NOTE: This analysis is not covered in the paper

ANALYSIS OF EXPERIMENTS BY TYPE

To explore to what extent the limitations observed vary across experiments with different objectives, we classified the experiments and reassessed them by type. Table 1 provides an overview of the classification of the 194 experiments into three groups: evaluation, generalization, and optimization.

Table 1. Types of experiments found

Category	Count	Percentage
Optimization	67	35%
Evaluation	90	46%
Generalization	25	13%
Optimization+Evaluation	10	5%
Evaluation+Generalization	2	1%

Evaluation experiments aim at comparing the proposed DNN with a baseline of expected-values set by the researchers, DNNs proposed by other researchers, other techniques that do not rely on DNNs, or human performance. We find that 46% of the experiments aim to perform such an evaluation. **Generalization experiments** aim to assess a DNN under a different dataset, most commonly under a different test set but also sometimes under a different training set used to define the model parameters, or a new user context. We find that 13% of the experiments fall into this category. For example, [AP6] proposes a DNN for predicting developer actions (represented as a sequence of image regions). One of the experiments runs the DNN for predicting actions for developers and programming languages different from the ones in the training set. **Optimization experiments** aim at exploring and eventually identifying the best DNN configuration, within some allocated resources, through the manipulation of a large number of variables, from the model hyperparameters to the deep learning algorithm. We find a large number of optimization experiments (35%). We also find **combinations of optimization and evaluation experiments** (the same experiment compares other approaches and variants of the proposed approach) in 5% of the cases. Finally, 1% of experiments **combine evaluation and generalization** (the same experiment compares the DNN with other approaches while it is being assessed under a different dataset, or new user context).

We find that the limitations observed earlier remain mostly the same across experiment types. One noticeable difference, however, is that the optimization experiments are the ones with the most missing response variables (18% vs. 0% for the other types of experiments) and factors and treatments specification (10% vs. 0% and 4% for evaluation and generalization experiments). Since the response variable is often associated with accuracy, this is an easily fixed oversight. However, missing factors and treatments seem more problematic since it undermines these experiments' objective to identify the best model configuration, and the factors and treatments are key in defining such configuration. We provide the full breakdown of the data.

Characterization of DNN experiments per type

		Fully Addressed				Partially Addressed				Missing			
		Optimization	Evaluation	Generalization	All	Optimization	Evaluation	Generalization	All	Optimization	Evaluation	Generalization	All
Hypotheses	**Research**	60%	84%	84%	76%	0%	0%	0%	0%	40%	16%	16%	24%
Variables identification	**Model hyperparameters**	3%	12%	0%	7%	90%	78%	96%	86%	7%	10%	4%	8%
	Model parameters	3%	2%	0%	2%	0%	0%	0%	0%	97%	98%	100%	98%
	DL algorithm	24%	31%	20%	26%	73%	66%	80%	72%	3%	3%	0%	3%
	Training hyperparameters	19%	21%	16%	19%	73%	70%	72%	73%	7%	9%	12%	8%
	Training data	73%	66%	80%	70%	25%	29%	20%	27%	1%	6%	0%	4%
Operationalization	**Factors and treatments**	37%	0%	12%	14%	52%	100%	84%	81%	10%	0%	4%	4%
	Response variables	70%	81%	80%	76%	12%	19%	20%	18%	18%	0%	0%	6%
Design	**Choice of design**	0%	0%	0%	0%	58%	73%	84%	70%	42%	27%	16%	30%
	Instrumentation	3%	1%	0%	2%	94%	99%	100%	97%	3%	0%	0%	1%
Objects selection	**Test sets characteristics**	57%	57%	84%	59%	16%	26%	12%	20%	27%	18%	4%	22%
Analysis & interpretation	**Descriptive statistics**	4%	12%	20%	10%	22%	36%	44%	34%	73%	52%	36%	56%
	Inferential statistics	6%	16%	20%	12%	0%	1%	0%	1%	94%	83%	80%	87%
Validity evaluation	**Validity threats**	0%	1%	8%	2%	87%	76%	68%	79%	13%	23%	24%	20%

Summary of characterization: All papers

		Hypotheses	Variables identification					Operationalization		Design		Population	Analysis		Validity evaluation
		Research	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics	
EXPERIMENTS:	**194**														
COUNT M:		47	15	190	5	16	7	8	12	58	2	42	109	169	38
COUNT PA:		0	166	0	139	142	52	158	34	136	189	38	66	1	153
COUNT FA		147	13	4	50	36	135	28	148	0	3	114	19	24	3
% M		24%	8%	98%	3%	8%	4%	4%	6%	30%	1%	22%	56%	87%	20%
% PA		0%	86%	0%	72%	73%	27%	81%	18%	70%	97%	20%	34%	1%	79%
% FA		76%	7%	2%	26%	19%	70%	14%	76%	0%	2%	59%	10%	12%	2%
TOTAL		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
OPTIMIZATION:	**67**														
COUNT M:		27	5	65	2	5	1	7	12	28	2	18	49	63	9
COUNT PA:		0	60	0	49	49	17	35	8	39	63	11	15	0	58
COUNT FA		40	2	2	16	13	49	25	47	0	2	38	3	4	0
% M		40%	7%	97%	3%	7%	1%	10%	18%	42%	3%	27%	73%	94%	13%
% PA		0%	90%	0%	73%	73%	25%	52%	12%	58%	94%	16%	22%	0%	87%
% FA		60%	3%	3%	24%	19%	73%	37%	70%	0%	3%	57%	4%	6%	0%
TOTAL		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
EVALUATION:	**90**														
COUNT M:		14	9	88	3	8	5	0	0	24	0	16	47	75	21
COUNT PA:		0	70	0	59	63	26	90	17	66	89	23	32	1	68
COUNT FA		76	11	2	28	19	59	0	73	0	1	51	11	14	1
% M		16%	10%	98%	3%	9%	6%	0%	0%	27%	0%	18%	52%	83%	23%
% PA		0%	78%	0%	66%	70%	29%	100%	19%	73%	99%	26%	36%	1%	76%
% FA		84%	12%	2%	31%	21%	66%	0%	81%	0%	1%	57%	12%	16%	1%
TOTAL		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
GENERALIZATION:	**25**														
COUNT M:		4	1	25	0	3	0	1	0	4	0	1	9	20	6
COUNT PA:		0	24	0	20	18	5	21	5	21	25	3	11	0	17
COUNT FA		21	0	0	5	4	20	3	20	0	0	21	5	5	2
% M		16%	4%	100%	0%	12%	0%	4%	0%	16%	0%	4%	36%	80%	24%
% PA		0%	96%	0%	80%	72%	20%	84%	20%	84%	100%	12%	44%	0%	68%
% FA		84%	0%	0%	20%	16%	80%	12%	80%	0%	0%	84%	20%	20%	8%
TOTAL		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Characterization of ICSE papers

Venue	Paper #	Experiment	Type	Hypotheses	Variables identification					Operationalization		Design		Population	Analysis		Validity evaluation	Artifact	
				Research	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics		Availability	Badge
ICSE'18	AP1	E1	Optimization	M	PA	M	PA	PA	FA	M	M	M	PA	PA	M	M	M	Yes	No
		E2	Evaluation	M	PA	M	PA	PA	FA	PA	FA	PA	PA	PA	M	M	M		
		E3	Generalization	M	PA	M	PA	PA	FA	M	FA	M	PA	PA	PA	M	M		
		E4	Generalization	M	PA	M	PA	PA	FA	PA	FA	PA	PA	PA	PA	FA	M		
ICSE'18	AP2	E1	Evaluation	M	PA	M	PA	PA	FA	PA	FA	M	PA	FA	PA	PA	PA	Yes	No
ICSE'19	AP3	E1	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	M	PA	FA	M	M	PA		
ICSE'19	AP4	E1	Optimization	M	PA	M	PA	PA	FA	FA	PA	M	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	M	PA	FA	M	M	PA		
ICSE'19	AP5	E1	Optimization+evaluation	FA	PA	M	PA	PA	PA	PA	FA	M	PA	M	M	M	M	Yes	Available Reusable
		E2	Evaluation	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	M	M	M		
		E3	Evaluation	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	M	M	M		
		E4	Optimization	FA	PA	M	PA	PA	PA	M	FA	PA	PA	M	M	M	M		
ICSE'19	AP6	E1	Optimization	FA	PA	M	PA	PA	FA	FA	FA	PA	PA	FA	PA	M	M	No	No
		E2	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	M	M		
		E3	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	M	M		
		E4	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	M	M		
		E5	Evaluation	FA	M	M	M	M	M	PA	PA	M	PA	M	M	M	M		
ICSE'19	AP7	E1	Optimization	M	PA	M	PA	PA	PA	FA	M	M	PA	M	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	PA	PA	PA	PA	PA	M	M	M	PA		
		E3	Generalization	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	M	M	PA		
ICSE'19	AP8	E1	Evaluation	FA	PA	FA	PA	FA	FA	PA	FA	M	PA	PA	M	M	PA	Yes	Available
		E2	Evaluation	FA	PA	FA	PA	FA	FA	PA	FA	M	PA	PA	M	M	PA		
		E3	Optimization	FA	PA	FA	PA	FA	FA	PA	FA	M	PA	PA	M	M	PA		
		E4	Optimization	FA	PA	FA	PA	FA	FA	PA	FA	M	PA	PA	M	M	PA		
ICSE'19	AP9	E1	Evaluation	FA	PA	M	PA	PA	FA	PA	PA	M	PA	PA	M	M	PA	Yes	No
		E2	Optimization	FA	PA	M	PA	PA	FA	PA	PA	M	PA	PA	M	M	PA		
ICSE'19	AP10	E1	Optimization	M	PA	M	PA	PA	FA	PA	M	PA	PA	PA	M	M	PA	Yes	Available
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	PA	FA	FA	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	PA	FA	M	PA		
		E4	Optimization	FA	PA	M	PA	PA	FA	FA	FA	PA	PA	PA	FA	M	PA		
ICSE'20	AP11	E1	Evaluation	M	PA	M	PA	PA	FA	PA	PA	M	PA	FA	M	FA	M	Yes	No
		E2	Evaluation	M	PA	M	PA	PA	FA	PA	FA	M	PA	FA	FA	FA	M		
ICSE'20	AP12	E1	Evaluation	M	PA	M	PA	M	PA	PA	FA	PA	PA	M	M	M	PA	Yes	No
		E2	Generalization	M	PA	M	PA	M	PA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Generalization	M	PA	M	PA	M	PA	PA	FA	PA	PA	FA	M	M	PA		
		E4	Optimization	M	PA	M	PA	M	PA	PA	FA	PA	PA	FA	M	M	PA		
ICSE'20	AP13	E1	Optimization	M	PA	M	PA	PA	FA	FA	FA	PA	PA	FA	PA	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E4	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E5	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
ICSE'20	AP14	E1	Optimization	M	M	M	PA	PA	FA	FA	FA	PA	PA	M	M	M	PA	Yes	No
		E2	Evaluation	FA	M	M	PA	PA	FA	PA	FA	PA	PA	M	M	M	PA		
		E3	Optimization	FA	M	M	PA	PA	FA	PA	FA	M	PA	M	M	M	PA		
ICSE'20	AP15	E1	Optimization+evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	M	PA	M	PA	Yes	Available
		E2	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	M	PA	M	PA		
		E3	Optimization+evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	M	PA	M	PA		
ICSE'20	AP16	E1	Optimization	M	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E4	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E5	Evaluation	M	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	FA	PA		
ICSE'20	AP17	E1	Optimization	FA	PA	M	FA	PA	FA	FA	FA	M	PA	PA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	FA	PA	FA	PA	PA	PA	PA	PA	M	M	PA		
		E3	Evaluation	FA	PA	M	FA	PA	FA	PA	FA	PA	PA	PA	M	M	PA		
ICSE'21	AP18	E1	Evaluation	FA	PA	M	FA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA	Yes	No
ICSE'21	AP19	E1	Evaluation	FA	M	M	PA	M	M	PA	FA	PA	PA	FA	PA	M	PA	Yes	No

Characterization of ICSE papers

Venue	Paper #	Experiment	Type	Hypotheses	Variables identification					Operationalization		Design		Population	Analysis		Validity evaluation	Artifact	
				Research	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics		Availability	Badge
ICSE'21	AP20	E1	Optimization	M	PA	M	FA	FA	PA	FA	M	M	PA	M	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	FA	FA	PA	PA	FA	PA	PA	M	PA	M	PA		
		E3	Optimization	FA	PA	M	FA	FA	PA	PA	FA	PA	PA	M	PA	M	PA		
ICSE'21	AP21	E1	Evaluation	FA	PA	M	PA	FA	PA	PA	FA	M	PA	M	M	M	M	Yes	No
ICSE'21	AP22	E1	Optimization	M	PA	M	PA	FA	FA	FA	M	PA	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E4	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E5	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
ICSE'21	AP24	E1	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E4	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E5	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E6	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
ICSE'21	AP23	E1	Optimization+evaluation	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	PA	M	PA	Yes	No
		E2	Optimization+evaluation	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	PA	M	PA		
		E3	Optimization	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	M	PA	M	PA		
		E4	Evaluation	M	PA	M	PA	PA	PA	PA	PA	PA	PA	M	PA	M	PA		

EXPERIMENTS: **78**

COUNT M:	21	5	74	1	6	2	3	5	21	0	23	50	72	16
COUNT PA:	0	73	0	70	63	19	66	8	57	78	17	21	1	62
COUNT FA	57	0	4	7	9	57	9	65	0	0	38	7	5	0
% M	27%	6%	95%	1%	8%	3%	4%	6%	27%	0%	29%	64%	92%	21%
% PA	0%	94%	0%	90%	81%	24%	85%	10%	73%	100%	22%	27%	1%	79%
% FA	73%	0%	5%	9%	12%	73%	12%	83%	0%	0%	49%	9%	6%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

OPTIMIZATION: **29**

COUNT M:	10	2	27	0	1	0	2	5	9	0	8	21	29	3
COUNT PA:	0	27	0	26	23	6	18	2	20	29	7	7	0	26
COUNT FA	19	0	2	3	5	23	9	22	0	0	14	1	0	0
% M	34%	7%	93%	0%	3%	0%	7%	17%	31%	0%	28%	72%	100%	10%
% PA	0%	93%	0%	90%	79%	21%	62%	7%	69%	100%	24%	24%	0%	90%
% FA	66%	0%	7%	10%	17%	79%	31%	76%	0%	0%	48%	3%	0%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

EVALUATION: **34**

COUNT M:	7	3	32	1	3	2	0	0	10	0	9	23	29	7
COUNT PA:	0	31	0	29	27	7	34	6	24	34	8	8	1	27
COUNT FA	27	0	2	4	4	25	0	28	0	0	17	3	4	0
% M	21%	9%	94%	3%	9%	6%	0%	0%	29%	0%	26%	68%	85%	21%
% PA	0%	91%	0%	85%	79%	21%	100%	18%	71%	100%	24%	24%	3%	79%
% FA	79%	0%	6%	12%	12%	74%	0%	82%	0%	0%	50%	9%	12%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

GENERALIZATION: **10**

COUNT M:	4	0	10	0	2	0	1	0	1	0	1	5	9	5
COUNT PA:	0	10	0	10	8	3	9	0	9	10	2	2	0	5
COUNT FA	6	0	0	0	0	7	0	10	0	0	7	3	1	0
% M	40%	0%	100%	0%	20%	0%	10%	0%	10%	0%	10%	50%	90%	50%
% PA	0%	100%	0%	100%	80%	30%	90%	0%	90%	100%	20%	20%	0%	50%
% FA	60%	0%	0%	0%	0%	70%	0%	100%	0%	0%	70%	30%	10%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Characterization of ESEC/FSE papers

				Hypotheses		Variables identification					Operationalization		Design		Population	Analysis			Artifact	
Venue	Paper #	Experiment	Type	Research	Statistical	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics	Validity evaluation	Availability	Badge
FSE'18	AP25	E1	Optimization	M	M	PA	M	FA	PA	FA	M	PA	PA	PA	FA	M	M	M	Yes	No
		E2	Evaluation	M	M	PA	M	FA	PA	FA	PA	PA	PA	PA	FA	PA	M	M		
		E3	Evaluation	M	M	PA	M	FA	PA	FA	PA	PA	PA	PA	FA	PA	M	M		
FSE'18	AP26	E1	Optimization	M	M	PA	M	FA	PA	PA	M	PA	M	PA	PA	M	M	M	Yes	No
		E2	Evaluation	M	M	PA	M	FA	PA	PA	PA	FA	M	PA	PA	M	M	M		
		E3	Evaluation	M	M	PA	M	FA	PA	PA	PA	FA	M	PA	PA	M	M	M		
		E4	Evaluation	M	M	PA	M	FA	PA	PA	PA	PA	M	PA	PA	M	M	M		
FSE'18	AP27	E1	Evaluation	FA	M	PA	M	PA	PA	PA	PA	PA	M	PA	M	M	M	PA	No	No
FSE'18	AP28	E1	Evaluation	FA	M	PA	M	PA	PA	PA	PA	FA	M	PA	PA	M	M	PA	No	No
		E2	Evaluation	FA	M	PA	M	PA	PA	PA	PA	FA	M	PA	PA	PA	M	PA		
		E3	Evaluation	FA	M	PA	M	PA	PA	PA	PA	FA	M	PA	PA	M	M	PA		
FSE'19	AP29	E1	Evaluation	FA	M	PA	M	PA	PA	FA	PA	PA	PA	PA	PA	PA	M	PA	Yes	Yes
FSE'19	AP30	E1	Optimization	M	M	PA	M	PA	M	PA	PA	M	M	PA	M	M	M	PA	No	No
		E2	Evaluation	FA	M	PA	M	PA	M	PA	PA	FA	PA	PA	M	FA	M	PA		
		E3	Evaluation	FA	M	PA	M	PA	M	M	PA	FA	PA	PA	M	PA	M	PA		
FSE'19	AP31	E1	Evaluation	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	FA	M	M	PA	No	No
		E2	Optimization	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	FA	M	M	PA		
FSE'19	AP32	E1	Evaluation	FA	M	PA	M	M	M	FA	PA	PA	PA	PA	FA	PA	FA	PA	Yes	No
		E2	Optimization	FA	M	PA	M	M	M	FA	PA	PA	PA	PA	FA	PA	FA	PA		
		E3	Optimization	FA	M	PA	M	M	M	FA	PA	PA	PA	PA	FA	PA	FA	PA		
FSE'20	AP33	E1	Evaluation	FA	M	PA	M	PA	PA	PA	PA	FA	PA	PA	PA	M	M	M	Yes	No
		E2	Optimization	FA	M	PA	M	PA	PA	PA	PA	FA	PA	PA	PA	M	M	M		
		E3	Evaluation	FA	M	PA	M	PA	PA	M	PA	FA	M	PA	M	M	M	M		
FSE'20	AP34	E1	Optimization+evaluation	M	M	PA	M	PA	PA	M	PA	PA	PA	PA	PA	M	M	M	No	No
		E2	Evaluation	M	M	PA	M	PA	PA	M	PA	PA	M	PA	PA	M	M	M		
FSE'20	AP35	E1	Optimization	M	M	PA	M	PA	PA	FA	M	M	M	PA	M	M	M	PA	Yes	No
		E2	Optimization+evaluation	M	M	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	FA	PA		
FSE'20	AP36	E1	Evaluation	FA	M	PA	M	PA	PA	FA	PA	FA	PA	FA	PA	PA	M	PA	Yes	Yes
		E2	Optimization	FA	M	PA	M	PA	PA	FA	PA	FA	M	FA	PA	M	M	PA		
		E3	Optimization	FA	M	PA	M	PA	PA	FA	FA	M	M	FA	PA	M	M	PA		
		E4	Generalization	FA	M	PA	M	PA	PA	FA	PA	FA	PA	PA	PA	M	M	PA		
FSE'20	AP37	E1	Optimization	M	M	M	M	FA	M	FA	M	M	M	PA	FA	M	M	M	Yes	No
		E2	Optimization	FA	M	M	M	FA	PA	FA	PA	PA	PA	PA	FA	PA	M	M		
		E3	Evaluation	FA	M	M	M	FA	M	FA	PA	PA	M	PA	FA	M	M	M		
		E4	Generalization	FA	M	M	M	FA	M	FA	FA	PA	M	PA	FA	M	M	M		
FSE'20	AP38	E1	Optimization	M	M	M	M	PA	PA	PA	M	M	M	M	M	M	M	M	Yes	No
		E2	Evaluation	FA	M	M	M	PA	PA	PA	PA	FA	PA	PA	PA	PA	M	M		
		E3	Evaluation	FA	M	M	M	PA	PA	PA	PA	FA	PA	PA	PA	PA	M	M		
		E4	Evaluation	FA	M	M	M	PA	PA	PA	PA	FA	PA	PA	PA	PA	M	M		
		E5	Evaluation	FA	M	M	M	PA	PA	PA	PA	FA	PA	PA	PA	PA	M	M		
		E6	Evaluation	M	M	M	M	M	M	PA	PA	FA	M	PA	PA	M	M	M		
FSE'21	AP42	E1	Evaluation	FA	M	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	M	M	PA	Yes	No
		E2	Optimization	FA	M	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Optimization	FA	M	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	M	M	PA		
FSE'21	AP43	E1	Evaluation	FA	M	PA	M	PA	PA	PA	PA	FA	PA	PA	FA	M	M	PA	Yes	No
		E2	Generalization	FA	M	PA	M	PA	PA	PA	FA	FA	M	PA	FA	M	M	PA		
		E3	Optimization	FA	M	PA	M	PA	PA	PA	PA	FA	M	PA	FA	M	M	PA		
		E4	Optimization	FA	M	PA	M	PA	PA	PA	FA	FA	M	PA	FA	M	M	PA		
FSE'21	AP40	E1	Evaluation	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	M	M	M	PA	Yes	Yes
		E2	Evaluation	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	M	M	M	PA		
		E3	Optimization	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	M	M	M	PA		
		E4	Evaluation+Generalization	FA	M	PA	M	FA	PA	PA	PA	FA	PA	PA	M	M	M	PA		
		E5	Optimization	M	M	PA	M	FA	PA	PA	FA	FA	PA	PA	M	FA	M	PA		
FSE'21	AP41	E1	Optimization	M	M	PA	M	PA	PA	M	FA	M	M	M	M	M	M	PA	Yes	Yes
		E2	Evaluation	FA	M	PA	M	PA	PA	FA	PA	FA	PA	PA	M	M	M	PA		
		E3	Evaluation+Generalization	FA	M	PA	M	PA	PA	FA	PA	FA	M	PA	M	M	M	PA		
		E4	Optimization	FA	M	PA	M	PA	PA	FA	PA	FA	PA	PA	M	M	M	PA		
		E5	Optimization	FA	M	PA	M	PA	PA	FA	PA	FA	M	PA	M	M	M	PA		
		E6	Optimization	M	M	PA	M	PA	PA	FA	PA	FA	M	PA	M	M	M	PA		
		E7	Optimization	M	M	PA	M	PA	PA	FA	FA	FA	PA	PA	M	M	M	PA		
FSE'21	AP44	E1	Evaluation	FA	M	FA	M	FA	FA	FA	PA	PA	PA	PA	FA	PA	M	PA	Yes	No
FSE'21	AP45	E1	Evaluation	FA	M	PA	M	PA	FA	FA	PA	FA	PA	PA	FA	PA	FA	PA	Yes	No
		E2	Optimization	FA	M	PA	M	PA	FA	FA	FA	FA	PA	PA	FA	PA	M	PA		
		E3	Generalization	FA	M	PA	M	PA	FA	FA	PA	FA	PA	PA	FA	PA	FA	PA		
		E6	Generalization	FA	M	PA	M	PA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA		

Characterization of ESEC/FSE papers

				Hypotheses		Variables identification					Operationalization		Design		Population	Analysis			Artifact	
Venue	Paper #	Experiment	Type	Research	Statistical	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics	Validity evaluation	Availability	Badge
FSE'21	AP39	E1	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	M	M	PA	Yes	Yes
		E2	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E3	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E4	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E5	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E6	Optimization	FA	M	FA	M	FA	FA	FA	FA	FA	M	PA	FA	M	M	PA		
		E7	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	M	PA	FA	M	M	PA		
		E8	Evaluation	FA	M	FA	M	FA	FA	FA	PA	FA	M	PA	FA	PA	M	PA		

EXPERIMENTS: 73

COUNT M:	19	73	10	73	4	10	5	5	6	29	2	19	46	67	22
COUNT PA:	0	0	54	0	39	47	29	59	16	44	68	21	25	0	51
COUNT FA	54	0	9	0	30	16	39	9	51	0	3	33	2	6	0
% M	26%	100%	14%	100%	5%	14%	7%	7%	8%	40%	3%	26%	63%	92%	30%
% PA	0%	0%	74%	0%	53%	64%	40%	81%	22%	60%	93%	29%	34%	0%	70%
% FA	74%	0%	12%	0%	41%	22%	53%	12%	70%	0%	4%	45%	3%	8%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

OPTIMIZATION: 26

COUNT M:	10	26	3	26	2	4	1	5	6	13	2	10	21	24	6
COUNT PA:	0	0	22	0	14	18	9	14	5	13	22	4	4	0	20
COUNT FA	16	0	1	0	10	4	16	7	15	0	2	12	1	2	0
% M	38%	100%	12%	100%	8%	15%	4%	19%	23%	50%	8%	38%	81%	92%	23%
% PA	0%	0%	85%	0%	54%	69%	35%	54%	19%	50%	85%	15%	15%	0%	77%
% FA	62%	0%	4%	0%	38%	15%	62%	27%	58%	0%	8%	46%	4%	8%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

EVALUATION: 38

COUNT M:	7	38	6	38	2	5	3	0	0	13	0	7	19	36	14
COUNT PA:	0	0	24	0	18	23	18	38	9	25	37	15	18	0	24
COUNT FA	31	0	8	0	18	10	17	0	29	0	1	16	1	2	0
% M	18%	100%	16%	100%	5%	13%	8%	0%	0%	34%	0%	18%	50%	95%	37%
% PA	0%	0%	63%	0%	47%	61%	47%	100%	24%	66%	97%	39%	47%	0%	63%
% FA	82%	0%	21%	0%	47%	26%	45%	0%	76%	0%	3%	42%	3%	5%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

GENERALIZATION: 5

COUNT M:	0	5	1	5	0	1	0	0	0	2	0	0	3	4	1
COUNT PA:	0	0	4	0	4	2	1	3	1	3	5	1	2	0	4
COUNT FA	5	0	0	0	1	2	4	2	4	0	0	4	0	1	0
% M	0%	100%	20%	100%	0%	20%	0%	0%	0%	40%	0%	0%	60%	80%	20%
% PA	0%	0%	80%	0%	80%	40%	20%	60%	20%	60%	100%	20%	40%	0%	80%
% FA	100%	0%	0%	0%	20%	40%	80%	40%	80%	0%	0%	80%	0%	20%	0%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Characterization of TSE papers

Venue	Paper #	Experiment	Type	Hypotheses	Variables identification					Operationalization		Design		Population	Analysis		Validity evaluation	Artifact	
				Research	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics		Availability	Badge
TSE'19	AP46	E1	Optimization	M	PA	M	FA	FA	FA	FA	FA	M	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E3	Optimization	FA	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E4	Generalization	FA	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E5	Generalization	FA	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E6	Evaluation	FA	PA	M	FA	FA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
TSE'20	AP47	E1	Optimization	M	PA	M	PA	PA	FA	FA	FA	M	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
		E4	Generalization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	PA	M	PA		
TSE'20	AP48	E1	Optimization	M	PA	M	PA	PA	FA	FA	FA	M	PA	FA	M	M	PA	Yes	No
		E2	Optimization	M	PA	M	PA	PA	FA	FA	FA	M	PA	FA	M	M	PA		
		E3	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E4	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E5	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
		E6	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
TSE'20	AP49	E1	Optimization	M	FA	M	FA	PA	FA	FA	FA	M	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	FA	M	FA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Evaluation	FA	FA	M	FA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E4	Evaluation	FA	FA	M	FA	PA	FA	PA	FA	PA	PA	FA	FA	FA	PA		
TSE'21	AP50	E1	Evaluation	FA	PA	M	PA	FA	FA	PA	FA	PA	PA	FA	PA	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	FA	FA	PA	PA	PA	PA	FA	PA	M	PA		
		E3	Optimization	FA	PA	M	PA	FA	FA	FA	PA	PA	PA	FA	PA	M	PA		
		E4	Evaluation	FA	PA	M	PA	FA	FA	PA	FA	M	PA	FA	M	M	PA		
		E5	Optimization	M	PA	M	PA	FA	FA	FA	FA	M	PA	FA	PA	M	PA		
TSE'21	AP51	E1	Optimization	M	PA	M	PA	PA	FA	FA	M	PA	PA	FA	M	M	PA	Yes	No
		E2	Evaluation	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
		E3	Optimization	FA	PA	M	PA	PA	FA	PA	FA	PA	PA	FA	M	M	PA		
TSE'21	AP52	E1	Evaluation+Optimization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	M	PA	Yes	No
		E2	Evaluation+Optimization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	M	PA		
		E3	Evaluation+Optimization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	M	PA		
		E4	Generalization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	FA	PA		
		E5	Generalization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	M	PA		
		E6	Generalization	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	PA	M	PA		
TSE'21	AP55	E1	Evaluation	FA	PA	M	FA	PA	FA	PA	FA	PA	PA	FA	PA	M	FA	Yes	No
		E2	Generalization	FA	PA	M	FA	PA	FA	PA	FA	PA	PA	FA	PA	M	FA		
		E3	Generalization	FA	PA	M	FA	PA	FA	PA	FA	PA	PA	FA	PA	M	FA		
TSE'21	AP54	E1	Evaluation	FA	PA	M	PA	PA	FA	PA	PA	PA	PA	FA	M	M	PA	Yes	No
		E2	Generalization	FA	PA	M	PA	PA	FA	PA	PA	M	PA	FA	M	M	PA		
TSE'21	AP53	E1	Evaluation	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	FA	PA	FA	PA	No	No
		E2	Optimization	FA	PA	M	PA	PA	PA	PA	FA	PA	PA	FA	PA	FA	PA		
		E3	Generalization	FA	PA	M	PA	PA	PA	FA	FA	PA	PA	FA	PA	M	PA		
		E4	Optimization	FA	PA	M	PA	PA	PA	FA	FA	PA	PA	FA	PA	M	PA		

Characterization of TSE papers

Venue	Paper #	Experiment	Type	Hypotheses	Variables identification					Operationalization		Design		Population	Analysis		Validity evaluation	Artifact	
				Research	Model hyperparameters	Model parameters	DL algorithm	Training hyperparameters	Training data	Factors and treatments	Response variables	Choice of design	Instrumentation	Test dataset	Descriptive statistics	Inferential statistics		Availability	Badge
EXPERIMENTS:		**43**																	
			COUNT M:	7	0	43	0	0	0	0	1	8	0	0	13	30	0		
			COUNT PA:	0	39	0	30	32	4	33	10	35	43	0	20	0	40		
			COUNT FA	36	4	0	13	11	39	10	32	0	0	43	10	13	3		
			% M	16%	0%	100%	0%	0%	0%	0%	2%	19%	0%	0%	30%	70%	0%		
			% PA	0%	91%	0%	70%	74%	9%	77%	23%	81%	100%	0%	47%	0%	93%		
			% FA	84%	9%	0%	30%	26%	91%	23%	74%	0%	0%	100%	23%	30%	7%		
			TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		
OPTIMIZATION:		**12**																	
			COUNT M:	7	0	12	0	0	0	0	1	6	0	0	7	10	0		
			COUNT PA:	0	11	0	9	8	2	3	1	6	12	0	4	0	12		
			COUNT FA	5	1	0	3	4	10	9	10	0	0	12	1	2	0		
			% M	58%	0%	100%	0%	0%	0%	0%	8%	50%	0%	0%	58%	83%	0%		
			% PA	0%	92%	0%	75%	67%	17%	25%	8%	50%	100%	0%	33%	0%	100%		
			% FA	42%	8%	0%	25%	33%	83%	75%	83%	0%	0%	100%	8%	17%	0%		
			TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		
EVALUATION:		**18**																	
			COUNT M:	0	0	18	0	0	0	0	0	1	0	0	5	10	0		
			COUNT PA:	0	15	0	12	13	1	18	2	17	18	0	6	0	17		
			COUNT FA	18	3	0	6	5	17	0	16	0	0	18	7	8	1		
			% M	0%	0%	100%	0%	0%	0%	0%	0%	6%	0%	0%	28%	56%	0%		
			% PA	0%	83%	0%	67%	72%	6%	100%	11%	94%	100%	0%	33%	0%	94%		
			% FA	100%	17%	0%	33%	28%	94%	0%	89%	0%	0%	100%	39%	44%	6%		
			TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		
GENERALIZATION:		**10**																	
			COUNT M:	0	0	10	0	0	0	0	0	1	0	0	1	7	0		
			COUNT PA:	0	10	0	6	8	1	9	4	9	10	0	7	0	8		
			COUNT FA	10	0	0	4	2	9	1	6	0	0	10	2	3	2		
			% M	0%	0%	100%	0%	0%	0%	0%	0%	10%	0%	0%	10%	70%	0%		
			% PA	0%	100%	0%	60%	80%	10%	90%	40%	90%	100%	0%	70%	0%	80%		
			% FA	100%	0%	0%	40%	20%	90%	10%	60%	0%	0%	100%	20%	30%	20%		
			TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		