Automatic IMRAD Classification of Citation Contexts: Comparing Text Representations and Machine Learning Classifiers

Francis Lareau[1,2,3]

[1]Université de Sherbrooke, 2500, boulevard de l'Université Sherbrooke (Québec), J1K 2R1, Canada
[2]Université du Québec à Montréal, 1430, rue Saint-Denis, Montréal (Québec), H2X 3J8, Canada
[3]Centre interuniversitaire de recherche sur la science et la technologie, 1430, rue Saint-Denis, Montréal (Québec), H2X 3J8, Canada

Abstract
Identifying the rhetorical function of citations within scientific articles is essential to advance citation analysis and develop intelligent writing assistance tools. Citation contexts serve distinct argumentative roles depending on their location within the standardized IMRAD structure (Introduction, Methods, Results, Discussion). To our knowledge, no comprehensive evaluation of machine learning approaches for automated IMRAD citation classification exists. This study systematically evaluated 172 combinations of representation-classifiers in a corpus of 303,947 annotated citation contexts from the scientific literature to predict the placement of sections. We compared nine text representation methods—ranging from classical bag-of-words to fine-tuned BERT variants and OpenAI embeddings—with five classifier categories (linear, neural, tree-based, instance-based, and unsupervised). The results show that domain-specific fine-tuning substantially enhances performance, with BERT fine-tuned on IMRAD classification achieving a weighted F-measure of 0.753 when paired with linear classification. Linear classifiers generally outperformed neural and other approaches, suggesting that IMRAD-specific linguistic patterns create linearly separable feature spaces. Embeddings derived from an earlier OpenAI model achieved competitive performance (F1=0.722), indicating that LLM-derived representations capture relevant structural information. Interestingly, embeddings from a newer OpenAI model yielded worse performance (F1=0.709), indicating that newer models do not uniformly improve structural representation quality. Unsupervised clustering methods performed poorly (F1 = 0.685), confirming that IMRAD section inference requires labelled training data. These findings establish the feasibility of automatic citation context classification and highlight the importance of task-specific model adaptation for scientific text analysis. The work has implications for bibliometric studies, intelligent authoring systems, and research evaluation frameworks seeking to quantify the rhetorical impact of citations across article sections.

Keywords
argument mining, citation context, IMRAD, OpenAI, embeddings, text representation, BERT, machine learning

1. Introduction

The IMRAD structure—Introduction, Methods, Results, and Discussion—has become the standardized format for scientific communication in most disciplines. This standardization facilitates knowledge dissemination and allows readers to locate specific information within research articles rapidly. Within this structure, citations serve a critical rhetorical function: they ground arguments in existing knowledge, establish scholarly credibility, and signal the scientific contribution being advanced.

Citation contexts—the passages in a text containing one or more references [1]—are not uniformly distributed across IMRAD sections; rather, they serve distinct argumentative functions depending on their location. Prior work has shown that citation contexts exhibit systematic lexical variation across IMRAD sections, notably in the distribution of verbs surrounding citations. Complementary large-scale analyses have also demonstrated that the position and age of references follow an almost invariant pattern along the IMRAD structure, with introductions and discussions concentrating most citations. Our study builds on these insights by treating IMRAD section identification as a supervised classification problem at the sentence level [2, 3]. The ability to automatically predict the IMRAD section

SCOLIA '26: Second International Workshop on Scholarly Information Access (SCOLIA), April 2, 2026, Delft, The Netherlands
lareau.francis@courrier.uqam.ca (F. Lareau)
0000-0002-0352-5246 (F. Lareau)

from which a citation context originates could have important applications in research evaluation, document analysis, and the design of intelligent tools for scientific writing.

Despite the structural importance of IMRAD organization in scientific writing, little is known about the computational feasibility of identifying citation context placement within this structure. Previous studies have examined citation contexts broadly, including whether cited works are peripheral or fundamental to the citing texts' arguments [4, 5, 6]. To our knowledge, no one has systematically evaluated multiple machine-learning approaches for IMRAD classification. This study addresses this gap by comparing text representation techniques and classification methods to predict affiliation of citation contexts with IMRAD sections.

This research is guided by the following primary research question: **Can machine learning models reliably predict the IMRAD section from which a citation context originates?** We test three specific hypotheses. First, we posit that citation contexts contain sufficient linguistic and structural markers to enable accurate IMRAD classification. Second, we hypothesize that advanced text representation methods inspired by large language models, such as GPT-based embeddings, outperform classical approaches including bag-of-words and doc2vec. Third, we expect neural network classifiers to demonstrate superior performance compared to linear, tree-based, instance-based, and unsupervised approaches. To test these hypotheses, we compared nine distinct text representation methods with five classification methodologies (45 combined models total) on a corpus of 303,947 annotated citation contexts from scientific articles. We employed the In-text Reference Corpus (InTeReC), a well-established dataset with explicit IMRAD section labels, enabling robust evaluation across representation-classifier combinations.

2. Methods

As illustrated in Figure 1, our experimental pipeline consists of corpus preprocessing, vector representation of the text, training classification families of models, and subsequent evaluation.

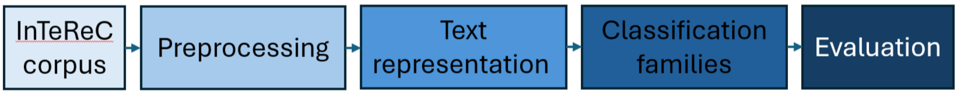

Figure 1: Experimental workflow.

2.1. Data and Corpus Description

The **In-text Reference Corpus** (InTeReC) served as the foundation for this study [7]. InTeReC comprises 314,023 citation contexts extracted from 90,071 open-access scientific articles published in seven Public Library of Science (PLOS) journals between 2001 and 2013. Each article was automatically partitioned into IMRAD sections using XML tags, and sentences containing explicit citations were extracted and annotated with their source section (Introduction, Methods, Results, Discussion). The corpus exhibits class imbalance, with Discussion sections being most represented (37.5%), followed by Introduction (23.5%), Methods (17.7%), and Results (9.0%). We removed 10,076 sentences labelled with overlapping categories from the original corpus (Methods & Results, Results & Discussion), resulting in the final annotated dataset of 303,947 sentences. The corpus was split into training (80%, n=243,157) and evaluation sets (20%, n=60,790). All experiments were conducted with a fixed random seed so that corpus splits and model initializations remained identical across configurations. We ran each representation–classifier combination once, rather than repeating training with multiple seeds. This design ensures a fair, controlled comparison between models while keeping the overall computational cost low.

2.2. Representation Methods

We used classical and modern vectorial text representations. The first, **BOW**, is a classic “bag of words" representation in which words are simply counted. The second, **DBOW**, is a doc2vec representation of the dbow variant. This model is generated using a simple neural network that learns to predict a masked word from its context. The third, **OAI1**, is an early model developed by OpenAI formally known as “text-embedding-ada-002". The fourth, **OAI2**, is the newer model of the OpenAI family named “text-embedding-3-large". Finally, the last one is a series of BERT-type representations generated using Transformers pre-trained on an extensive corpus of text. Bert is a bidirectional model that considers both the context and the word order in the text. We used the original representations of the basic BERT model - **BERT Base** – and the large BERT model – **BERT Large**. Then, three (3) fine-tuned representation models are used. The first, **BERT NLISTS**, is based on BERT Large and has been fine-tuned on natural language inference (NLI) and semantic textual similarity (STS) tasks [8]. The second, **BERT MPNet**, is based on BERT Base and is fine-tuned on a masked and permutated language modelling network (MPNet) task [9]. The last, **BERT IMRAD**, is based on the BERT-base model and is fine-tuned for IMRAD classification using BERT for sequence classification.[1]

2.3. Classification Methods

Various machine learning classification techniques enable us to perform the task at hand.

Clustering models group similar data points into clusters based on their similarity. The principle underlying these classifiers is to identify emergent groupings in the data automatically. This classification does not require labelled training data and can therefore be used for unsupervised learning tasks.[2] Note that clustering-based classification can be used in cases where labelled training data is available, such as to evaluate whether the classes or labels assigned to the data can match automatically detectable clusters without a priori knowledge of the classes to be discovered. When this type of approach solves a classification problem, it indicates that the feature space can be readily decomposed into implicit clusters corresponding to the classes of interest. We tested Gaussian mixture and K-means [12].

Instance-based classification models are a class of algorithms whose classification principle is the similarity of new cases to instances observed during training. These models are generally nonparametric, in that they require no assumptions about the distribution of the underlying data. These use all cases of the training dataset when modelling, and new data points are ranked based on their similarity to the training data. We tested the methods of the K-nearest neighbors [13], radius neighbors [14], and nearest centroid [15].

Linear classification models use linear decision boundaries to separate classes in feature space. Additionally, kernel-based linear classification enables the classification of non-linearly separable data in the original feature space by mapping the input data to a higher-dimensional space in which linear separation is possible. We tested stochastic gradient descent [16], support vector machines [17] (with and without kernel-based approaches), logistic regression [18], Ridge regression [17, 19], and passive-aggressive classification [20].[3]

Tree-based classification models use branched decision structures to classify data. The principle of tree classifiers is to recursively divide the space of features into subsets that are more homogeneous with respect to the target variable. We tested the decision trees [21], random forests [22], extremely randomized trees [23], and gradient boosting [24].

Neural classification models are a type of machine learning inspired by the structure of the human brain. The principle of neural classification is to approximate a function that maps input data to output labels by adjusting the weights of connections between neurons. This is achieved by iteratively

[1]We use Gensim [10] for BOW and DBOW; OpenAI’s API for OAI, and the Hugging Face API for BERT-type corpus representation models [11].

[2]Clustering-based classification does not identify which category corresponds to which label. However, we approximate the performance on the labelled data by assigning each label to the cluster that globally maximizes the overall F1 score.

[3]Although some algorithms contain ‘regression’ in their names, we use their standard classification variants that output discrete IMRAD labels.

propagating the input data through the network and adjusting the weights based on the errors between the predicted and actual labels. We tested the perceptron [25], the multilayer perceptron [26], and BERT for sequence classification.[4]

2.4. Evaluation Methodology

Model performance was assessed using the weighted F-measure (F1). This metric was chosen to address class imbalance and to provide a single interpretable performance score. The weighted F1 $F^{(w)}$ aggregates the class-specific F_1-scores by weighting each class according to its relative frequency in the dataset. For each class c, the weight is defined as $n_c/\sum_{j=1}^{C} n_j$, where n_c is the number of instances in class c and C is the total number of classes. The class-specific F_1 score is computed as the harmonic mean of precision and recall. The resulting weighted F-measure is therefore given by:

$$F^{(w)} = \sum_{c=1}^{C} \left(\frac{n_c}{\sum_{j=1}^{C} n_j} \right) \cdot \left(\frac{2\,\mathrm{Precision}_c\,\mathrm{Recall}_c}{\mathrm{Precision}_c + \mathrm{Recall}_c} \right)$$

We computed the weighted F1 for all possible combinations of representation and classification methods, yielding 172 models.[5] For each family of models (a given representation method and classifier type), we summarize performance by the maximum weighted F1 within that family. We further report the overall best-performing configuration as the maximum of these family-wise maxima.

3. Results

Analysis of the 45 joint representation-classifier type combinations revealed substantial variation in performance (Table 1). The maximum weighted F1 scores ranged from 0.292 (clustering with DBOW representation) to 0.753 (linear classification with BERT IMRAD representation).

Representation Model Performance: The highest-performing representation methods, based on the maximum F1 score across families of classifiers, were BERT IMRAD (F1=0.753), followed by OAI1 *text-embedding-ada-002* (F1=0.722), BERT Base (F1=0.711), BERT Large (F1=0.7094), and OAI2 *text-embedding-3-large* (F1=0.709). Conversely, the poorest-performing representation methods, identified using the minimum of the maximum F1 scores across families of classifiers, were DBOW (F1=0.410), BERT MPNet (F1=0.627), BOW (F1=0.635), and BERT NLISTS (F1=0.689). Notably, fine-tuning BERT directly on the IMRAD classification task yielded substantially superior performance compared to non-fine tuned BERT models, which themselves outperformed models fine-tuned on transfer learning tasks (MPNet, NLISTS). The relatively strong performance of the non-fine tuned OpenAI embedding model—without task-specific fine-tuning—suggests that LLM-derived representations capture valuable domain-general linguistic patterns relevant to IMRAD classification. Interestingly, the most recent OpenAI model exhibited inferior performance on this task compared to the earliest model (F1=0.709 vs 0.722).

Classification Method Performance: Performance varied substantially across classification approaches when assessed using the maximum of maxima across representation methods. Linear classifiers achieved the highest performance (F1=0.753), closely followed by tree-based classifiers (F1=0.752). Neural classifiers ranked third (F1=0.745), while instance-based classifiers exhibited lower performance (F1=0.732). Clustering methods performed notably poorly, with a maximum F1 score of 0.685. The superior performance of supervised methods relative to unsupervised clustering suggests that IMRAD section placement involves class-specific patterns that require labelled training data to capture effectively.

Overall, the optimal configuration paired BERT IMRAD with linear classification (F1=0.753). This model outperformed other top configurations, such as linear classification with early OpenAI embeddings (F1=0.722) or neural classification with BERT IMRAD (F1=0.745).

[4]We use Hugging Face for BERT for sequence classification, and Scikit-learn with default parameters for other methods

[5]Nine (9) representation techniques multiplied by nineteen classification methods, plus one end-to-end model.

Table 1
Maximum of the weighted f-measures of the 45 joint models, i.e., nine (9) representation models multiplied by five (5) classification types. The maximum of maxima is used to order the performance of representation (left to right) and classification (top to bottom) models.

	BERT IMRAD	OAI1	BERT Base	BERT Large	OAI2	BERT NLISTS	BOW	BERT MPNet	DBOW
Linear	0.753	0.722	0.711	0.7094	0.7090	0.689	0.635	0.627	0.410
Tree-based	0.752	0.633	0.642	0.650	0.640	0.595	0.587	0.533	0.388
Neural	0.745	0.658	0.652	0.633	0.655	0.640	0.588	0.573	0.436
Instance-based	0.732	0.590	0.614	0.612	0.611	0.576	0.556	0.533	0.378
Clustering	0.685	0.435	0.445	0.470	0.443	0.419	0.379	0.360	0.292

4. Discussion

4.1. Key Findings and Interpretation

Our results provide several important insights into IMRAD classification of citation contexts:

Efficacy of Domain-Specific Fine-Tuning: The marked superiority of BERT IMRAD (F1=0.753) over other BERT variants and general-purpose embeddings demonstrates that task and domain-specific fine-tuning substantially enhances performance for specialized text classification tasks. Fine-tuning BERT for this IMRAD classification task produced a more separable representation space than either generic BERT models or fine-tuning on unrelated tasks (NLISTS, MPNet). This finding suggests that the benefits of transfer learning depend critically on the alignment between the fine-tuning objective and the target task.

Prevalence of Linearity in the Feature Space: The slightly superior performance of linear classifiers relative to neural approaches (F1 = 0.753 vs. 0.745) suggests that the fine-tuned BERT representation yields a feature space where relatively simple, approximately linear decision boundaries already capture most of the signal needed for IMRAD classification. This does not imply that neural networks are fundamentally less capable, but rather that, in our setting, linear models may be easier to optimize reliably: linear classifiers with convex loss functions have a single global optimum and are well handled by standard solvers, whereas neural networks involve non-convex objectives and can converge to sub-optimal solutions under default hyperparameters and limited tuning.

Limited Benefits of Unsupervised Approaches: The poor performance of clustering methods (F1=0.685) indicates that the IMRAD structure cannot be reliably identified through unsupervised learning. This suggests that while implicit partitions may exist in the feature space, they do not align with IMRAD categories without explicit labelled training data. Although citation contexts across distinct IMRAD sections may display similar surface-level linguistic patterns, accurately identifying their differences requires supervised learning methods capable of capturing subtle discriminative features.

LLM-Derived Embeddings as a Practical Alternative: The strong out-of-the-box performance of OpenAI's *text-embedding-ada-002* (F1=0.722) suggests that embeddings derived from large language models capture sufficient linguistic information for reasonable IMRAD classification without task-specific fine-tuning. While this approach underperforms BERT IMRAD, it offers a practical advantage for researchers lacking the computational resources or labelled data for fine-tuning. Furthermore, our results indicate that increased model recency and architectural sophistication do not necessarily yield superior performance for this task. Specifically, the more recent *text-embedding-3-large* model did not outperform the oldest model in our experiments (F1 = 0.709 vs. 0.722).

4.2. Limitations and Considerations

Several limitations warrant acknowledgment. First, this study evaluated models on a single corpus (PLOS journals, 2001-2013); generalization to other disciplines, time periods, or publication venues remains untested. Second, citation contexts in scientific articles may differ systematically across genres (e.g., editorials, reviews, policy documents, etc.). Third, the class imbalance in the InTeReC corpus may advantage majority classes (Discussion). Fourth, we rely on a single random seed and one run per model, so our scores do not capture variability arising from stochastic training effects. Future work should complement our findings with multi-seed experiments and confidence intervals.

4.3. Implications for Research and Practice

These findings hold significance for multiple communities. For Scientific Text Mining, the demonstrated feasibility of automatic IMRAD classification enables new analyses of how citations function across article sections, potentially revealing disciplinary differences in citation practices and argumentative structures. For Argument Mining, automated IMRAD classification provides a structural foundation for moving beyond generic citation analysis toward richer rhetorical network graphs. For Intelligent Writing Assistance, these models could be integrated into writing tools to provide authors with real-time feedback about citation context placement, supporting more rhetorically effective positioning of citations. For Bibliometric Studies, automated IMRAD classification could enhance citation analysis by contextualizing citations within their structural location, enabling more granular investigation of research influence and knowledge transfer patterns.

4.4. Future Research Directions

Several avenues merit further exploration, including cross-domain evaluation to assess whether models trained on PLOS articles generalize to citations from other domains. Another promising direction is the fine-tuning of large language model embeddings, particularly OpenAI embeddings, to determine whether task-specific adaptation can reduce the performance gap with BERT IMRAD. Ensemble methods also warrant investigation, as combining linear and neural classifiers may balance interpretability with expressive power. In addition, systematic error analysis could help identify linguistic phenomena responsible for misclassifications and guide refinements in feature engineering. Finally, evaluating temporal generalization would clarify whether models trained on older articles remain effective on contemporary publications, thereby testing the temporal stability of IMRAD-specific linguistic markers.

5. Conclusions

This study demonstrates that machine learning can effectively predict the IMRAD section from which a citation context originates, with optimal performance achieved through domain-specific fine-tuning of BERT combined with linear classification (F1=0.753). The strong performance of fine-tuned models underscores the value of task-specific adaptation, while the competitive performance of non-fine tuned LLM-derived embeddings indicates that modern language models capture relevant linguistic structure without explicit fine-tuning. Linear classifier superiority suggests that IMRAD classification leverages relatively interpretable linguistic patterns, contrasting with the opacity of deep neural approaches. These findings advance our understanding of how machine learning can be applied to structure recognition in scientific writing and open pathways for practical applications in research evaluation, writing assistance, and bibliometric analysis.

Acknowledgments

This study was funded by the Canadian Social Sciences and Humanities Research Council (grant number 756-2024-0557). Parts of this work were developed within the framework of the author's PhD thesis.

Declaration on Generative AI

During the preparation of this work, the author used Grammarly to check grammar, spelling and improve prose. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] H. Small, Citation context analysis, in: B. Dervin, J. Voigt (Eds.), Progress in Communication Sciences, volume 3 of *Progress in communication sciences*, Ablex Publishing Corporation, Michigan, 1982, pp. 287–310. URL: https://books.google.ca/books?id=4kViAAAAMAAJ, issue: vol. 3.

[2] M. Bertin, I. Atanassova, A study of lexical distribution in citation contexts through the IMRaD standard, in: C. W. Proceedings (Ed.), Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval co-located with 36th European Conference on Information Retrieval (ECIR 2014), volume 1143 of *Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval co-located with 36th European Conference on Information Retrieval (ECIR 2014)*, 2014, pp. 5–12. URL: https://hal.science/hal-01940845.

[3] M. Bertin, I. Atanassova, Y. Gingras, V. Larivière, The invariant distribution of references in scientific articles 67 (2016) 164–177. URL: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23367. doi:10.1002/asi.23367, publisher: John Wiley & Sons, Ltd.

[4] M. H. Anderson, R. K. Lemken, An Empirical Assessment of the Influence of March and Simon's Organizations: The Realized Contribution and Unfulfilled Promise of a Masterpiece, Journal of Management Studies 56 (2019) 1537–1569. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/joms.12527. doi:10.1111/joms.12527, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joms.12527.

[5] A. Burton-Jones, What have we learned from the Smart Machine?, Information and Organization 24 (2014) 71–105. URL: https://www.sciencedirect.com/science/article/pii/S1471772714000104. doi:10.1016/j.infoandorg.2014.03.001.

[6] M. Lounsbury, E. J. Carberry, From King to Court Jester? Weber's Fall from Grace in Organizational Theory, Organization Studies 26 (2005) 501–525. URL: https://doi.org/10.1177/0170840605051486. doi:10.1177/0170840605051486, publisher: SAGE Publications Ltd.

[7] M. Bertin, I. Atanassova, InTeReC: In-text Reference Corpus for Applying Natural Language Processing to Bibliometrics, volume 2080 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 54–62. URL: http://ceur-ws.org/Vol-2080/paper6.pdf.

[8] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: http://arxiv.org/abs/1908.10084, arXiv:1908.10084 [cs].

[9] K. Song, X. Tan, T. Qin, J. Lu, T.-Y. Liu, MPNet: Masked and Permuted Pre-training for Language Understanding, 2020. URL: http://arxiv.org/abs/2004.09297. doi:10.48550/arXiv.2004.09297, arXiv:2004.09297 [cs].

[10] R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: M. Valletta (Ed.), Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks, University of Malta, 2010, pp. 46–50. URL: https://is.muni.cz/publication/884893/en/Software-Framework-for-Topic-Modelling-with-Large-Corpora/Rehurek-Sojka.

[11] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-Art Natural Language Processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://aclanthology.org/2020.emnlp-demos.6. doi:10.18653/v1/2020.emnlp-demos.6.

[12] S. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory 28 (1982) 129–137. doi:10.1109/TIT.1982.1056489, conference Name: IEEE Transactions on Information Theory.

[13] E. Fix, J. J. L. Hodges, Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties, Technical Report ADA800276, California Univ. Berkeley, 1951. URL: https://apps.dtic.mil/sti/citations/ADA800276, section: Technical Reports.

[14] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1967) 21–27. URL: http://ieeexplore.ieee.org/document/1053964/. doi:10.1109/TIT.1967.1053964.

[15] J. J. Rocchio, Relevance feedback in information retrieval, in: G. Salton (Ed.), The Smart retrieval system - experiments in automatic document processing, Prentice-Hall, Englewood Cliffs, NJ, 1971, pp. 313–323. URL: https://www.bibsonomy.org/bibtex/c18d843e34fe4f8bd1d2438227857225.

[16] H. Robbins, S. Monro, A Stochastic Approximation Method, The Annals of Mathematical Statistics 22 (1951) 400–407. URL: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-22/issue-3/A-Stochastic-Approximation-Method/10.1214/aoms/1177729586.full. doi:10.1214/aoms/1177729586, publisher: Institute of Mathematical Statistics.

[17] V. Vapnik, Pattern recognition using generalized portrait method, Automation and Remote Control (1963) 774–780. URL: https://www.semanticscholar.org/paper/Pattern-recognition-using-generalized-portrait-Vapnik/7cabbdf6a7288d15e26fa6ea504009bab3d1edf4.

[18] J. Berkson, Application of the Logistic Function to Bio-Assay, Journal of the American Statistical Association 39 (1944) 357–365. URL: https://www.jstor.org/stable/2280041. doi:10.2307/2280041, publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[19] A. E. Hoerl, R. W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics 12 (1970) 55–67. URL: https://www.jstor.org/stable/1267351. doi:10.2307/1267351, publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

[20] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, Y. Singer, Online Passive-Aggressive Algorithms, Journal of Machine Learning Research 7 (2006) 551–585. URL: http://jmlr.org/papers/v7/crammer06a.html.

[21] J. N. Morgan, J. A. Sonquist, Problems in the Analysis of Survey Data, and a Proposal, Journal of the American Statistical Association 58 (1963) 415–434. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500855. doi:10.1080/01621459.1963.10500855.

[22] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32. URL: https://doi.org/10.1023/A:1010933404324. doi:10.1023/A:1010933404324.

[23] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Machine Learning 63 (2006) 3–42. URL: https://doi.org/10.1007/s10994-006-6226-1. doi:10.1007/s10994-006-6226-1.

[24] J. H. Friedman, Greedy function approximation: A gradient boosting machine., The Annals of Statistics 29 (2001). URL: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full. doi:10.1214/aos/1013203451.

[25] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychological Review 65 (1958) 386–408. doi:10.1037/h0042519, place: US Publisher: American Psychological Association.

[26] D. E. Rumelhart, G. Hinton, R. J. Williams, Learning Internal Representations by Error Propagation, Technical Report ADA164453, University of San Diego, California, 1985. URL: https://apps.dtic.mil/sti/citations/ADA164453, section: Technical Reports.