Multilevel Machine Learning

Research Report

Paolo Colussi

p.colussi@uu.nl

Utrecht University

Supervisors:

Word count:
2486

1 Introduction

Multilevel (hierarchical) data, in which lower-level units are nested within higher-level units, are ubiquitous—for example, children in schools or patients in hospitals. Units belonging to the same group typically share contextual conditions and latent characteristics, which makes them more similar to each other than to units belonging to different groups. This poses challenges, since many commonly used statistical and machine learning methods assume that observations are independent and identically distributed (i.i.d.), an assumption that does not hold in the presence of clustering (Snijders and Bosker 2012). Ignoring this structure in statistical analysis violates independence assumptions and can lead to misleading results (H. Wu 2006; Liang and Zeger 1986). When within-group dependence is overlooked, models tend to underestimate uncertainty, overstate the amount of information in the data, and learn patterns that reflect group-specific systematic differences rather than generalise relationships (Snijders and Bosker 2012; Shahbazi and Azadeh-Fard 2025). These issues can be amplified in high-dimensional settings, where flexible machine learning methods such as decision trees or random forests are often preferred but still assume independent observations (Hastie, Tibshirani, and Friedman 2009; Hajjem, Bellavance, and Larocque 2011).

Traditional approaches to hierarchical data rely on linear and generalised linear mixed-effects models, which account for clustering through random effects and provide valid inference under appropriate assumptions (Gelman and Hill 2006; Snijders and Bosker 2012). Extensions incorporating regularisation, which prevents overfitting, such as mixed-model LASSO variants, have been proposed to address high-dimensional settings (Schelldorfer, Meier, and Bühlmann 2014; L. Yang and T. T. Wu 2023). By contrast, many machine learning methods—including random forests and neural networks—typically assume i.i.d. observations (James et al. 2021; Hastie, Tibshirani, and Friedman 2009). To bridge these frameworks, Hajjem, Bellavance, and Larocque (2011) introduced an expectation maximisation (EM) based algorithm to estimate random effects conventionally while using regression trees or random forests for the fixed part of linear regressions. This was later extended to generalised linear regressions (Hajjem, Larocque, and Bellavance 2017). Related EM-derived tree methods include RE-EM trees (Sela and Simonoff 2012) and the M-CART (Lin and Luo 2019). Other multilevel tree methods which do not rely on EM have been developed (Fokkema et al. 2018; Fontana et al. 2021) and approaches based on Bayesian or probabilistic trees (Speiser et al. 2020; Wundervald, Parnell, and Domijan 2023). These ideas generalise to forests through MERF (Mixed effect random forest) and its successors (Hajjem, Bellavance, and Larocque 2014; Ngufor et al. 2019; Pellagatti et al. 2021; Bergonzoli, Rossi, and Masci 2024), as well as Bayesian forest extensions (Speiser et al. 2019). Boosting-based mixed-effects models have also been proposed (Sigrist 2022; Sigrist 2024). Beyond tree-based algorithms, few multilevel ML methods exist, with neural-network adaptations representing a rare exception (Crane-Droesch 2017; Zhang et al. 2021).

Despite recent multilevel extensions of machine learning, not much has been done to systematically compare these models. The consequences of applying standard machine learning algorithms to clustered data remain only partly understood, especially for predictive accuracy and variable-importance reliability (Hastie, Tibshirani, and Friedman 2009; Register 2025). Comparative studies between classical mixed-effects models and machine learning methods that do or do not account for clustering are limited and often restricted to narrow applications (Shahbazi and Azadeh-Fard 2025; Register 2025; D'Ottaviano and W. Yang 2022; Mangino and Finch 2021). Moreover, the robustness of existing approaches to violations of assumptions about fixed effects, random effects, or the data-generating process is not well established. In high-dimensional settings, it is also poorly understood, particularly when clustering interacts with sparsity, multicollinearity or complex nonlinear relationships. Consequently, we still lack a clear picture of how these methods respond to hierarchical structure and what types of inferential or modelling errors arise when it is ignored.

In light of these limitations, this study asks: *under which conditions do multilevel machine-learning models outperform multilevel regression models, and how does ignoring clustered data affect the performance of machine-learning methods in binary classification?* As a reference I also include standard regression models and their multilevel expansion. The study focuses on binary classification. The rest of the report is organised in two main sections. The methodology section describes the modelling framework: logistic regression, multilevel logistic regression, tree-based models, GMERT/GMERF, the simulation design, and evaluation metrics. The simulation section presents results across linear, nonlinear, interaction, multicollinearity, and

high-dimensional scenarios, followed by potential extensions.

2 Methodology

To address the research question, I conduct a simulation study comparing single-level and multilevel versions of different modelling approaches. The analysis focuses on binary outcomes and considers four main families of methods: generalized linear models (logistic regression), generalized linear multilevel models (logistic mixed-effects regression), tree-based algorithms (classification trees and random forests), and their multilevel extensions (GMERT and GMERF). Each method is evaluated across a set of controlled data-generating mechanisms that vary the strength and form of the hierarchical structure as well as the dimensionality of the predictor space. This design makes it possible to compare models under comparable conditions and to examine how explicitly modelling the multilevel structure affects their behaviour across different scenarios.

2.1 Logistic Regression

The first technique considered is logistic regression, which provides the baseline single-level model for binary outcomes $y_i \in \{0, 1\}$. It assumes that all observations i are independent and relates the probability of success to a linear predictor η_i

$$\eta_i = x_i^\top \beta,$$

where x_i is a for each element i a column vector of p features, and β the p regression coefficients. Through the logit link

$$Pr(y_i = 1 \mid x_i) = p_i = \text{logit}^{-1}(\eta_i),$$

where $\text{logit}(p) = \log(\frac{p}{1-p})$ and $\text{logit}^{-1}(\eta_i) = \frac{1}{1+\exp(\eta_i)}$. Under this formulation,

$$y_i \sim \text{Bernoulli}(p_i),$$

and parameter estimation is carried out by maximizing the binomial log-likelihood with respect to β (Gelman and Hill 2006). Logistic regression serves as a natural benchmark: it ignores clustering entirely, allowing comparison with methods that explicitly incorporate the hierarchical structure.

2.2 Generalised Linear Multilevel Model

To account for clustering, logistic regression can be extended by introducing group-specific random effects. For observation i in cluster j, the linear predictor becomes

$$\eta_{ij} = x_{ij}^\top \beta + z_{ij}^\top b_j,$$

where $b_j \sim \mathcal{N}(0_p, \mathbf{D})$ is a vector of random effects for cluster j, such as a random intercept or random slopes, capturing how that cluster deviates from the population-level coefficients. The vector z_{ij} is the corresponding design vector indicating which random effects apply to observation i; for instance, $z_{ij} = 1$ in a random-intercept model, or $z_{ij} = (1, x_{ij})$ when both a random intercept and slope are included.

The outcome is modelled as

$$y_{ij} \sim \text{Bernoulli}\big(\text{logit}^{-1}(\eta_{ij})\big),$$

and inference targets both the fixed effects β and the covariance matrix $\mathbf{D}$ governing the distribution of the random effects. Because the likelihood involves integrating over b_j, it lacks a closed-form expression; estimation therefore relies on approximations such as Laplace or adaptive Gaussian quadrature, which provide maximum-likelihood (or restricted maximum-likelihood) estimates under the assumed hierarchical structure (Gelman and Hill 2006)

Logistic mixed-effects models offer a principled way to model within-cluster dependence and form the classical benchmark for hierarchical binary data.

2.3   Tree-based techniques

Classification trees partition the predictor space into terminal regions $R_1, \dots, R_T$ by recursively splitting the data. At each step, the algorithm selects the predictor and cutpoint that best improve node purity, typically measured by the Gini index

$$G(R) = \hat{p}_R(1 - \hat{p}_R),$$

where $\hat{p}_R$ is the proportion of observations with $y = 1$ in region R. The predicted probability for any x in a terminal node is simply $\hat{p}_R$ (Hastie, Tibshirani, and Friedman 2009).

Random forests extend trees by averaging predictions across many bootstrapped trees, each grown using random subsets of predictors and a random sample of the with replacement. For a new observation x, the forest estimate is

$$\hat{p}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{p}_{T_b}(x),$$

which reduces variance and yields more stable predictions.

Both trees and forests model nonlinearities and interactions without distributional assumptions but treat observations as independent (Hastie, Tibshirani, and Friedman 2009). They therefore form the single-level machine-learning baselines for comparison with multilevel extensions such as GMERT and GMERF.

2.4   GMERT and GMERF

Multilevel extensions of tree-based methods aim to embed the hierarchical structure of generalized linear mixed-effects models within a non-parametric framework. In a GLMM for binary outcomes, the linear predictor can be written as

$$\eta_{ij} = \underbrace{f(x_{ij})}_{fixed\ part} + \underbrace{z_{ij}^\top b_j}_{random\ part},$$

where $f(x_{ij})$ represents the fixed part, describing the population-level relationship between predictors and the response, and $b_j \sim N(0, D)$ represents the random part, modelling cluster-specific deviations. Standard GLMMs restrict the fixed part to the parametric form $f(x_{ij}) = x_{ij}^\top \beta$. GMERT (Generalised Mixed Effect Regression Tree) replaces this linear component with a regression tree, allowing the fixed part to take an arbitrary piecewise-constant structure while preserving the usual random-effects representation.

To estimate the model, Hajjem et all. (2017) adapt the penalized quasi-likelihood (PQL) algorithm used in GLMMs. At each iteration, the model is linearized around the current fitted values, yielding a pseudo-response y_{ij}^* and working weights. Conditioned on the current estimate of the tree, the random effects are updated through the mixed-model equations that arise from the approximate linear model

$$y_{ij}^* = f(x_{ij}) + z_{ij}^\top b_j + \varepsilon_{ij},$$

3

treating the fixed part as known. Once the random effects are updated, the fixed part is re-estimated by fitting a weighted regression tree to the adjusted data $y_{ij}^* - z_{ij}^\top b_j^{(t+1)}$, in the same spirit as the MERT algorithm for continuous outcomes. Variance components such as D are then updated using the closed-form expressions derived from the mixed-effects approximation. These steps—PQL linearization, random-effects update, tree refitting, variance-component update—are iterated until the linear predictor converges.

This procedure contrasts with standard GLMM inference, where the fixed effects are obtained through parametric likelihood optimization (e.g., Laplace approximation or adaptive Gaussian quadrature) (Breslow and Clayton 1993). In GMERT, the optimization step for the fixed part is replaced entirely by recursive partitioning applied to the PQL-adjusted data, meaning that the structure of $f(\cdot)$ is learned non-parametrically at every iteration rather than being determined by a pre-specified linear predictor.

By embedding a regression tree inside the mixed-effects estimation cycle, GMERT produces a model that retains the hierarchical interpretation of a GLMM while allowing the fixed part to adapt flexibly to non-linearities and interactions. GMERF (Generalised Mixed Effect Random Forest) follows the same logic, replacing the tree step by a random forest to further stabilize and regularize the non-parametric fixed-effects estimate (Hajjem, Larocque, and Bellavance 2017; Pellagatti et al. 2021).

2.5   Simulating data

To compare the four modelling approaches under controlled conditions, clustered binary data are generated from a logistic mixed-effects data-generating process. For observation i in cluster j,

$$y_{ij} \sim Bernoulli(logit^{-1}(f(x_{ij}) + b_{0j} + b_{1j}x_{ij,1}),$$

where $(b_{0j}, b_{1j})^\top \sim N(0, D)$ induces within-cluster dependence through a random intercept and a random slope on x_1. Across all scenarios, the hierarchical structure, number of clusters, cluster size, and random-effects covariance are held fixed. The scenarios differ only in the specification of the fixed-effects component $f()$, allowing the impact of increasing fixed-effect complexity to be isolated.

The simulation design is summarized in Table 1 and Table 2.

Table 1: Cluster structure and predictor generation

Component	Specification
Number of clusters (G)	30
Observations per cluster (n_i)	60
Random effects	Random intercept and random slope on x_1
Random-effects distribution	$(b_{0j}, b_{1j})^\top \sim \mathcal{N}(0, D)$
x_1	Uniform$(-2, 2)$
x_2	Normal (scenario-specific mean, unit variance)
x_3	Bernoulli(0.5)
x_4	Bernoulli(0.5) (Scenario 3 only)
Correlated predictors (Scenario 4)	$(x_1, x_2, x_3^*) \sim \mathcal{N}(0, \Sigma_X)$, $x_3 = \mathbb{I}(x_3^* > 0)$

Table 2: Fixed-effects specifications and design features across simulation scenarios

Scenario	Fixed-effects structure	Design feature
Baseline	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	Correctly specified linear GLMM
Nonlinear	$\beta_0 + \beta_1 x_1 + \beta_2 \log(x_2) + \beta_3 x_3$	Nonlinear covariate effect
Interaction	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 x_4 + \beta_3 x_3$	Interaction between predictors
Multicollinearity	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	Strong correlation among predictors

All models are evaluated using cluster-balanced 10-fold cross-validation to prevent information leakage across clusters.

2.6 Evaluation metrics

Performance is evaluated using accuracy, F1 scores, and a simple prediction bias measure. Accuracy summarizes overall classification performance but can obscure differences in how models handle the event and non-event classes (Hastie, Tibshirani, and Friedman 2009). I therefore report two F1 scores—one for the majority class and one for the minority class—each defined as

$$F1 = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}},$$

which provides a balanced measure of classification quality when precision–recall trade-offs differ across models.

To assess systematic over- or under-prediction, I compute a relative bias index

$$\text{Bias} = \frac{\sum \widehat{y}}{\sum y} - 1.$$

Positive values indicate a tendency to over-predict the event, negative values the reverse.

3 Preliminary results

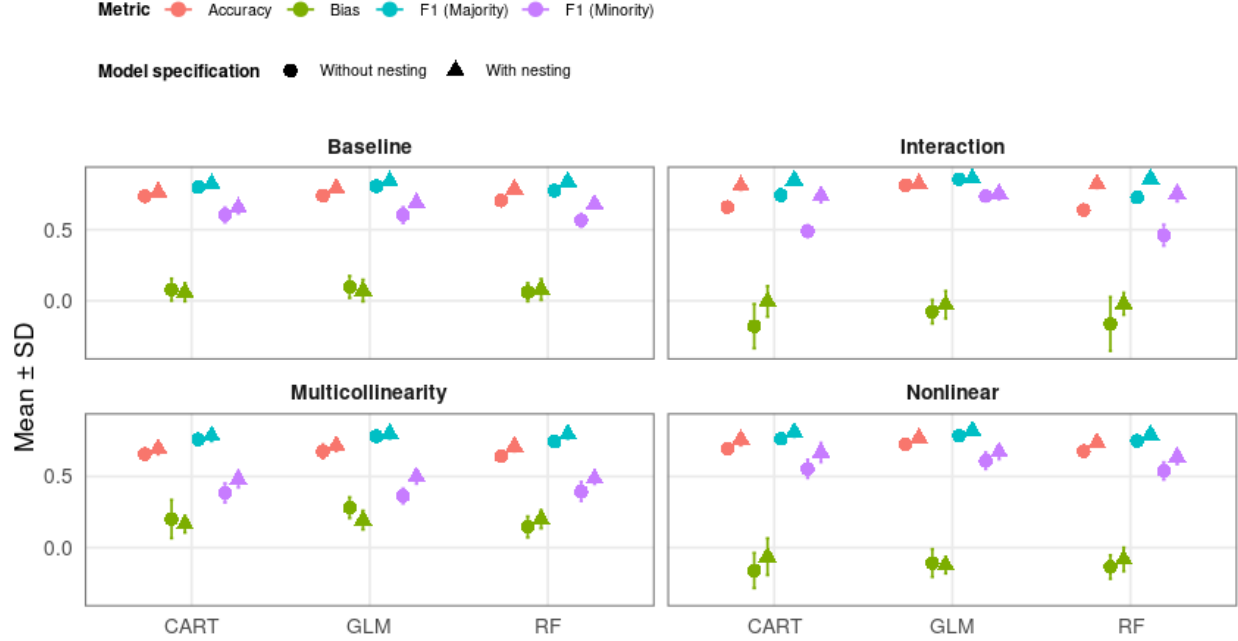Figure 1: Model performance across simulation scenarios

Table 3: F1 scores across simulation scenarios

Model	Baseline	Nonlinear	Interaction	Multicollinearity
Logit	0.805	0.786	0.856	0.782
CART	0.800	0.765	0.743	0.761
RF	0.775	0.750	0.727	0.745
GLMM	0.840	0.822	0.860	0.802
GMERT	0.819	0.814	0.849	0.781
GMERF	0.837	0.780	0.859	0.796

The results show a consistent pattern across modelling strategies for clustered data. Figure 1 reports model performance across the different scenarios. Since all evaluation metrics display very similar patterns, the analysis focuses on the majority-class F1 score, summarized in Table 3. When the fixed-effects structure is simple, performance differences are limited: the GLMM performs best, followed closely by GMERT and GMERF. The small gap between CART and random forests reflects the low dimensionality of the predictor space, which limits the benefits of feature subsampling.

As the data-generating process departs from this baseline through nonlinearities, interactions, or multicollinearity, differences become more pronounced. Single-level tree-based methods show a clearer loss in performance, whereas models incorporating random effects remain substantially more stable.

Across all settings, single-level machine-learning methods tend to perform worse than their multilevel extensions, even when the predictor space is small and the data-generating process is relatively simple. Introducing random effects through GMERT and GMERF is associated with systematic improvements, suggesting that accounting for within-cluster dependence is a key driver of predictive performance. At the same time, differences between multilevel machine-learning models and the GLMM vary across scenarios, indicating that the

benefits of flexible fixed-effect modelling depend on the structure of the underlying signal. These results motivate further investigation into the conditions under which multilevel machine-learning approaches provide substantive advantages, and into how they might be extended or refined in more complex or high-dimensional settings.

4 Next steps

4.1 Real data application

As a complement to the simulation study, the proposed methods could be applied to large-scale administrative data from Statistics Netherlands (CBS). One possible application concerns residential vacancy data, with a binary vacancy indicator and covariates describing dwellings, households, and socio-demographic characteristics across multiple hierarchical levels. Another potential use case involves survey process data, such as response or dropout indicators, covering roughly seven million units nested within samples and surveys from 2018–2024, with predictors available at each level. These applications would allow exploration of how different modelling strategies behave in complex, high-dimensional real-world settings and how explicitly modelling clustering and higher-level features affects predictive performance and stability.

4.2 More methods

Further extensions may explore additional multilevel modelling strategies. Hierarchical neural networks could capture complex nonlinear patterns while accounting for clustering (Zhang et al. 2021). Boosting-based mixed-effects models offer an alternative that combines stagewise fitting with random effects (Sigrist 2024), while Gaussian process approaches naturally encode multilevel dependence (Sigrist 2022). In high-dimensional contexts, penalized mixed-effects models such as GLMM-LASSO provide a principled alternative to standard regression (Schelldorfer, Meier, and Bühlmann 2014), even if they are not machine-learning methods in a strict sense.

4.3 More scenarios

A further extension concerns simulations in high-dimensional settings, which are particularly relevant for assessing multilevel methods when the number of predictors is large. These scenarios are not implemented at this stage due to computational constraints. Future work could explore alternative simulation designs, including settings with sparsity, different cluster sizes, or additional higher-level covariates, to examine how information at multiple hierarchical levels interacts with model complexity and multilevel corrections.

5 Conclusions

The results suggest that explicitly modelling clustering plays a central role in improving predictive performance in hierarchical settings, even when the predictor space is relatively small. Building on this evidence, the analysis will be extended to high-dimensional scenarios, where issues such as sparsity, multicollinearity, and complex interactions are likely to interact more strongly with the multilevel structure. These extensions will be explored first through targeted simulation designs and then through applications to high-dimensional administrative data. This two-step approach will make it possible to assess how multilevel machine-learning methods and classical mixed-effects models compare when both hierarchical dependence and high dimensionality are present, and to evaluate the robustness of the observed patterns beyond controlled settings.

References

Bergonzoli, Giulia, Lidia Rossi, and Chiara Masci (June 5, 2024). *Ordinal Mixed-Effects Random Forest*. DOI: 10.48550/arXiv.2406.03130. arXiv: 2406.03130 [stat]. Pre-published.

Breslow, N. E. and D. G. Clayton (Mar. 1993). "Approximate Inference in Generalized Linear Mixed Models". In: *Journal of the American Statistical Association* 88.421, pp. 9–25. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1993.10594284.

Crane-Droesch, Andrew (May 18, 2017). *Semiparametric Panel Data Models Using Neural Networks*. DOI: 10.48550/arXiv.1702.06512. arXiv: 1702.06512 [stat]. Pre-published.

D'Ottaviano, Fabio and Wenzhao Yang (Nov. 2, 2022). "On Missing Random Effects in Machine Learning". In: *Communications in Statistics - Simulation and Computation* 51.11, pp. 6320–6331. ISSN: 0361-0918, 1532-4141. DOI: 10.1080/03610918.2020.1801729.

Fokkema, M. et al. (Oct. 2018). "Detecting Treatment-Subgroup Interactions in Clustered Data with Generalized Linear Mixed-Effects Model Trees". In: *Behavior Research Methods* 50.5, pp. 2016–2034. ISSN: 1554-3528. DOI: 10.3758/s13428-017-0971-x.

Fontana, Luca et al. (July 2021). "Performing Learning Analytics via Generalised Mixed-Effects Trees". In: *Data* 6.7, p. 74. ISSN: 2306-5729. DOI: 10.3390/data6070074.

Gelman, Andrew and Jennifer Hill (Dec. 18, 2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st ed. Cambridge University Press. ISBN: 978-0-521-86706-1 978-0-521-68689-1 978-0-511-79094-2. DOI: 10.1017/CBO9780511790942.

Hajjem, Ahlem, François Bellavance, and Denis Larocque (Apr. 2011). "Mixed Effects Regression Trees for Clustered Data". In: *Statistics & Probability Letters* 81.4, pp. 451–459. ISSN: 01677152. DOI: 10.1016/j.spl.2010.12.003.

– (June 3, 2014). "Mixed-Effects Random Forest for Clustered Data". In: *Journal of Statistical Computation and Simulation* 84.6, pp. 1313–1328. ISSN: 0094-9655, 1563-5163. DOI: 10.1080/00949655.2012.741599.

Hajjem, Ahlem, Denis Larocque, and François Bellavance (July 2017). "Generalized Mixed Effects Regression Trees". In: *Statistics & Probability Letters* 126, pp. 114–118. ISSN: 01677152. DOI: 10.1016/j.spl.2017.02.033.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-84857-0 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7.

James, Gareth et al. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. New York, NY: Springer US. ISBN: 978-1-0716-1417-4 978-1-0716-1418-1. DOI: 10.1007/978-1-0716-1418-1.

Liang, Kung-Yee and Scott L. Zeger (1986). "Longitudinal Data Analysis Using Generalized Linear Models". In: *Biometrika* 73.1, pp. 13–22. ISSN: 0006-3444, 1464-3510. DOI: 10.1093/biomet/73.1.13.

Lin, Shuqiong and Wen Luo (2019). "A New Multilevel CART Algorithm for Multilevel Data with Binary Outcomes". In: *Multivariate Behavioral Research* 54.4, pp. 578–592. ISSN: 1532-7906. DOI: 10.1080/00273171.2018.1552555. PMID: 30644764.

Mangino, Anthony A. and W. Holmes Finch (Dec. 2021). "Prediction With Mixed Effects Models: A Monte Carlo Simulation Study". In: *Educational and Psychological Measurement* 81.6, pp. 1118–1142. ISSN: 0013-1644. DOI: 10.1177/0013164421992818. PMID: 34565818.

Ngufor, Che et al. (Jan. 2019). "Mixed Effect Machine Learning: A Framework for Predicting Longitudinal Change in Hemoglobin A1c". In: *Journal of Biomedical Informatics* 89, pp. 56–67. ISSN: 1532-0480. DOI: 10.1016/j.jbi.2018.09.001. PMID: 30189255.

Pellagatti, Massimo et al. (2021). "Generalized Mixed-Effects Random Forest: A Flexible Approach to Predict University Student Dropout". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 14.3, pp. 241–257. ISSN: 1932-1872. DOI: 10.1002/sam.11505.

Register, Brennan (2025). "Comparing the Effectiveness of Standard vs. Multilevel Machine Learning Algorithms on Hierarchical Data". In: HDL: 1903/34077.

Schelldorfer, Jürg, Lukas Meier, and Peter Bühlmann (Apr. 3, 2014). "GLMMLasso: An Algorithm for High-Dimensional Generalized Linear Mixed Models Using ℓ_1-Penalization". In: *Journal of Computational and Graphical Statistics* 23.2, pp. 460–477. ISSN: 1061-8600, 1537-2715. DOI: 10.1080/10618600.2013.773239.

Sela, Rebecca J. and Jeffrey S. Simonoff (Feb. 1, 2012). "RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data". In: *Machine Learning* 86.2, pp. 169–207. ISSN: 1573-0565. DOI: 10.1007/s10994-011-5258-3.

Shahbazi, Marzieh Amiri and Nasibeh Azadeh-Fard (Sept. 1, 2025). "Hierarchical Data Modeling: A Systematic Comparison of Statistical, Tree-Based, and Neural Network Approaches". In: *Machine Learning with Applications* 21, p. 100688. ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2025.100688.

Sigrist, Fabio (July 25, 2022). *Latent Gaussian Model Boosting*. DOI: 10.48550/arXiv.2105.08966. arXiv: 2105.08966 [cs]. Pre-published.

– (Nov. 5, 2024). *Gaussian Process Boosting*. DOI: 10.48550/arXiv.2004.02653. arXiv: 2004.02653 [cs]. Pre-published.

Snijders, Tom A. B. and Roel J. Bosker (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2. ed. Los Angeles, Calif.: SAGE. 354 pp. ISBN: 978-1-84920-200-8 978-1-84920-201-5.

Speiser, Jaime Lynn et al. (Feb. 15, 2019). "BiMM Forest: A Random Forest Method for Modeling Clustered and Longitudinal Binary Outcomes". In: *Chemometrics and intelligent laboratory systems : an international journal sponsored by the Chemometrics Society* 185, pp. 122–134. ISSN: 0169-7439. DOI: 10.1016/j.chemolab.2019.01.002. PMID: 31656362.

– (Apr. 2, 2020). "BiMM Tree: A Decision Tree Method for Modeling Clustered and Longitudinal Binary Outcomes". In: *Communications in Statistics - Simulation and Computation* 49.4, pp. 1004–1023. ISSN: 0361-0918, 1532-4141. DOI: 10.1080/03610918.2018.1490429.

Wu, Hulin (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience. 369 pp. ISBN: 978-0-471-48350-2 978-0-470-00966-6.

Wundervald, Bruna, Andrew Parnell, and Katarina Domijan (Apr. 24, 2023). *Hierarchical Embedded Bayesian Additive Regression Trees*. DOI: 10.48550/arXiv.2204.07207. arXiv: 2204.07207 [stat]. Pre-published.

Yang, Luoying and Tong Tong Wu (June 2023). "Model-Based Clustering of High-Dimensional Longitudinal Data via Regularization". In: *Biometrics* 79.2, pp. 761–774. ISSN: 1541-0420. DOI: 10.1111/biom.13672. PMID: 35428983.

Zhang, Lei et al. (Jan. 2021). "Hierarchical Deep-Learning Neural Networks: Finite Elements and Beyond". In: *Computational Mechanics* 67.1, pp. 207–230. ISSN: 0178-7675, 1432-0924. DOI: 10.1007/s00466-020-01928-9.