



**HAL**  
open science

## ICDAR 2023 CROHME: Competition on Recognition of Handwritten Mathematical Expressions

Yejing Xie, Harold Mouchère, Foteini Simistira Liwicki, Sumit Rakesh, Rajkumar Saini, Masaki Nakagawa, Cuong Tuan Nguyen, Thanh-Nghia Truong

► **To cite this version:**

Yejing Xie, Harold Mouchère, Foteini Simistira Liwicki, Sumit Rakesh, Rajkumar Saini, et al.. ICDAR 2023 CROHME: Competition on Recognition of Handwritten Mathematical Expressions. 17th International Conference Document Analysis and Recognition - ICDAR 2023, Aug 2023, San José, CA., United States. pp.553-565, 10.1007/978-3-031-41679-8\_33 . hal-04264727

**HAL Id: hal-04264727**

**<https://hal.science/hal-04264727v1>**

Submitted on 30 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ICDAR 2023 CROHME: Competition on Recognition of Handwritten Mathematical Expressions

Yejing Xie<sup>1</sup>, Harold Mouchère<sup>1</sup>, Foteini Simistira Liwicki<sup>2</sup>, Sumit Rakesh<sup>2</sup>, Rajkumar Saini<sup>2</sup>, Masaki Nakagawa<sup>3</sup>, Cuong Tuan Nguyen<sup>4</sup>, and Thanh-Nghia Truong<sup>3</sup>

<sup>1</sup> Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

<sup>2</sup> Lulea University of Technology, Sweden

<sup>3</sup> Tokyo Univ. of Agri. & Tech., Japan

<sup>4</sup> FPT University, Vietnam

`chrome2023@univ-nantes.fr`

**Abstract.** This paper overviews the 7th edition of the Competition on Recognition of Handwritten Mathematical Expressions. ICDAR 2023 CROHME proposes three tasks with three different modalities: on-line, off-line and bimodal. 3905 new handwritten equations have been collected to propose new training, validation and test sets for the two modalities. The complete training set includes previous CROHME training set extended with complementary off-line (from OffRaSHME competition) and on-line samples (generated). The evaluation is conducted using the same protocol as the previous CROHME, allowing a fair comparison with previous results. This competition allows for the first time the comparison of the on-line and off-line systems on the same test set. Six participating teams have been evaluated. Finally the same team won all 3 tasks with more than 80% of expression recognition rate.

**Keywords:** mathematical expression recognition · handwriting recognition · bimodal · evaluation · dataset.

## 1 Introduction

Handwritten mathematical recognition is an important and challenging task with numerous real-world applications. The CROHME (Competition on Recognition of On-line Handwritten Mathematical Expressions) competition has been organized since 2011 to foster research and development in this field. The competition provides a platform for researchers to compare and evaluate their methods for recognizing on-line and off-line handwritten mathematical expressions. CROHME has contributed to significant advances in this area by promoting the development of new approaches and datasets, as well as facilitating the exchange of ideas among researchers.

At the very beginning CROHME (2011 [5], 2012 [6], 2013 [8], 2014 [9], 2016 [7]) focused only on on-line content. Recently, CROHME 2019 [4] has considered also off-line content using the rendering of the on-line signal to produce a perfect image of the handwritten expressions. This was a first step towards the off-line domain. After that, other competitions and dataset as OffSRAHME [15] appeared focusing on the off-line domain with more realistic content. This time the competition includes bimodal expressions that combine both on-line and off-line data.

In this paper, we provide an overview of the ICDAR 2023 CROHME competition, describing the three tasks, the corresponding datasets and the evaluation metrics. We also provide the results of the participating systems with a short description. We conclude by discussing the main trends and challenges in handwritten mathematical recognition.

## 2 Tasks

**Task 1. On-line Handwritten Formula Recognition.** As the main task in the previous CROHME, participants utilise InkML format data, which has series of handwritten strokes collected by a tablet or similar device (as in Figure 1), and should convert it to Symbol Label Graph (SymLG), which is a specific Symbol Layout Tree since CROHME 2019. Note that the participants were advised to use strictly only the on-line information and not the off-line information.

**Task 2. Off-line Handwritten Formula Recognition.** Real scanned images (as in Figure 3) or rendered images from InkML (as in Figure 2) are used and should be converted to SymLG. In the same manner as for Task 1, the participants were advised to use strictly only the off-line information and not the on-line information.

**Task 3. Bi-modal Handwritten Formula Recognition.** Both the on-line strokes and off-line images which come from the same acquisition were used for the bi-modal system. While similar with the 2 tasks mentioned above, the outputs of bi-modal system are also SymLG. Note that the participants in this task could use both the on-line and off-line information.

All three tasks involved in the above system are ranked according to the rate of completely correct recognition, but each task is ranked separately.

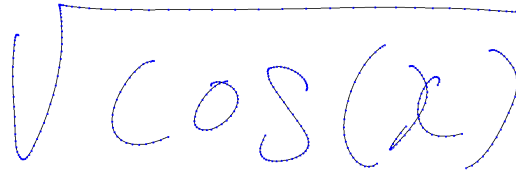
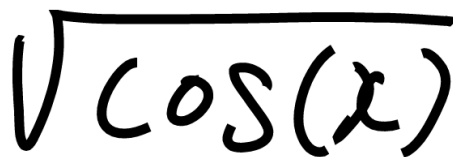
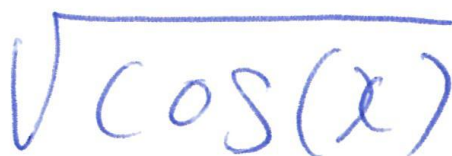


Fig. 1: On-line handwritten formula. Strokes are sequence of points (in Inkml file format).



$$\sqrt{\cos(x)}$$

Fig. 2: Off-line handwritten formula rendered from Inkml (PNG file format).



$$\sqrt{\cos(x)}$$

Fig. 3: Off-line handwritten formula scanned from real paper (400 dpi, PNG file format).

### 3 Datasets and formula encodings

In this section, we describe the data sources, data collection, data format and data encoding that we used in this competition. As indicated in Table 1, CROHME 2023 collects new bimodal data both from off-line real images and on-line strokes, as well as merges data from the previous CROHME. On this basis, it is supplemented by the off-line dataset OffRaSHME and large scale of artificial on-line data.

#### 3.1 Handwritten Formulas Data Format.

In general, CROHME 2023 provides 2 types of input handwritten formula data, which are InkML for on-line and PNG format image for off-line, and the bi-modal is combining these 2 formats of one same handwritten formula. All 3 tasks used Symbol Label Graph (SymLG) as the ground truth for system training and also as the final output of the system for further performance evaluation.

**InkML:** Strokes are defined by lists of  $(x, y)$  coordinates, representing sampled points which are collected by on-line input equipment. Figure 1 gives an example of on-line version of an expression, each point is visible and connected to the next point to see the strokes. The specific strokes compose the symbols, and these symbols compose the handwritten formulas. Each InkML (except test

Table 1: CROHME 2023 Data Sets. On-line data are only inkml files, off-line are images, bimodal are inkml and image for each equation.

Tasks	Train	#	Validation	#	Test	#
Task 1						
On-line	Train 2019	9 993	Test 2016	1147		
	Validation 2019	986				
	artificial	145 108				
	new samples	1045	new samples	555	new samples	2 300
	<b>Total=</b>	<b>157 132</b>	<b>Total=</b>	<b>1 702</b>	<b>Total=</b>	<b>2 300</b>
Task 2						
Off-line	rendered	9975	rendered	1147		
	real	1604				
	OffRaSHME	10 000				
	new samples	1045	new samples	555	new samples	2 300
	<b>Total=</b>	<b>20 979</b>	<b>Total=</b>	<b>1 702</b>	<b>Total=</b>	<b>2 300</b>
Task 3:						
Bimodal	InkML + rendered	9 975	InkML + rendered	1147		
	InkML + real	1 604				
	new samples	1045	new samples	555	new samples	2 300
	<b>Total=</b>	<b>10 979</b>	<b>Total=</b>	<b>1 702</b>	<b>Total=</b>	<b>2 300</b>

set), in addition to giving information on the position of the strokes, also provides stroke-level annotation information to indicate the relationship between strokes with the assistance of MathML (XML-based representation) in presentation mode. Thus the symbol segmentations and labels are available in this type of data. Similarly, the Label Graph (LG), which is a CSV-based representation for Stroke-level Label Graph used in previous CROHME is also provided

**Image:** Due to the different sources of data, there are two different images types. Indeed, in first competitions several types of acquisition devices have been used: Annoto pens (which produce on-line and off-line signal), tablet-pc with pen based sensitive screen, and numeric white board. So a number of early CROHME data, which are only available as on-line data, lack real off-line images, so these were rendered automatically from the on-line data with size  $1000 \times 1000$  pixels and 5 pixels of edge padding. At the same time, as OffSRaHME [15], we scanned all the formulas that existed in real writing on paper; these real off-line images are not gray-scale and have no fixed size due to the different sizes of handwritten formulas scale.

**Symbol Label Graph:** With the development of end-to-end handwritten mathematical recognition system, the stroke level segmentation of each symbol is no longer produced. Therefore, CROHME 2023 does not demand stroke level

Label Graph (LG) files as the final output of the system, only Symbol Label Graph (SymLG) is required as formula structure representation for evaluation. SymLG and the related metrics are detailed in section 4.

### 3.2 Formula Data Selection and Collection.

Firstly a corpus of expressions is built then these expressions are written by volunteers.

**Formula Data Selection.** The selection of the mathematical expressions is an important step because it selects the language domain of the competition. To build this list of expressions, we have followed the same process as in the past contests, as described in [10]. Using a public set of  $\text{\LaTeX}$  sources of scientific papers [1] all math expressions have been extracted and filtered using the grammar IV defined in previous context. Then the expressions are selected in this list to keep the term frequency and expressions length frequency comparable to the training set. 4000 expressions have been pre-selected.

**Handwritten Formula Collection.** The 2023 new data collection was done over 150 of participants in three sites: France, Japan and Sweden with the Wacom Intuio device, which is allowed to record the pen trajectory while writing on a standard paper. Each A4 size paper has 5 or 8 formulas written on it. Most of the formulas need to be written in rectangular boxes in order to easily separate the formulas on the same sheet of paper. We also tried to let the participants write down a series of formulas freely, without any limitation of boxes.

After the collection, each expression have been individually checked. A part of them has been discarded because of issues in the acquisition (miss matching of on-line and off-line), connecting symbols (the same stroke shared by 2 symbols), or scratching... leading to a new set of 3900 handwritten expressions.

### 3.3 Generated artificial HME

Two pattern generation sets are provided for Task 1. The first set consists of syntactic patterns generated by parsing the structure of handwritten mathematical expressions (HMEs) to identify the syntactic role of each component within an HME [12]. After that, each HME is decomposed to identify all valid sub-HMEs based on their syntactic role. These sub-HMEs are considered as new meaningful HMEs. Moreover, we also interchange sub-HMEs that have the same syntactic role inside each HME to get new HME patterns. The generated syntactic patterns differ from the original patterns, and we have provided approximately 76,224 syntactic patterns in CROHME 2023.

The second set consists of synthetic patterns generated from  $\text{\LaTeX}$  sequences. We create an XML layout from each  $\text{\LaTeX}$  sequence and update the layout with the handwritten symbols extracted from CROHME 2019 training set. We generated approximately 68,884 HME patterns from  $\text{\LaTeX}$  sequences of the CROHME 2019 Wikipedia corpus and collected  $\text{\LaTeX}$  sequences.

The table 1 describes clearly the usage of the different data sets in the competition.

## 4 Evaluation metrics

### 4.1 Symbol Layout Graph: SymLG

As CROHME 2023 concentrates on formula recognition for a variety of input formats, the evaluation metric of formula recognition is always the same symbol-level evaluation for different tasks. The unified SymLG representation forgets which strokes belong to which symbol and allows all systems, whether they produce stroke-level or symbol-level results, to have an identical standard that can be compared directly. The Figure 4(a) gives an example of a SymLG for an expression containing 4 symbols. This graph can be generated from the  $\text{\LaTeX}$  string, from the mathML representation (available in InkML files) or from Stroke level Label Graphs.

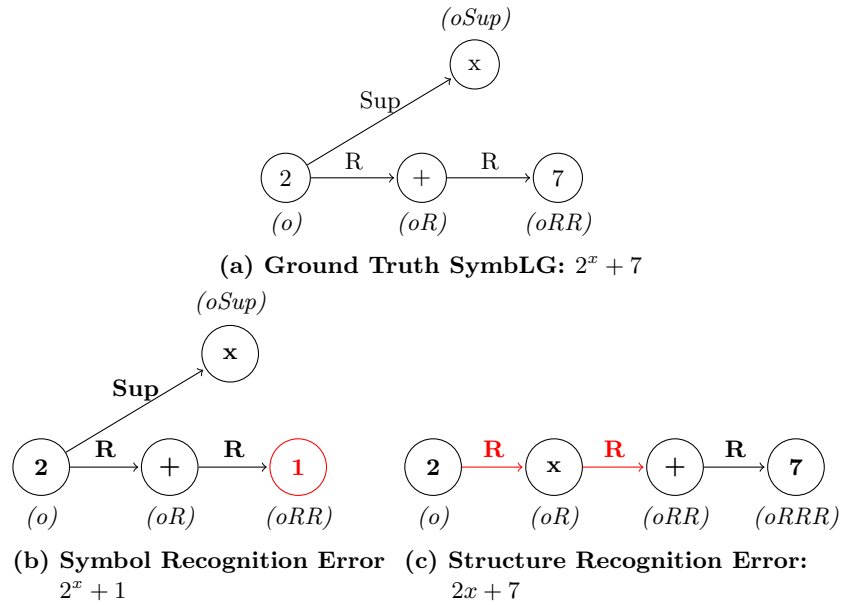


Fig. 4: A visualisation example for formula recognition evaluation with SymLG. Compared with the Ground Truth (a) ' $2^x + 7$ ', (b) ' $2^x + 1$ ' and (c) ' $2x + 7$ ' are 2 possible mistakes of system recognition. (b) has a symbol recognition error, which means that '1' and '7' are mistaken, while (c) has two structure recognition errors, which means that the relationship between '2' and 'x', '2' and '+' are misunderstood, which leads to new errors of mismatching nodes  $oRR$  and  $oRRR$ .

### 4.2 Metrics

As previous CROHME, initially described in [10], updated in [4], we consider expression and structure recognition rates as the most important metrics.

For each system, we generated a table giving the percentage of expressions with matching MathML trees, and with at most 3 incorrect symbols or relations. And the completely correct recognition rates were used to rank systems for all Tasks. It should be noticed that the evaluation metric applied on SymLG loses a part of its advantages.

Indeed, originally designed for stroke level comparison, the tool lgeval first aligned the node identifiers. For one expression, stroke identifiers are unique and the same whatever the recognized output. In Symbol level LG, the node identifiers are based on the position of the symbol in the relation tree, starting from the origin. This identification leads to add new errors during the evaluation in the case of relation errors.

For example, in Figure 4(c) the symbols  $x$ ,  $+$  and  $7$  are correctly recognized, but because of the relation error at the beginning of the expression, all following nodes are renamed and miss-matching with the ground-truth. So, in case of perfect recognition, every nodes are matching and no errors will be produced. In case of only symbol recognition errors (without structure errors), the matching is pertinent and the provided errors are meaning full. In case of a relation recognition error, all the symbols from the concerned sub-expression will be in error.

## 5 Participating methods

As clearly shown in Table 4, there are in total 6 teams joined in 3 tasks. Sunia, YP\_OCR and TUAT participated all the 3 tasks, while DPRL\_RIT only participated in task 1, PERO and UIT@AIClub\_Tensor only participated in task 2. We need to mention that, probably because of our technique problem, the system description of team UIT@AIClub\_Tensor is missed, thus the descriptions of the 5 team are shown as follow.

**Sunia PTE.LTD**<sup>5</sup> A standard encoder-decoder models are used to translate a bitmap image or a sequence of offsets and pen-up flags to a sequence of tokens in a compact language where each mathematical expression has a unique representation. The encoder consists of stacked CNN/BLSTM layers, and the decoder relies on an attention mechanism that can model convergence and estimate the location of each symbol. For on-line recognition, preprocessing steps such as reordering (for some of the combined models which were trained on spatially sorted strokes and thus stroke-order free), resampling, and normalization are performed to ensure that our system is not too sensitive to stroke order, device, location, and size. Model ensembling and LL(1) grammar parsing are also employed during beam search to boost the accuracy and avoid illegal output. We have augmented the training set by assembling synthetic samples from the subexpressions in the official training set and applying local and global distortions. For data usage, based on on-line official data, more synthetic samples were generated with stroke-level annotations by official labels of these artificial

<sup>5</sup> <https://patentscope.wipo.int/search/zh/detail.jsf?docId=DE375155214>



samples. And for off-line and bimodal task, rendered version of on-line samples are also used in addition to the official train set.

**YP\_OCR, CVTE Research:** The system utilized an attention-based encoder-decoder structure, where the encoder adopted the DenseNet architecture. In the decoding stage, both a bidirectional tree decoder and a regular decoder are applied, and two different models are obtained. For Task 1, the InkML files were rendered uniformly before feeding them into the models with different decoders. Additionally, a multi-stage curriculum learning training strategy is employed to train the models. For Task 2, the models is trained by only using the official training dataset and alleviated the problem of limited training data by using data augmentation methods. Finally, the trained models are fused in different stages to obtain better results and integrated the results of both decoders through beam search during the inference stage.

**PERO System:** The system consists of an optical model (OM) and a language model (LM). The OM is based on a CRNN architecture trained using the CTC loss function. It operates on images with a normalized height of 128 pixels and arbitrary length. The LM is an LSTM network trained to predict the next LaTeX token given its predecessors. To train the OM, the provided off-line dataset and also the provided synthetic data are used. As training data for the LM, the ground truth provided in the datasets and also publicly available equations from Wikipedia<sup>1</sup> are used. The OM is trained to produce a LaTeX code. The LM decodes logits generated by the OM using prefix search decoding and the hypothesis representing a valid LaTeX code with the highest probability is taken as a result. Finally, LaTeX codes are converted into SymLG using the provided tools.

**Team DPRL RIT QD-GGA** [3] system is extended, a visual parser that (1) creates a line-of-sight (LOS) graph over strokes, (2) scores stroke segmentation, symbol class, and relationship hypotheses with a multi-task CNN, (3) segments and classifies symbols, and (4) selects relationships using a maximum spanning tree. We have improved visual features and speed. Input images are now 64x64 for all stroke and formula window images, with formula windows centered on the target/parent stroke. The 2 nearest neighbors [14] of strokes define context windows. A modified graph attention network is used with cropped strokes and their two cropped nearest neighbors as additional context features. Spatial pyramidal pooling (SPP) [2] was added to avoid spatial information loss. These features have improved symbol segmentation and classification rates, while relationships require more work. Memory optimizations provide inference times of 25.9 ms/formula on a desktop system with two Nvidia GeForce GTX 1080 Ti GPUs (12GB each).

**Team TUAT** Two models and their combination are provided for on-line, off-line, and bi-modal recognition tasks. For the on-line recognition model, a deep BLSTM network is used for jointly classifying symbols and relations. Then, a 2D-CFG is used to parse the symbols into mathematical structure [11]. For the off-line recognition model, an end-to-end deep neural network that is trained using weakly supervised learning is used, and a symbol classifier is added to the

encoder-decoder model to improve the localization and classification of the CNN features. More details of this method can be found in the paper [13]. For the combination of on-line and off-line models, firstly several candidates with their probability scores during the Beam Search process are generated for each on-line and off-line recognition model. Then, their prediction scores are combined using a trained weight. The weight is chosen by the best combination recognition rate on the validation set. For the data, the systems are trained using the given dataset as provided by the competition.

## 6 Results

In this section, we summarize all of the results in CROHME 2023. Table 2 and 3 present respectively results on the new test set 2023 and on the previous test set 2019. There are in total 6 teams participating in 3 tasks, 3 of them participate in all of these 3 tasks, while the other 3 teams only selected one of on-line or off-line task.

Table 2: Formula Recognition Results (Test set 2023)

	Structure + Symbol Labels			Structure Correct rate
	Correct	$1 \leq \text{s.err}$	$2 \leq \text{s.err}$	
<b>Task 1: On-line</b>				
Sunia*	<b>82.34</b>	<b>90.26</b>	<b>92.47</b>	<b>92.41</b>
YP_OCR	72.55	83.57	86.22	86.60
TUAT	41.10	54.52	60.04	56.85
DPRL_RIT	38.19	53.39	58.39	59.98
<b>Task 2: Off-line</b>				
Sunia*	<b>70.81</b>	<b>81.74</b>	<b>86.13</b>	<b>86.95</b>
YP_OCR	67.86	80.86	85.13	85.99
PERO	58.37	71.22	75.57	75.55
TUAT	51.02	63.17	67.48	64.59
UIT@AIClub_Tensor	14.83	22.87	28.57	32.19
<b>Task 3: Bi-modal</b>				
Sunia*	<b>84.12</b>	<b>91.43</b>	<b>93.70</b>	<b>94.13</b>
YP_OCR	72.55	83.57	86.22	86.60
TUAT	53.76	68.83	74.22	70.81

**Sunia** team obtained all the highest recognition rate in on-line, off-line and bi-modal tasks, and is ahead of **YP\_OCR** who also took part in all of 3 tasks.

Compared to the Test set 2019, Test set 2023 had lower recognition rates for the same participants. AS this is a global observation, we can conclude that the new test set is harder than the previous one. However, Sunia, which won CROHME 2023, had a better recognition rate than the systems that participated in CROHME 2019 on the same data. It is a significant indication of the

Table 3: Formula Recognition Results (Testset 2019)

	Structure + Symbol Labels			Structure
	Correct	$1 \leq \text{s.err}$	$2 \leq \text{s.err}$	Correct rate
<b>Task 1: On-line</b>				
Sunia*	<b>88.24</b>	<b>93.08</b>	<b>93.99</b>	<b>94.25</b>
YP_OCR	84.74	90.99	92.33	92.66
TUAT	56.88	71.89	76.31	70.73
DPRL_RIT	40.70	59.47	67.06	70.23
<i>2019 winner</i>	<i>80.73</i>	<i>88.99</i>	<i>90.74</i>	<i>91.49</i>
<b>Task 2: Off-line</b>				
Sunia*	<b>77.73</b>	<b>87.41</b>	<b>90.08</b>	<b>90.58</b>
YP_OCR	73.39	85.74	88.49	89.16
PERO	67.39	77.23	81.07	78.98
TUAT	50.13	63.55	66.89	63.30
UIT@AIClub_Tensor	38.28	52.29	58.80	59.13
<i>2019 winner</i>	<i>77.15</i>	<i>86.82</i>	<i>88.99</i>	<i>89.49</i>
<b>Task 3: Bi-modal</b>				
Sunia*	<b>86.91</b>	<b>92.16</b>	<b>93.58</b>	<b>93.74</b>
YP_OCR	84.74	90.99	92.33	92.66
TUAT	57.88	73.81	82.82	72.14

continuous developments in Handwritten Mathematical Expression Recognition over recent years.

Since the 3 tasks in this competition use exactly the same evaluation metrics, it is interesting to conduct a comparison between different tasks. In general, bi-modal systems have the best performance, which has a more enriched data dimension therefore fits the common sense. In most cases, on-line systems have better capabilities than off-line systems, except on-line and off-line systems from TUAT. This illustrates that on-line data with the combination of spatial and temporal information is more conducive than off-line data which only consider the spatial information. However, we do not exclude that generated artificial on-line HME greatly increases the sample size of the on-line dataset, and improving the performance of the deep learning system. Therefore, it is also necessary to further explore the usage of artificial data for off-line datasets, which can greatly increase the data volume of off-line datasets and save manual annotation costs.

A summary of confusion histograms that tabulate errors in symbols and symbol pairs as defined in [10] is presented in Table 4. It is important to note that the most common errors are confusion with absent symbols which are not shown in the table. Indeed, the confusion with absent symbols is always caused by the errors in structure, different structures can not be matched symbol by symbol, that raises errors in the position and classification of the symbols, as explained in section 4.2. The statistics in the table show that symbols 1, 2, and – are the most difficult to recognize, which have commonalities across systems, maybe because these symbols are the most frequent. For symbol pairs, most of

systems are hard to identify  $-1$ ,  $\frac{1}{2}$ , and  $\frac{1}{2}$ , there is also specificity in the different systems.

Table 4: Most frequent symbol and sub-structure recognition errors (Test set 2023). **E** columns give the GT symbol or sub-structure which have been mis-recognized. **#** columns give the number of occurrences of this errors.

	Symbols						Symbol Pairs					
	$1^{st}$		$2^{nd}$		$3^{rd}$		$1^{st}$		$2^{nd}$		$3^{rd}$	
	<b>E</b>	<b>#</b>	<b>E</b>	<b>#</b>	<b>E</b>	<b>#</b>	<b>E</b>	<b>#</b>	<b>E</b>	<b>#</b>	<b>E</b>	<b>#</b>
<b>Task1: On-line</b>												
Sunia	1	84	)	71	2	68	-1	34	--	27	$\frac{1}{2}$	26
YP_OCR	1	193	2	152	-	132	-1	39	(x	26	$\frac{1}{2}$	26
TUAT	1	405	2	380	-	362	$\frac{1}{2}$	70	-1	70	--	59
DPRL_RIT	1	599	2	558	-	540	$\frac{1}{2}$	112	-1	109	$\frac{1}{2}$	104
<b>Task2: Off-line</b>												
Sunia	1	181	2	157	-	136	-1	27	--	26	$\frac{1}{2}$	21
YP_OCR	1	213	x	172	2	168	x)	38	-1	36	(x	33
PERO	1	331	-	225	2	214	-1	57	= 1	52	$\frac{1}{2}$	37
TUAT	1	291	-	213	2	211	$\frac{1}{2}$	51	= 1	50	--	48
UIT@AIClub_Tensor	-	986	1	982	2	978	$\frac{1}{2}$	169	$\frac{1}{2}$	165	-1	148
<b>Task3: Bi-modal</b>												
Sunia	1	70	2	56	x	46	x)	14	ab	14	dx	12
YP_OCR	1	193	2	152	-	132	-1	39	(x	26	$\frac{1}{2}$	26
TUAT	1	337	2	254	-	245	$\frac{1}{2}$	58	= 1	55	--	55

## 7 Conclusion

CROHME 2023 provides 1,045 new expressions in train set, 555 in validation set and 2,300 in test set of handwritten formulas in bi-modal with manual annotations, as well as large-scale of artificial data. Compared to the CROHME 2019 results, the CROHME 2023 winner had a superior performance, especially in on-line task. But the winner’s improvements are not as impressive as those in 2019. CROHME 2023 participants all adopted the deep learning approaches, training models with large amounts of data. However, several system continue to use structural constraints (as LL(1) or CFG grammars). Furthermore, statistical language models are now integrated by some participating systems. It is worth noting that, the winner team Sunia, used data augmentation strategy to increase the amount of provided official data. In addition, most on-line systems have better performance than off-line systems, there is still a significant improvement potential for Handwritten Mathematical Expression Recognition systems, especially off-line.

On the basis of this topic, we have more in-depth ongoing discussions scheduled for the future. Nowadays, deep learning models, especially those with encoder-

decoder structure, have become the dominant solution for Handwritten Mathematical Expression Recognition. These systems are sensitive to samples quantity and quality of data sets, therefore, it will be interesting to extend the amount of existing data set by exploring more cost-effective annotation methods as well as gaining more data with the help of synthetic algorithms. Exploring non-supervised training strategy would also be benefit. In addition, more benchmarking experiments are possible to explore the highlights and weaknesses of each system, in order to provide more guidance for them.

## Acknowledgments

The organization team of CROHME 2023 would like to thank all the writers who participated in the input of the thousands of equations, the colleagues who helped with the annotation, the participants for their constructive feedback. We particularly thank Da-Han Wang and the OffRASHME team for sharing their data.

## References

1. Gehrke, J., Ginsparg, P., Kleinberg, J.: Overview of the 2003 KDD cup. *ACM SIGKDD Explorations Newsletter* **5**(2), 149–151 (2003)
2. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* **37**(9), 1904–1916 (2015)
3. Mahdavi, M., Zanibbi, R.: Visual parsing with query-driven global graph attention (qd-gga): preliminary results for handwritten math formula recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 570–571 (2020)
4. Mahdavi, M., Zanibbi, R., Mouchère, H., Viard-Gaudin, C., Garain, U.: ICDAR 2019 CROHME + TFD: Competition on Recognition of Handwritten Mathematical Expressions and Typeset Formula Detection. In: *15th IAPR International Conference on Document Analysis and Recognition (ICDAR 2019)*. Sydney, Australia (Sep 2019)
5. Mouchère, H., Viard-Gaudin, C., Kim, D.H., Kim, J.H., Utpal, G.: Crohme2011: Competition on recognition of online handwritten mathematical expressions. In: *ICDAR*. Beijing, China (2011)
6. Mouchère, H., Viard-Gaudin, C., Kim, D.H., Kim, J.H., Utpal, G.: Icfhr 2012 - competition on recognition of on-line mathematical expressions (crohme 2012). In: *ICFHR*. Bari, Italy (2012)
7. Mouchère, H., Viard-Gaudin, C., Zanibbi, R., Garain, U.: Icfhr2016 competition on recognition of handwritten mathematical expressions (crohme2016). In: *ICFHR*. Shenzhen, Chine (2016)
8. Mouchère, H., Viard-Gaudin, C., Zanibbi, R., Garain, U., Kim, D.H., Kim, J.H.: Icdar 2013 crohme: Third international competition on recognition of online handwritten mathematical expressions. In: *ICDAR*. Washington, DC, USA (August 2013)

9. Mouchère, H., Viard-Gaudin, C., Zanibbi, R., Utpal, G.: Icfhr 2014 - competition on recognition of on-line mathematical expressions (crohme 2014). In: ICFHR. Crete, Greece (2014)
10. Mouchère, H., Zanibbi, R., Garain, U., Viard-Gaudin, C.: Advancing the state of the art for handwritten math recognition: the CROHME competitions, 2011–2014. *IJDAR* **19**(2), 173–189 (2016). <https://doi.org/10.1007/s10032-016-0263-5>
11. Nguyen, C.T., Truong, T.N., Nguyen, H.T., Nakagawa, M.: Global context for improving recognition of online handwritten mathematical expressions. In: Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part II 16. pp. 617–631. Springer (2021)
12. Truong, T.N., Nguyen, C.T., Nakagawa, M.: Syntactic data generation for handwritten mathematical expression recognition. *Pattern Recognition Letters* **153**, 83–91 (2022)
13. Truong, T.N., Nguyen, C.T., Phan, K.M., Nakagawa, M.: Improvement of end-to-end offline handwritten mathematical expression recognition by weakly supervised learning. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 181–186. IEEE (2020)
14. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
15. Wang, D.H., Yin, F., Wu, J.W., Yan, Y.P., Huang, Z.C., Chen, G.Y., Wang, Y., Liu, C.L.: Icfhr 2020 competition on offline recognition and spotting of handwritten mathematical expressions-offrashme. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 211–215. IEEE (2020)