iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73756

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Supervised Learning Approaches for Robust Predictive Modelling in Data Science

Varsharani T. Dond[1], Mohini K. Vaidya[2]
[1]*Assistant Professor, PVG's College of Science and Commerce, Maharashtra, India*
[2]*Assistant Professor, Haribhai V. Desai College, Maharashtra, India*

Abstract: Supervised learning remains the dominant paradigm for predictive modeling in data science, yet real-world deployments frequently fail due to fragile data pipelines, distributional shift, and optimistic evaluation. This article surveys supervised learning approaches with a focus on robustness—defined as the stability of predictive performance under perturbations to data, environment, or assumptions. We organize the model space into seven families: linear and generalized linear models; tree-based models; kernel methods; instance-based methods; probabilistic generative models; neural networks; and ensemble learning. For each family we discuss inductive biases, optimization, computational complexity, calibration, and typical failure modes. We then synthesize a method-agnostic workflow spanning dataset auditing, leakage prevention, feature engineering, resampling, hyperparameter tuning, model selection, and post-hoc reliability analysis (calibration, uncertainty, and drift monitoring). Robustness strategies—regularization, data augmentation, adversarial training, cost-sensitive learning, resampling for class imbalance, monotonic constraints, conformal prediction, and causal sensitivity analysis—are reviewed with practical guidance. Case vignettes from healthcare, finance, and operations illustrate trade-offs between accuracy, interpretability, and reliability. The paper concludes with open research directions, including integrating causal structure into supervised objectives, leveraging self-supervised pretraining for tabular data, distributionally robust optimization, and aligning evaluation with societal impact.

Keywords: supervised learning, robustness, predictive modeling, model selection, calibration, uncertainty quantification, distribution shift, class imbalance, regularization, conformal prediction

I. INTRODUCTION

Data-driven decision systems have accelerated across domains such as healthcare, finance, logistics, and public policy. Supervised learning—learning a function mapping inputs to labeled outputs—forms the backbone of prediction at scale. While algorithmic advances have reduced the gap between research and practice, model fragility remains a critical barrier to trust. Optimistic validation, unrecognized label noise, covariate shift, and data leakage frequently inflate expected generalization, leading to degraded performance once deployed.

Consequently, model robustness—the capacity to maintain predictive quality amid data imperfections and environmental changes—has become a primary design objective rather than a secondary concern.

This paper makes three contributions. First, we provide a structured review of major supervised learning families through the lens of robustness, highlighting inductive biases and failure modes. Second, we propose a pragmatic, auditable workflow for robust predictive modeling from raw data to monitored deployment. Third, we identify emerging research directions that combine statistical rigor with real-world constraints—notably distributionally robust optimization, conformal prediction, and causal perspectives.

We adopt a problem-agnostic perspective but ground discussion in tabular, time-series, and simple text/vision scenarios commonly encountered in applied data science. Our emphasis is on classification and regression. We assume the reader is familiar with basic probability and optimization.

II. BACKGROUND AND TAXONOMY OF SUPERVISED LEARNING

Supervised learning estimates a function (f:) given pairs ((x_i, y_i)). For regression, (=); for classification, () is a finite label set. A learning algorithm selects (f_) to minimize expected loss ([(f_(x), y)]), approximated by empirical risk with regularization. Robustness implicates bias–variance trade-offs, capacity control, loss functions, and optimization stability.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

A. Families of Models

1) Linear and Generalized Linear Models (GLMs). Ordinary least squares, ridge, lasso, elastic net, and logistic/Poisson regression impose linear decision boundaries with penalized coefficients. They excel in small-(n)/large-(p) settings with collinearity control and facilitate interpretability.
2) Decision Trees and Tree Ensembles. Single trees (CART, C4.5) capture nonlinear feature interactions with axis-aligned splits; ensembles (Random Forests, Gradient Boosting, XGBoost, LightGBM, CatBoost) improve accuracy and robustness via bagging, boosting, and regularization.
3) Kernel Methods. Support Vector Machines (SVMs) and Kernel Ridge Regression project inputs into high-dimensional feature spaces via kernels (RBF, polynomial), optimizing margin-based objectives.
4) Instance-based Methods. (k)-Nearest Neighbors (kNN) and related methods defer generalization to prediction time, relying on local neighborhoods in a metric space.
5) Probabilistic Generative Models. Naïve Bayes and Linear/Quadratic Discriminant Analysis model class-conditional densities under simplifying assumptions.
6) Neural Networks. Multilayer perceptrons and deep architectures (CNNs, RNNs/transformers for sequences) learn hierarchical representations via stochastic optimization.
7) Ensembles and Stacking. Bagging, boosting, stacking/blending aggregate diverse learners to reduce variance and exploit complementary strengths.

B. Robustness Notions

1) Statistical robustness: bounded influence of outliers (e.g., Huber loss), heavy-tailed noise tolerance.
2) Algorithmic robustness: stability under data perturbations (e.g., regularization strength, early stopping).
3) Distributional robustness: resilience to covariate, label, or concept shift; worst-case performance bounds under uncertainty sets (e.g., DRO).
4) Operational robustness: reliability across pipeline changes, missingness patterns, and latency constraints; monitoring drift and calibration.

III. MODEL FAMILIES: INDUCTIVE BIASES, STRENGTHS, AND FAILURE MODES

A. Linear and Generalized Linear Models

1) Strengths: closed-form or convex optimization; straightforward regularization (L2 for shrinkage, L1 for sparsity); well-calibrated probabilities for logistic regression; interpretability (coefficients, odds ratios).
2) Failure modes: misspecification in presence of nonlinear interactions; sensitivity to outliers (mitigated by robust losses); multicollinearity; extrapolation beyond support.
3) Robustness levers: feature scaling; Huber/Tukey losses; ridge/elastic net; interaction terms and splines; monotonic constraints; Bayesian priors for shrinkage; robust standard errors.

B. Decision Trees and Ensembles

Single trees are interpretable but high-variance. Random Forests reduce variance via bootstrap aggregation and random feature subsetting, naturally handle missingness and mixed data types, and are relatively robust to outliers. Gradient Boosting Machines (GBMs**)** fit residuals sequentially, offering strong accuracy with careful regularization (learning rate, max depth, subsampling, L1/L2 penalties, monotonic constraints). CatBoost mitigates target leakage from categorical encoding.

1) Failure modes: overfitting (deep trees, high learning rate), sensitivity to noisy labels in boosting, leakage via target encoding, biased importance measures.
2) Robustness levers: early stopping with stratified cross-validation, shrinkage, subsampling, monotonic constraints, honest splitting, permutation importance, SHAP-based sanity checks, and out-of-bag (OOB) validation.

C. Kernel Methods (SVMs)

1) Strengths: margin maximization confers robustness to small perturbations; effective in high-dimensional spaces with limited samples; hinge loss resists some outliers.
2) Failure modes: scaling to very large datasets; kernel/(C)/gamma sensitivity; probability calibration often required.

3) Robustness levers: nested cross-validation for kernel/regularization selection; Platt scaling/Isotonic regression; approximate kernels (Nyström, random Fourier features) for scalability.

D. Instance-based (kNN)

1) Strengths: simple, nonparametric; adapts to complex decision boundaries with sufficient data; naturally captures local structure.
2) Failure modes: curse of dimensionality; distance metric sensitivity; inference latency; sensitivity to class imbalance and noise.
3) Robustness levers: metric learning; dimensionality reduction (PCA, UMAP); distance-weighted voting; cleaning rules (Edited/Condensed kNN); anomaly removal.

E. Probabilistic Generative Models

1) Naïve Bayes is robust under conditional independence and extremely data-efficient; LDA/QDA assume Gaussian class-conditional distributions.
2) Failure modes: violated independence or Gaussian assumptions; poorly calibrated probabilities when assumptions fail.
3) Robustness levers: semi-naïve variants (TAN), feature selection, variance regularization, Bayesian smoothing.

F. Neural Networks

1) Strengths: universal function approximation; scalable with hardware; strong performance on unstructured data; flexible multi-task objectives.
2) Failure modes: overparameterization leading to optimization instabilities; sensitivity to label noise and adversarial perturbations; calibration errors (overconfident probabilities).
3) Robustness levers: weight decay, dropout, data augmentation, mixup/cutmix, early stopping, sharpness-aware minimization, robust losses (label smoothing, generalized cross-entropy), adversarial training, and post-hoc calibration (temperature scaling). Optimizers such as Adam and SGD with momentum balance speed and generalization.

G. Ensembles and Stacking

1) Strengths: reduce variance, hedge against misspecification, and often improve calibration; useful for tabular data where heterogeneous signals exist.
2) Failure modes: leakage via blending folds; complexity and maintainability; diminishing returns without diversity.
3) Robustness levers: strict out-of-fold (OOF) blending, simple meta-learners, diversity-promoting base learners, and ensembling calibrated probabilities instead of raw scores.

IV. DATA-CENTRIC ROBUSTNESS: AUDITS, PREPROCESSING, AND FEATURE ENGINEERING

Robustness begins with data. Key steps:

1) Data audits. Characterize missingness (MCAR/MAR/MNAR), outliers, class imbalance, leakage risks, duplicate leakage across splits, and temporal/spatial autocorrelation. Visual profiling and drift baselines (e.g., PSI, KS) help.
2) Preventing leakage. Enforce causal time ordering; group-aware and time-series splits; avoid target leakage in encoders and feature creation; confine preprocessing within cross-validation folds.
3) Handling missingness. Use model-native handling (e.g., XGBoost, LightGBM) or imputation pipelines (median/most frequent, iterative imputation); indicator flags for informative missingness.
4) Feature engineering. Domain features, interaction terms, monotonic transforms (logit/log), and robust scaling (quantile/robust scaler) can stabilize models. For categoricals, prefer target encoding with OOF discipline; for high-cardinality, use hashing or CatBoost’s ordered statistics.
5) Resampling for imbalance. Stratified sampling; cost-sensitive losses; class weighting; synthetic oversampling (SMOTE variants) within training folds only.
6) Label quality. Estimate label noise via consensus, weak supervision, or confident learning; consider noise-robust losses or relabeling workflows.

V. MODEL SELECTION, TUNING, AND EVALUATION

A. Resampling Schemes

1) Holdout with stratification for speed; k-fold cross-validation for stable estimates; nested CV for unbiased model selection when hyperparameter search is extensive.
2) Time-series CV (rolling-origin, purged K-fold) to respect temporal leakage.
3) GroupKFold when units (patients, customers) appear multiple times.

B. Metrics and Calibration

Select metrics aligned with decisions: RMSE/MAE for regression; accuracy, ROC-AUC, PR-AUC, F1 for classification; cost curves when misclassification costs are asymmetric. Always report calibration (Brier score, reliability diagrams). For imbalanced data, ROC-AUC can be misleading—prefer PR-AUC and cost-sensitive analyses.

C. Hyperparameter Tuning

Adopt coarse-to-fine search: defensible defaults → random search → Bayesian optimization. Guardrails: bounded search spaces, early stopping, and repeated stratified CV. Track compute budgets and carbon cost. Prefer simpler models when performance is statistically indistinguishable (Occam’s razor).

D. Uncertainty Quantification

Combine predictive intervals (quantile regression, conformal prediction), parameter uncertainty (bootstrapping, Bayesian inference), and stability analysis (jackknife, leave-one-group-out). Aggregate across resamples to characterize variance and to detect brittle pipelines.

E. Statistical Significance and Effect Sizes

Use McNemar’s test for paired classification, Diebold–Mariano for forecast comparison, and bootstrap confidence intervals for metric differences. Report effect sizes and minimum detectable effects to avoid p-hacking.

VI. ROBUSTNESS STRATEGIES BY FAILURE MODE

A. Noisy Labels and Outliers

1) Robust losses (Huber, Tukey biweight, generalized cross-entropy).
2) Label smoothing or soft labels from teacher models.
3) Early-learning regularization: stop before memorization, monitor training dynamics.
4) Data cleaning loops**:** uncertainty or influence functions to flag mislabeled samples.

B. Class Imbalance and Rare Events

1) Class weights or focal loss; threshold-moving with cost curves.
2) OOF SMOTE/ADASYN; anomaly detection for extreme imbalance; evaluation on PR-AUC and recall@k.

C. Covariate and Concept Shift

1) Drift detection (PSI, KL/KS tests); retraining triggers.
2) Domain adaptation: importance weighting, covariate shift correction; representation learning with invariant risk minimization.
3) Distributionally Robust Optimization (DRO): optimize worst-case risk over uncertainty sets (e.g., Wasserstein balls).
4) Conformal prediction to maintain coverage under mild exchangeability assumptions.

D. Missing Data and Measurement Error

1) Multiple imputation; model-native missing handling; noisy feature models.
2) Sensitivity analysis across plausible imputation mechanisms.

E. Interpretability, Fairness, and Governance

1) Global: coefficients, partial dependence, accumulated local effects, SHAP with caution (feature correlation caveats).
2) Local: LIME/SHAP for instance-level explanations with stability checks.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

3) Fairness auditing: group metrics (TPR/FPR parity, calibration), counterfactual tests, and remediation (reweighing, constraint-aware optimization).
4) Governance: model cards, datasheets, reproducibility checklists, and human-in-the-loop signoff.

VII. CASE VIGNETTES

A. Healthcare: Sepsis Early Warning (Binary Classification)
1) Data: ICU EHR time-series summarized into tabular features.
2) Approach: Baseline logistic regression with L2 and calibrated probabilities; GBM with monotonic constraints honoring clinical priors (e.g., higher lactate → higher risk).
3) Robustness: Grouped time-based CV to avoid patient leakage; calibration with isotonic regression; conformal risk control to flag uncertain predictions for clinician review.
4) Outcome: Slight AUROC gain for GBM over logistic, but improved recall@k with conformal triage; GBM adopted with human oversight.

B. Finance: Credit Default Prediction (Imbalanced)
1) Data: Loan applications with high-cardinality categoricals.
2) Approach: CatBoost with ordered target statistics; cost-sensitive thresholding to meet portfolio constraints.
3) Robustness: Reject inference requires calibrated probabilities and monotonic constraints on income/DTI; OOF encoding to prevent leakage; PR-AUC reporting.
4) Outcome: 8–12% lift in recall at fixed precision relative to legacy scorecard while maintaining interpretability via monotone partial dependence.

C. Operations: Demand Forecasting (Regression)
1) Data: Multi-seasonal retail time series with promotions and holidays.
2) Approach: Gradient boosting with lag/rolling features; quantile regression for P50/P90.
3) Robustness: Purged time-series CV; holiday leakage prevention; conformal intervals for service-level planning.
4) Outcome: Reduced stockouts and overstock through probabilistic forecasts; governance via model cards and drift dashboards.

VIII. REPRODUCIBLE AND AUDITABLE WORKFLOW
1) Problem framing: objective, constraints, decision threshold, harm analysis.
2) Data access & lineage: immutable snapshots; documented joins and filters; leakage checklist.
3) Preprocessing pipelines: encapsulated transformations fit only on training folds; schema checks.
4) Modeling: baselines first; hypothesis-driven feature engineering; controlled hyper parameter search.
5) Evaluation: stratified/nested CV; calibration checks; uncertainty and stability analysis.
6) Deployment: versioned artifacts; shadow mode; A/B or interleaved tests with guardrails.
7) Monitoring: performance, drift, calibration, data quality; retraining policies; incident response.
8) Documentation: model cards, datasheets; decisions and exceptions log; ethics review.

IX. DISCUSSION
No single supervised learner dominates across problems—the No Free Lunch intuition persists. Robust predictive modeling hinges less on algorithmic novelty and more on disciplined data work, conservative validation, and alignment with decision costs. Tree ensembles and regularized linear models remain strong baselines on tabular data; neural networks lead on high-dimensional unstructured modalities. Kernel methods occupy a sweet spot for medium-sized, high-dimensional problems where margins matter. Regardless of model, calibration and uncertainty quantification are necessary for safe decision-making.
A significant practical challenge is preventing and detecting leakage. Many published gains evaporate under stricter resampling or when time-ordering is respected. Another challenge is resolving the tension between accuracy and interpretability. Monotonic constraints, generalized additive models with pairwise interactions (GA2Ms), and post-hoc explanations partially bridge the gap but require care to avoid misleading narratives. Finally, robustness must include socio-technical considerations: fairness across groups, transparency for stakeholders, and governance for accountability.

X. FUTURE DIRECTIONS

1) Distributionally Robust Optimization (DRO): tractable uncertainty sets (f-divergence, Wasserstein) and connections to regularization for tabular tasks.
2) Conformal prediction at scale: efficient, adaptive coverage under covariate shift.
3) Causal representation learning: incorporating invariances that support robust extrapolation and counterfactual reasoning.
4) Self-supervised pretraining for tabular data: masked modeling and contrastive objectives to improve sample efficiency.
5) Neural–symbolic hybrids and monotone deep networks: embedding domain constraints to prevent pathological behavior.
6) Responsible ML by design: standardized robustness and harm audits required for deployment in regulated domains.

XI. CONCLUSION

Robust supervised learning in data science is less about finding a universally best algorithm and more about constructing a reliable end-to-end system. By aligning inductive biases with data properties, adopting leakage-safe evaluation, and quantifying uncertainty and calibration, practitioners can substantially improve real-world performance. Emerging techniques—DRO, conformal prediction, causal regularization, and self-supervised pretraining—promise further gains in reliability. The workflow and comparative guidance presented here aim to support Scopus-ready research and industry deployments alike.

REFERENCES

(Note: Ensure capitalization and punctuation are consistent with APA 7th; add DOIs/URLs where appropriate.)
[1] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
[3] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
[4] Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple Classifier Systems (pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1
[5] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
[6] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
[7] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01
[8] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. https://doi.org/10.1006/jcss.1997.1504
[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
[10] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580. (Dropout early report)
[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[12] Holland, P. W. (1986). Statistics and causal inference. Journal of the American Statistical Association, 81(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354
[13] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In International Conference on Learning Representations. https://arxiv.org/abs/1412.6980
[14] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In International Joint Conference on Artificial Intelligence (pp. 1137–1145).
[15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.
[16] Kull, M., Silva Filho, T., & Flach, P. (2017). Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration. Electronic Journal of Statistics, 11(2), 5052–5080. https://doi.org/10.1214/17-EJS1338SI
[17] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 30.
[18] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R., & Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523), 1094–1111. https://doi.org/10.1080/01621459.2017.1307116
[19] Liu, Y., Qi, Y., Li, J., & Tao, D. (2020). Adversarial examples: Attacks and defenses for deep learning. Springer. (For overview)
[20] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
[21] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In International Conference on Learning Representations.
[22] Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press.
[23] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
[24] Quinlan, J. R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.
[25] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD, 1135–1144. https://doi.org/10.1145/2939672.2939778
[26] Rousseeuw, P. J., & Leroy, A. M. (1987). Robust regression and outlier detection. Wiley.

[27] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 1929–1958.
[28] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.
[29] Tibshirani, R. J., Athey, S., Friedberg, R., Hadad, V., Miner, L. E., & Wager, S. (2020). Package ‘grf’: Generalized random forests. Journal of Computational and Graphical Statistics, 29(3), 629–653.
[30] Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In Contributions to Probability and Statistics (pp. 448–485). Stanford University Press.
[31] Vapnik, V. N. (1998). Statistical learning theory. Wiley.
[32] Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (3rd ed.). Academic Press. (For skill scores & forecast verification)
[33] Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software, 77(1), 1–17. https://doi.org/10.18637/jss.v077.i01
[34] Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. Proceedings of the Eighth ACM SIGKDD, 694–699. https://doi.org/10.1145/775047.775151
[35] Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. International Conference on Learning Representations.
[36] Zhang, Y., & Yang, Q. (2017). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 29(12), 431–447.
[37] (Add any domain-specific references or recent robust tabular deep learning papers as appropriate.)