “female”

Table 5: Translation of diastolic value into a template-based sentence: “Pat. hat einen diastolischen Blutdruck von ***X***mmHg.” (Pat. has a diastolic blood pressure of ***X***mmHg.)

diastolic value	X

Table 6: Translation of age into a template-based sentence: “Pat. ist ein ***age_type*** im Alter von ***age***.” (Pat. is a ***age_type*** in the age of ***age***.)

age	$\leq$1	>1 & $\leq$3	>3 & <18	$\geq$18 & $\leq$40	>40 & $\leq$65	>65
age_type	Baby baby	Kleinkind toddler	Kind child	Erwachsener adult	Erwachsener mittleren Alters middle-aged adult	Senior senior

Table 7: Translation of pulse value into a template-based sentence: “Pat. hat einen ***pulse_type*** Puls von ***pulse value***.” (Pat. has a ***pulse_type*** pulse of ***pulse value***.)

pulse value	$\leq$60	>60 & <100	$\geq$100 & $\leq$120	>120
pulse_type	“zu niedrigen” “too low”	“normalen” “normal”	“erhöhten” “elevated”	“stark erhöhten” “highly elevated”

Table 8: Translation of temperature value into a template-based sentence: “Pat. ist ***temp_type*** mit einer Körpertemperatur von ***temp value*** Grad Celsius.” (Pat. is ***temp_type*** with a body temperature of ***temp value*** degrees Celsius.)

temperature value	$\leq$35	>35 & <37.5	$\geq$37.5 & <38.5	$\geq$38.5 & <41	$\geq$41
temp_type	“unterkühlt” “undercooled”	“normal” “normal”	“überwärmt” “overheated”	“heiß” “hot”	“sehr heiß” “very hot”

Table 9: Translation of spo2 value into a template-based sentence: “Pat. hat eine ***spo2_type*** Sauerstoffsättigung von ***spo2 value***%.” (Pat. has a ***spo2_type*** oxygen saturation of ***spo2 value***%.)

spo2 value	<90	$\geq$90 & <95	$\geq$95
spo2_type	“sehr niedrige” “very low”	“niedrige” “low”	“normal” “normal”

Table 10: Translation of blood sugar value into a template-based sentence: “Pat. hat einen ***bs_type*** Blutzuckerspiegel von ***bs value***mg/dl.” (Pat. has a ***bs_type*** blood sugar level of ***bs value***mg/dl.)

bs value	$\leq$54	>54 & <70	$\geq$70 & $\leq$100	>100 & <306	$\geq$306
bs_type	“zu niedrigen” “too low”	“niedrigen” “low”	“normalen” “normal”	“erhöhten” “increased”	“zu hohen” “too high”

Table 11: Translation of systolic value into a template-based sentence: “Pat. hat einen ***systolic_type*** systolischen Blutdruck von ***systolic value***mmHg.” (Pat. has a ***systolic_type*** systolic blood pressure of ***systolic value***mmHg.)

systolic value	<90	≥90 & <100	≥100 & ≤120	>120 & ≤140	>140
systolic_type	“zu niedrigen” “too low”	“niedrigen” “low”	“normalen” “normal”	“hohen” “high”	“zu hohen” “too high”

Table 12: Translation of heart frequency value into a template-based sentence: “Pat. hat eine ***hf_type*** Herzfrequenz von ***heart frequency value***.” (Pat. has a ***hf_type*** heart frequency of ***heart frequency value***.)

heart frequency value	<40	≥40 & ≤60	>60 & ≤100	>100 & <140	≥140 & <160	≥160
hf_type	“zu niedrige” “too low”	“niedrige” “low”	“normale” “normal”	“hohe” “high”	“zu hohe” “too high”	“extrem hohe” “extremely high”

Table 13: Frequency of diagram (diagnosis) labels in the dataset

diagram label		# in dataset
Unwohlsein bei Erwachsenen	Discomfort in adults	2480
Stürze	Falls	2104
Extremitätenprobleme	Limb problems	1735
Atemproblem bei Erwachsenen	Respiratory problem in adults	1369
Abdominelle Schmerzen bei Erwachsenen	Abdominal pain in adults	1123
Thoraxschmerz	Thoracic pain	1120
Kopfverletzung	Head injury	808
Urologisches Problem	Urological problem	765
Wunden	Wounds	586
Herzklopfen	Palpitations	554
Kollaps	Collapse	537
Rückenschmerz	Back pain	402
Betrunkener Eindruck	Drunken impression	392
Durchfälle und Erbrechen	Diarrhea and vomiting	244
Generelle Indikatoren	General indicators	234
Gastrointestinale Blutung	Gastrointestinal bleeding	207
Angriff (Zustand nach)	Attack (condition after)	179
Überdosierung und Vergiftung	Overdose and poisoning	143
Körperstammverletzung	Trunk injury	142
Schweres Trauma	Severe trauma	127
Diabetes	Diabetes	118
Nackenschmerz	Neck pain	116
Allergie	Allergy	106
Auffälliges Verhalten	Abnormal behavior	98
Kopfschmerz	Headache	89
Besorgte Eltern	Concerned parents	68
Atemproblem bei Kindern	Breathing problem in children	67
Selbstverletzung	Self-harm	57
Krampfanfall	Seizure	56
Psychiatrische Erkrankung	Psychiatric illness	48
Abdominelle Schmerzen bei Kindern	Abdominal pain in children	45
Unwohlsein bei Kindern	Malaise in children	41
Abszesse und lokale Infektionen	Abscesses and local infections	38
Bisse und Stiche	Bites and stings	32
Verbrennungen und Verbrühungen	Burns and scalds	30
Fremdkörper	Foreign bodies	24
Gesichtsprobleme	Facial problems	24
Asthma	Asthma	21
Hodenschmerz	Testicular pain	20
Halsschmerz	Sore throat	12
Hautausschläge	Skin rashes	9
Unwohlsein bei Neugeborenen	Discomfort in newborns	7
Chemikalienkontakt	Chemical contact	7
Augenprobleme	Eye problems	6
Vaginale Blutung	Vaginal bleeding	3
Unwohlsein bei Säuglingen	Discomfort in infants	2
Ohrenprobleme	Ear problems	2

Table 14: Frequency of indication (discriminator) labels in the dataset

indication label	# in dataset
Recent problem	3931
Moderate pain	3116
Unstoppable minor bleeding	1322
Recent mild pain	1049
Low O2 saturation	625
Rapid onset	579
Report of unconsciousness	377
Abnormal cardiac history	334
Inappropriate history	321
New abnormal pulse	311
Altered state of consciousness can be fully explained by alcohol consumption	267
Cardiac pain	248
Swelling	242
Hot	214
–None	211
Gross misalignment	203
Persistent palpitations	191
Severe pain	152
Very low O2 saturation	151
Tarry stools or fresh blood accumulation	131
Altered state of consciousness	124
Colicky pain	121
Vomiting	118
Urinary retention	118
Conspicuous injury mechanism	108
Tendency to bleed	91
Pleural pain	84
Macrohematuria	74
Shock	71
Overheated	66
Abnormal psychiatric history	63
Wheezing	61
Report of acute vomiting of blood	58
Conspicuous hematological or metabolic anamnesis	58
Fresh neurological deficit	51
Suspected sepsis	50
Moderate risk of (future) self-harm	48
Cannot speak in complete sentences	48
Hyperglycemia	46
Inadequate breathing	42
Signs of dehydration	41
Compromised airway	39
Fresh or old blood stools	38
State of exhaustion	35
High risk of (future) self-harm	33
Moderate pain or itching	33
Conspicuous respiratory history	33
Direct neck trauma	32
Recent injury	32
Scalp hematoma	32
New state of confusion	29
Noticeable restlessness	29
Productive cough	28
Unstoppable major bleeding	27
Local infection	24
Vomiting of blood	23
Malposition	22
Dysuria	20
Unable to walk	19
Acute neurological deficit	19
Hypoglycemia	19
Smoke exposure	19
Local inflammation	18
Hypothermia	18
Recent mild pain or itching	15
Direct back trauma	13
Altered state of consciousness	12
Persistent vomiting	12
Extensive secretions or vesicle formation	11
Inhalation trauma	10
Impaired (distal) circulation	10
Facial edema	10
Scrotal swelling/redness	10
Low peak flow	8
Known or suspected immunosuppression	8
Acute respiratory distress	8
Moderate lethality	8
Report of overdose or intoxication	8
Inadequate history (of alcohol consumption)	7
Hyperglycemia with ketosis	7
Abnormal history of GI bleeding	6
Life-threatening hemorrhage	6
Report of head injury	6
Persistent seizure	5
Tongue edema	5
Electrical accident	5
No response to own asthma medication	5
Radiation of pain into the shoulder	5
Critical skin condition	5
Open fracture	5
Very low peak flow	5
Very hot	4
Pain radiating to the back	4
Overheated joint	3
Stridor	3
Moderately lethal animal bite	3
...	

Table 15: Frequency of triage labels in the dataset

triage label	# in dataset
red	187
orange	1551
yellow	8785
green	5684
blue	190

Table 16: Performance Metrics Top-3

	Diagnose						Indication						Triage (direct)						Triage (indirect)					
	recall		precision		F1		recall		precision		F1		recall		precision		F1		recall		precision		F1	
	w	m	w	m	w	m	w	m	w	m	w	m	w	m	w	m	w	m	w	m	w	m	w	m
MT expert dia	0.621	0.412	0.605	0.455	0.6	0.415	0.344	0.11	0.279	0.201	0.251	0.126	0.587	0.419	0.56	0.432	0.566	0.418	0.541	0.388	0.551	0.363	0.533	0.365
top2	0.795	0.552	0.792	0.587	0.791	0.561	0.541	0.209	0.437	0.302	0.427	0.219	0.884	0.712	0.887	0.639	0.884	0.664	0.509	0.326	0.468	0.358	0.433	0.298
top3	0.861	0.622	0.862	0.625	0.859	0.612	0.646	0.304	0.54	0.386	0.541	0.307	0.976	0.918	0.977	0.833	0.976	0.868	0.502	0.298	0.394	0.34	0.359	0.241
MT expert tri	0.599	0.377	0.571	0.45	0.565	0.394	0.367	0.106	0.291	0.171	0.264	0.111	0.579	0.45	0.573	0.399	0.574	0.415	0.55	0.399	0.567	0.334	0.545	0.351
top2	0.789	0.563	0.772	0.597	0.774	0.561	0.551	0.175	0.466	0.263	0.456	0.188	0.889	0.744	0.896	0.621	0.89	0.667	0.516	0.321	0.485	0.34	0.446	0.291
top3	0.863	0.67	0.856	0.672	0.857	0.658	0.647	0.294	0.572	0.367	0.569	0.294	0.979	0.95	0.979	0.821	0.978	0.871	0.479	0.298	0.426	0.356	0.377	0.262
MT expert & extra (feat.) dia	0.604	0.379	0.601	0.42	0.599	0.393	0.319	0.105	0.255	0.185	0.23	0.115	0.579	0.404	0.541	0.439	0.55	0.413	0.531	0.341	0.54	0.339	0.527	0.335
top2	0.778	0.554	0.779	0.576	0.776	0.551	0.532	0.205	0.425	0.298	0.421	0.214	0.876	0.709	0.876	0.66	0.876	0.681	0.52	0.33	0.484	0.374	0.454	0.311
top3	0.839	0.648	0.839	0.636	0.836	0.631	0.633	0.268	0.527	0.346	0.533	0.27	0.973	0.915	0.973	0.825	0.972	0.862	0.541	0.325	0.428	0.363	0.398	0.267
MT expert & extra (text) tri	0.627	0.405	0.593	0.474	0.591	0.422	0.357	0.106	0.301	0.186	0.269	0.121	0.586	0.418	0.571	0.408	0.575	0.403	0.542	0.41	0.56	0.338	0.535	0.354
top2	0.798	0.526	0.783	0.591	0.785	0.545	0.549	0.177	0.47	0.27	0.453	0.193	0.893	0.727	0.898	0.642	0.894	0.676	0.522	0.355	0.496	0.362	0.453	0.318
top3	0.868	0.611	0.862	0.646	0.862	0.616	0.646	0.267	0.575	0.348	0.568	0.274	0.98	0.944	0.981	0.804	0.98	0.855	0.531	0.337	0.426	0.367	0.384	0.274
MT expert & extra (feat.) disc	0.602	0.39	0.572	0.46	0.575	0.404	0.335	0.113	0.283	0.169	0.27	0.121	0.569	0.397	0.544	0.419	0.549	0.403	0.548	0.409	0.561	0.359	0.548	0.375
top2	0.77	0.508	0.747	0.57	0.751	0.516	0.528	0.198	0.459	0.263	0.46	0.207	0.887	0.75	0.891	0.675	0.888	0.706	0.544	0.364	0.52	0.385	0.495	0.346
top3	0.833	0.578	0.821	0.62	0.822	0.582	0.628	0.277	0.576	0.343	0.581	0.287	0.967	0.92	0.968	0.846	0.967	0.877	0.506	0.318	0.448	0.374	0.411	0.287
MT expert disc	0.547	0.352	0.51	0.416	0.503	0.363	0.353	0.16	0.335	0.179	0.331	0.152	0.561	0.4	0.546	0.396	0.55	0.393	0.551	0.385	0.565	0.357	0.554	0.367
top2	0.709	0.455	0.671	0.519	0.673	0.468	0.517	0.242	0.505	0.252	0.5	0.225	0.869	0.669	0.877	0.6	0.871	0.626	0.529	0.344	0.517	0.358	0.493	0.33
top3	0.782	0.528	0.755	0.565	0.757	0.526	0.639	0.317	0.636	0.31	0.628	0.293	0.953	0.845	0.955	0.807	0.953	0.824	0.532	0.333	0.47	0.353	0.419	0.283
MT extra (feat.) tri	0.605	0.367	0.575	0.442	0.563	0.384	0.349	0.103	0.291	0.18	0.254	0.113	0.582	0.45	0.552	0.402	0.558	0.392	0.527	0.442	0.544	0.314	0.52	0.333
top2	0.774	0.496	0.754	0.554	0.756	0.507	0.544	0.195	0.454	0.27	0.435	0.198	0.895	0.753	0.894	0.66	0.893	0.681	0.521	0.34	0.484	0.335	0.451	0.293
top3	0.855	0.626	0.846	0.65	0.848	0.625	0.632	0.259	0.547	0.337	0.535	0.26	0.983	0.957	0.983	0.818	0.981	0.858	0.519	0.329	0.423	0.354	0.384	0.269
MT expert & extra (feat.) tri	0.602	0.362	0.567	0.445	0.559	0.381	0.361	0.096	0.293	0.173	0.267	0.108	0.574	0.48	0.553	0.384	0.556	0.389	0.556	0.509	0.566	0.331	0.543	0.352
top2	0.788	0.538	0.769	0.59	0.772	0.545	0.541	0.189	0.452	0.272	0.435	0.195	0.888	0.781	0.891	0.629	0.888	0.674	0.517	0.341	0.487	0.347	0.448	0.301
top3	0.86	0.629	0.848	0.649	0.85	0.62	0.627	0.254	0.546	0.337	0.536	0.257	0.982	0.97	0.982	0.826	0.981	0.877	0.501	0.329	0.429	0.356	0.385	0.275
MT expert & extra (text) disc	0.574	0.376	0.538	0.452	0.539	0.392	0.363	0.15	0.327	0.183	0.328	0.157	0.561	0.393	0.548	0.39	0.55	0.389	0.552	0.363	0.565	0.345	0.552	0.35
top2	0.741	0.482	0.716	0.535	0.721	0.493	0.53	0.237	0.5	0.263	0.501	0.234	0.879	0.688	0.885	0.613	0.88	0.642	0.546	0.334	0.519	0.343	0.493	0.311
top3	0.821	0.602	0.81	0.623	0.812	0.6	0.641	0.292	0.625	0.309	0.624	0.28	0.959	0.838	0.96	0.811	0.959	0.819	0.537	0.324	0.461	0.344	0.41	0.267
MT extra (text) disc	0.541	0.34	0.48	0.377	0.477	0.337	0.344	0.159	0.324	0.143	0.323	0.136	0.563	0.422	0.52	0.378	0.52	0.381	0.546	0.425	0.555	0.344	0.549	0.368
top2	0.707	0.44	0.646	0.469	0.655	0.437	0.531	0.266	0.523	0.24	0.514	0.233	0.853	0.679	0.851	0.605	0.846	0.612	0.545	0.408	0.543	0.394	0.537	0.394
top3	0.784	0.527	0.738	0.551	0.747	0.518	0.655	0.344	0.654	0.29	0.64	0.292	0.937	0.839	0.938	0.757	0.935	0.768	0.529	0.336	0.49	0.362	0.445	0.303
MT extra (text) tri	0.615	0.407	0.579	0.48	0.577	0.421	0.348	0.107	0.295	0.169	0.266	0.112	0.582	0.419	0.576	0.368	0.576	0.379	0.55	0.418	0.57	0.32	0.546	0.342
top2	0.8	0.548	0.782	0.594	0.785	0.553	0.549	0.201	0.465	0.263	0.456	0.201	0.884	0.716	0.893	0.583	0.886	0.627	0.521	0.351	0.496	0.358	0.457	0.318
top3	0.87	0.607	0.863	0.642	0.863	0.61	0.653	0.29	0.577	0.33	0.576	0.276	0.979	0.951	0.98	0.801	0.978	0.848	0.509	0.32	0.43	0.35	0.381	0.264
MT expert & extra (text) dia	0.621	0.42	0.615	0.462	0.612	0.428	0.333	0.107	0.257	0.19	0.237	0.119	0.569	0.386	0.546	0.384	0.552	0.378	0.535	0.345	0.552	0.318	0.534	0.325
top2	0.799	0.56	0.802	0.553	0.797	0.541	0.555	0.201	0.438	0.277	0.444	0.208	0.877	0.709	0.884	0.597	0.879	0.635	0.524	0.343	0.493	0.353	0.468	0.316
top3	0.845	0.61	0.848	0.589	0.843	0.589	0.651	0.298	0.557	0.365	0.565	0.299	0.976	0.955	0.976	0.806	0.975	0.859	0.507	0.323	0.427	0.356	0.388	0.274
MT extra (text) dia	0.606	0.403	0.601	0.444	0.6	0.411	0.327	0.095	0.256	0.184	0.232	0.108	0.563	0.404	0.54	0.372	0.545	0.371	0.529	0.335	0.546	0.312	0.529	0.318
top2	0.789	0.541	0.794	0.539	0.788	0.529	0.557	0.185	0.445	0.285	0.446	0.198	0.873	0.707	0.877	0.618	0.874	0.651	0.537	0.365	0.49	0.359	0.462	0.323
top3	0.847	0.641	0.851	0.605	0.845	0.605	0.646	0.262	0.544	0.349	0.551	0.269	0.973	0.942	0.974	0.815	0.973	0.864	0.534	0.333	0.415	0.352	0.384	0.267
MT extra (feat.) dia	0.621	0.405	0.616	0.447	0.613	0.41	0.362	0.107	0.249	0.183	0.229	0.11	0.578	0.421	0.526	0.376	0.529	0.37	0.518	0.347	0.537	0.296	0.519	0.306
top2	0.805	0.58	0.796	0.579	0.793	0.548	0.564	0.196	0.413	0.283	0.421	0.198	0.874	0.725	0.876	0.655	0.874	0.682	0.524	0.368	0.496	0.353	0.467	0.325
top3	0.851	0.661	0.847	0.629	0.843	0.623	0.644	0.246	0.51	0.334	0.524	0.249	0.973	0.922	0.974	0.834	0.973	0.871	0.508	0.339	0.442	0.357	0.409	0.292
MT weighted dia	0.587	0.368	0.598	0.399	0.588	0.377	0.336	0.112	0.249	0.177	0.233	0.113	0.562	0.517	0.549	0.348	0.548	0.369	0.528	0.369	0.549	0.296	0.527	0.31
top2	0.716	0.495	0.727	0.502	0.717	0.485	0.523	0.193	0.396	0.274	0.4	0.194	0.886	0.767	0.895	0.548	0.885	0.595	0.515	0.363	0.501	0.339	0.463	0.314
top3	0.737	0.535	0.747	0.526	0.737	0.517	0.629	0.26	0.519	0.342	0.53	0.26	0.975	0.985	0.974	0.781	0.971	0.836	0.53	0.345	0.426	0.336	0.38	0.264
MT weighted tri	0.62	0.401	0.583	0.464	0.581	0.408	0.337	0.106	0.282	0.185	0.247	0.118	0.561	0.426	0.547	0.358	0.55	0.363	0.54	0.44	0.555	0.314	0.532	0.33
top2	0.79	0.534	0.77	0.579	0.772	0.534	0.552	0.196	0.458	0.278	0.445	0.204	0.885	0.778	0.889	0.597	0.884	0.643	0.526	0.387	0.495	0.341	0.453	0.308
top3	0.856	0.647	0.847	0.657	0.848	0.633	0.648	0.272	0.562	0.35	0.556	0.279	0.98	0.956	0.981	0.794	0.98	0.848	0.519	0.35	0.432	0.35	0.386	0.274
expert	0.617	0.409	0.612	0.453	0.608	0.416	0.358	0.142	0.302	0.178	0.303	0.147	0.571	0.601	0.551	0.348	0.55	0.362	0.541	0.383	0.549	0.375	0.545	0.379
top2	0.807	0.57	0.807	0.548	0.804	0.546	0.53	0.216	0.473	0.253	0.481	0.219	0.904	0.861	0.909	0.542	0.898	0.577	0.52	0.323	0.488	0.369	0.476	0.318
top3	0.863	0.622	0.868	0.583	0.863	0.59	0.648	0.272	0.608	0.297	0.615	0.267	0.983	0.964	0.983	0.77	0.981	0.829	0.499	0.316	0.441	0.386	0.406	0.29
expert & extra (feat.)	0.604	0.388	0.601	0.428	0.595	0.393	0.338	0.121	0.315	0.148	0.316	0.122	0.525	0.417	0.54	0.333	0.529	0.36	0.531	0.356	0.545	0.34	0.537	0.347
top2	0.783	0.522	0.788	0.511	0.782	0.504	0.529	0.213	0.512	0.238	0.51	0.208	0.684	0.699	0.692	0.584	0.682	0.624	0.532	0.341	0.517	0.369	0.497	0.334
top3	0.824	0.572	0.829	0.549	0.822	0.547	0.642	0.285	0.634	0.283	0.628	0.261	0.714	0.776	0.72	0.706	0.71	0.729	0.537	0.322	0.48	0.36	0.425	0.275
weighted	0.611	0.41	0.61	0.426	0.607	0.401	0.344	0.114	0.282	0.16	0.272	0.12	0.541	0.507	0.552	0.325	0.539	0.356	0.522	0.35	0.54	0.318	0.528	0.33
top2	0.743	0.538	0.74	0.52	0.736	0.506	0.546	0.205	0.472	0.244	0.478	0.202	0.688	0.741	0.694	0.572	0.683	0.63	0.528	0.34	0.513	0.354	0.488	0.324
top3	0.774	0.596	0.772	0.572	0.768	0.56	0.662	0.282	0.613	0.303	0.62	0.27	0.709	0.783	0.716	0.731	0.705	0.743	0.518	0.317	0.457	0.357	0.412	0.276
MT weighted disc	0.606	0.4	0.567	0.433	0.576	0.402	0.366	0.138	0.33	0.186	0.325	0.145	0.57	0.4	0.538	0.353	0.542	0.355	0.561	0.429	0.574	0.334	0.564	0.354
top2	0.78	0.512	0.757	0.531	0.764	0.511	0.537	0.217	0.505	0.261	0.505	0.221	0.874	0.711	0.874	0.604	0.869	0.632	0.515	0.348	0.506	0.338	0.491	0.325
top3	0.84	0.578	0.827	0.583	0.831	0.568	0.652	0.285	0.633	0.32	0.633	0.285	0.959	0.852	0.957	0.79	0.956	0.798	0.511	0.341	0.477	0.356	0.436	0.305
extra (text)	0.618	0.403	0.613	0.456	0.607	0.414	0.344	0.124	0.284	0.18	0.279	0.132	0.56	0.393	0.544	0.346	0.542	0.349	0.538	0.344	0.546	0.348	0.533	0.338
top2	0.808	0.576	0.802	0.574	0.798	0.554	0.542	0.196	0.471	0.265	0.472	0.21	0.904	0.863	0.91	0.567	0.899	0.609	0.534	0.325	0.486	0.358	0.454	0.294
top3	0.864	0.663	0.863	0.646	0.858	0.638	0.658	0.285	0.599	0.33	0.604	0.282	0.982	0.957	0.983	0.742	0.979	0.794	0.532	0.321	0.416	0.363	0.377	0.257
expert & extra (text)	0.594	0.388	0.595	0.435	0.589	0.397	0.374	0.12	0.288	0.196	0.267	0.128	0.551	0.385	0.546	0.333	0.545	0.345	0.549	0.332	0.548	0.349	0.542	0.335
top2	0.769	0.546	0.771	0.554	0.764	0.535	0.564	0.192	0.455	0.284	0.456	0.204	0.881	0.708	0.893	0.544	0.882	0.587	0.539	0.322	0.464	0.37	0.446	0.293
top3	0.803	0.601	0.805	0.581	0.797	0.574	0.666	0.258	0.571	0.338	0.578	0.265	0.973	0.964	0.973	0.817	0.972	0.875	0.516	0.297	0.365	0.346	0.341	0.226
normal	0.596	0.404	0.603	0.401	0.592	0.384	0.325	0.113	0.348	0.113	0.33	0.102	0.54	0.411	0.546	0.322	0.54	0.34	0.537	0.386	0.552	0.315	0.542	0.334
top2	0.798	0.58	0.802	0.538	0.794	0.541	0.474	0.222	0.488	0.198	0.472	0.192	0.806	0.684	0.81	0.521	0.802	0.559	0.524	0.352	0.505	0.335	0.492	0.321
top3	0.863	0.648	0.866	0.598	0.86	0.607	0.577	0.284	0.587	0.25	0.572	0.247	0.905	0.881	0.902	0.713	0.9	0.764	0.502	0.306	0.42	0.312	0.392	0.252
MT extra (feat.) disc	0.556	0.352	0.504	0.38	0.511	0.345	0.345	0.112	0.315	0.132	0.316	0.113	0.552	0.386	0.534	0.331	0.533	0.335	0.538	0.341	0.557	0.303	0.544	0.314
top2	0.733	0.462	0.691	0.487	0.703	0.455	0.523	0.191	0.5	0.229	0.501	0.193	0.875	0.72	0.882	0.585	0.872	0.623	0.527	0.345	0.524	0.336	0.506	0.327
top3	0.804	0.552	0.779	0.552	0.785	0.531	0.642	0.303	0.634	0.319	0.631	0.293	0.968	0.953	0.968	0.798	0.966	0.845	0.525	0.333	0.49	0.329	0.44	0.285
extra (feat.)	0.631	0.412	0.593	0.462	0.587	0.415	0.335	0.124	0.252	0.202	0.232	0.133	0.534	0.348	0.538	0.315	0.536	0.325	0.518	0.309	0.521	0.311	0.513	0.305
top2	0.806	0.539	0.793	0.576	0.795	0.546	0.545	0.189	0.418	0.272	0.417	0.201	0.754	0.68	0.76	0.577	0.756	0.617	0.516	0.304	0.458	0.335	0.433	0.273
top3	0.872	0.594	0.87	0.614	0.869	0.594	0.662	0.269	0.548	0.344	0.559	0.278	0.796	0.798	0.8	0.718	0.796	0.748	0.536	0.306	0.38	0.336	0.355	0.228