BIT-ADMIT: A Machine Learning Framework for Automating International Student Admission and Scholarship Decisions

Tesfaye Yeabsira Mersha[1], Nada Tehhaf[2], Sarah Khalfi[3], Mohamed Amine Sebbarr[4]
[1]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
Email: yeasbira.tesfaye@bit.edu.cn

***Abstract*—Universities process thousands of international applications annually, with admission and scholarship decisions traditionally relying on manual evaluations. This approach is time-consuming, resource-intensive, and prone to bias, reducing fairness and consistency. This paper presents BIT-ADMIT, a machine learning framework automating admission and scholarship decisions for international applicants at Chinese universities. We generated 2,000 synthetic student records modeled on Beijing Institute of Technology's application form, including program type, country of origin, GPA, CSCA scores, research alignment, publications, recommendations, interview performance, and language proficiency. The dataset excludes personally identifiable information and was validated via entropy analysis, data-drift checks, and bias audits. After preprocessing, feature engineering, and normalization, six classifiers were trained: Random Forest, XGBoost, CatBoost, Gradient Boosting, Logistic Regression, and SVC. XGBoost and Gradient Boosting achieved top performance, with F1-scores of 0.998 for admission and 0.985 for scholarship predictions. The framework is deployed as a FastAPI web service enabling real-time inference. Though trained on synthetic data, distribution alignment tests indicate robustness and potential generalizability. Future work includes validation with anonymized real applications and exploring fairness metrics and bias detection mechanisms.**

***Index Terms*—Machine learning, educational data mining, admission prediction, scholarship allocation, automated decision systems, international student admission**

I. Introduction

A. Background and Motivation

Higher education institutions worldwide face increasing pressure to process international student applications efficiently, fairly, and at scale. Traditional admission and scholarship evaluation pipelines remain largely manual, requiring substantial human resources and time. These processes are inherently subject to evaluator bias, potentially resulting in inconsistent decisions, delayed notifications, and suboptimal allocation of qualified applicants [1], [2]. Machine learning has demonstrated promise in improving predictive accuracy and supporting transparent, data-driven decision-making across various educational contexts [3], [4]. However, its application to international student admissions in Chinese universities remains largely unexplored, representing a significant gap in both research and practice.

This project was conducted as a school project and does not constitute a published research paper.

B. International Education in China

China has emerged as a major destination for international students, driven by rapid economic growth, modernization of higher education infrastructure, and government initiatives promoting global talent acquisition. Historically, foreign students primarily enrolled in language programs, but recent trends show substantial increases in technical and professional degree programs, including engineering, medicine, economics, and management. According to the Ministry of Education (MOE), in 2018, over 492,000 international students from 196 countries studied at 1,004 higher education institutions across China, with Beijing and Shanghai hosting the largest concentrations. Over the past decade, the number of international students pursuing higher-education degrees increased by 13%, while government scholarships for international students expanded nearly six-fold. English-taught programs (ETPs) have grown by 12% between 2021 and 2023, positioning China as the largest global provider of ETPs outside English-speaking countries [5], [6], [7], [8].

To standardize and improve applicant quality assessment, China recently introduced the China Scholastic Competency Assessment (CSCA), a standardized test for international students pursuing undergraduate degrees. CSCA provides comprehensive evaluation of academic and language abilities, offering universities an objective reference for admission and scholarship decisions. This development underscores a growing emphasis on data-driven assessment methodologies in Chinese higher education.

C. Research Gap and Contribution

Despite these advances, evaluation of international student applications in China remains predominantly manual. This approach presents several critical challenges: it is resource-intensive, time-consuming, and susceptible to evaluator bias, leading to inconsistent decisions, notification delays, and suboptimal allocation of high-quality applicants [8], [9]. Furthermore, there exists a lack of standardized, data-driven frameworks to support fair and efficient decision-making. To our knowledge, no prior studies have applied machine learning to access the admission and scholarship decisions for international applicants at Chinese universities, highlighting a clear research gap that this work addresses.

TABLE I
SUMMARY OF GENERATED APPLICANT DATA (N = 2000)

Feature	Categories / Range	Distribution
Program Category	Undergraduate/ Postgraduate/ Chinese Language / Dual Degree	800 / 700 / 300 / 200
Country	15 countries	133 each
Degree Language	English-taught / Chinese-taught	1400 / 600
Quality Class	Low / Mid / High	1000 / 700 / 300
Admission Decision	Admitted / Rejected	600 / 1400
Scholarship Tier	None / Partial / Full	1200 / 300 / 100
GPA	0–4 scale	Low: 2.6±0.4, Mid: 3.2±0.3, High: 3.7±0.15
Math/Physics Score	0–10 scale	Low: 4.5±1.8, Mid: 6.0±1.5, High: 8.0±1.0
Research Alignment	0–10 scale	Low: 3.8±1.8, Mid: 5.5±1.5, High: 7.5±1.2
Recommendation Strength	0–10 scale	Low: 5.8±1.2, Mid: 7.2±1.0, High: 8.5±0.8
Interview Score	0–100 scale	Low: 65±12, Mid: 78±10, High: 88±8
English Test	IELTS / TOEFL / DUOLINGO	800 / 800 / 400
English Score	Normalized 0–9 / 0–120 / 0–160	English Test and Quality Class dependent
Chinese Proficiency	HSK 1–6	Program and Quality Class dependent
Publications	0–∞	Poisson: λ=0.1 (Low), 0.5 (Mid), 1.5 (High)

By applying BIT-ADMIT in the international student recruitment in Chinese Universities context, this work provides a technical pipeline and empirical evidence that ML models can support in the process of admission and scholarship decisions. The framework demonstrates potential to reduce human bias, improve processing efficiency, and enable fair, data-driven decision-making in higher education, complementing recent initiatives such as CSCA and aligning with broader research emphasizing fairness, transparency, and operational efficiency.

II. METHODOLOGY

A. Data Generation and Validation

Due to the absence of publicly available datasets containing international student application data to Chinese universities and the confidential nature of application records, we generated a synthetic dataset for model development and evaluation. The synthetic data generation process was designed to closely mirror the structure and statistical properties of real application data, based on Beijing Institute of Technology's application form specifications.

The dataset comprises 2,000 synthetic applicant records with the following features: program category, specific program applied, country of origin, degree language (English-taught or Chinese-taught), previous GPA, CSCA(mathematics and physics) score, research alignment score, publication count, recommendation strength, interview score, English test type and score, Chinese proficiency level (HSK), admission decision (binary: admitted/rejected), and scholarship tier (categorical: none/partial/full). All personally identifiable information (PII) was excluded to ensure privacy compliance.

Table I summarizes the distribution characteristics of the generated dataset. The data generation process incorporated realistic constraints and relationships between features. For instance, applicants were classified into three quality tiers (low, mid, high) with distinct statistical properties for academic metrics. Low-quality applicants were intentionally overrepresented to ensure sufficient rejected cases for balanced model training, although in practice student applications typically exhibit higher quality distributions.

The overall admission rate is 34.25%, and among admitted students, 13.50% received partial scholarships while 3.45% received full scholarships. All rejected applicants are categorized as receiving no scholarship. We use a random state of 42 to ensure reproducibility.

Comprehensive exploratory data analysis was performed to validate the synthetic dataset.

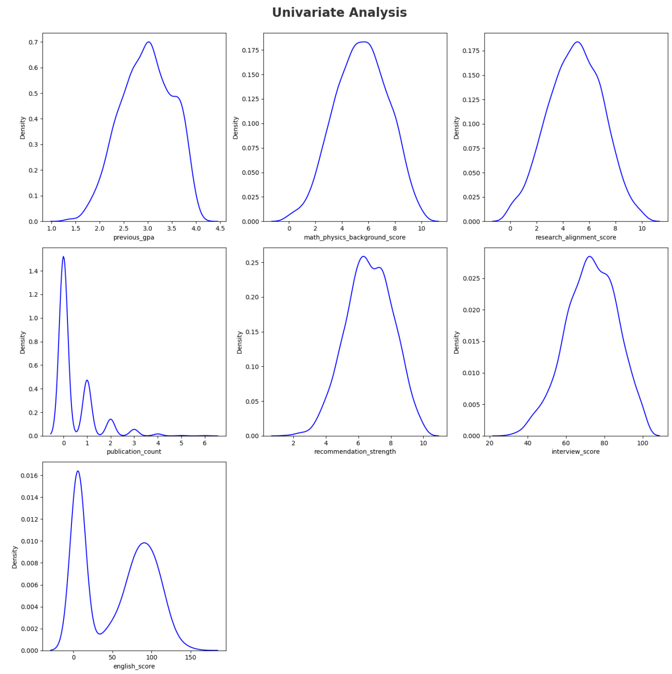

Fig. 1. Distributions of continuous numerical features including previous GPA, CSCA(math and physics) score, research alignment score, publication count, recommendation strength, and English score. The complete exploratory data analysis code and results are available in our GitHub repository.

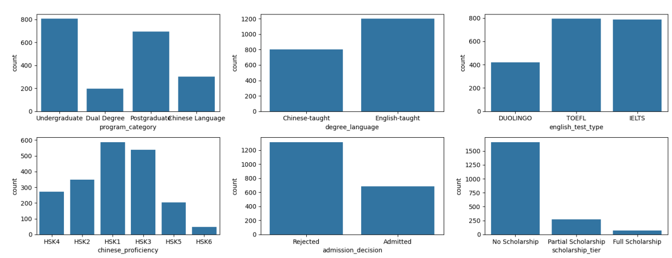

Fig. 2. Distributions of categorical features including program applied, degree language, English test type, Chinese proficiency level, admission decision, and scholarship tier. The complete exploratory data analysis code and results are available in our GitHub repository.

B. Data Preprocessing and Feature Engineering

The preprocessing pipeline consisted of several stages to prepare the raw data for model training. First, application ID was removed as it provides no predictive value and represents personally identifiable information. Helper features used exclusively for data generation (e.g., quality_class) were also excluded. All remaining features underwent standardization: strings were converted to lowercase, stripped of whitespace, and hyphens were replaced with underscores for consistency.

Chinese proficiency was converted from categorical HSK levels to numerical values (1–6). Target variables (admission_decision and scholarship_tier) were separated, encoded, and stored in a distinct dataframe for supervised learning.

The dataset was split into training (80%) and testing (20%) subsets using stratified sampling (random_state=42) to maintain class distribution balance across both sets. Feature engineering was performed to enhance model performance and capture domain-specific relationships:

- Publication count transformation: Applied log transformation ($\log_1 p$(publication_count)) to reduce skewness in the distribution and mitigate the influence of outliers.
- Language requirement flag: Created binary feature language_requirement_passed encoding whether applicants meet minimum English or Chinese proficiency thresholds:
  - English-taught programs: TOEFL $\geq$ 90, IELTS $\geq$ 6.5, or Duolingo $\geq$ 90
  - Chinese-taught programs: HSK level $\geq$ 4
- Weighted composite score: Developed a program-specific weighted scoring mechanism combining key applicant features:
  - Undergraduate programs: 40% previous GPA, 30% CSCA(math/physics), 10% recommendation strength, 20% interview score
  - Postgraduate programs: 40% previous GPA, 30% research alignment, 10% publications (capped at 5), 10% recommendation strength, 10% interview score
  - Chinese language and dual degree programs: 50% previous GPA, 20% recommendation strength, 30% interview score

The final feature set for model training comprised:

Numerical features: previous GPA, math/physics background score, research alignment score, recommendation strength, interview score, english score, publication count (log-transformed), weighted score, language requirement passed
Categorical features: program category, degree language, english test type, bit program applied, country, chinese proficiency
Target variables: admission decision (binary) and scholarship tier (multi-class)

C. Model Selection and Configuration

We evaluated six classical machine learning classifiers with a random seed of 42 set across all experiments to ensure reproducibility. The following configurations were implemented:

- Random Forest Classifier: Maximum tree depth limited to 10, minimum samples required to split an internal node set to 5, minimum samples at leaf nodes set to 2.
- XGBoost Classifier: 100 boosting iterations with maximum tree depth of 6, learning rate of 0.1, L1 regularization (reg_alpha) of 0.1, and L2 regularization (reg_lambda) of 0.1.
- CatBoost Classifier: 100 boosting iterations with tree depth of 6 and learning rate of 0.1. Verbose output suppressed during training.
- Gradient Boosting Classifier: 100 boosting iterations with maximum tree depth of 5 and learning rate of 0.1.
- Logistic Regression: Regularization strength C=1.0 with default L2 penalty, maximum 100 iterations for optimization convergence.
- Support Vector Classifier (SVC): Regularization parameter C=1.0, probability estimates enabled (probability=True) for ROC-AUC computation, unlimited iterations (max_iter=-1) allowing convergence.

All models were trained on the preprocessed training set and evaluated on the held-out test set.

III. Results

A. Model Performance Comparison

After training, all six models were evaluated on two distinct prediction tasks: (1) admission decision prediction (binary classification) and (2) scholarship tier prediction (multi-class classification). Performance was assessed using accuracy, F1-score, and AUC-ROC metrics on both training and test datasets to evaluate generalization capability.

For the admission decision prediction task (Table II), XGBoost Classifier achieved the highest performance with test accuracy of 0.9975, F1-score of 0.998, and AUC of 0.997. CatBoost and Gradient Boosting classifiers demonstrated comparable performance with test F1-scores of 0.994. Random Forest achieved moderate performance (F1-score of 0.973), while linear models (Logistic Regression and SVC) showed substantially lower performance with F1-scores of 0.888 and 0.859 respectively, indicating the non-linear nature of the decision boundary.

TABLE II
Admission Decision Prediction Performance

Model	Train Accuracy	Test Accuracy	Test F1	Test AUC
XGBoost	1.000	0.9975	0.998	0.997
CatBoost	0.999	0.9925	0.994	0.993
Gradient Boosting	1.000	0.9925	0.994	0.991
Random Forest	0.998	0.965	0.973	0.955
Logistic Regression	0.858	0.855	0.888	0.837
SVC	0.838	0.810	0.859	0.772

For scholarship tier prediction (Table III), XGBoost again demonstrated superior performance with test accuracy of 0.985, F1-score of 0.985, and AUC of 0.997. Gradient Boosting achieved a test F1-score of 0.978, while CatBoost obtained 0.972. Random Forest maintained reasonable performance (F1-score of 0.953), whereas linear models again showed limited effectiveness with F1-scores below 0.84.

The minimal gap between training and test performance across top-performing models (XGBoost, Gradient Boosting, CatBoost) indicates effective generalization without significant overfitting. The consistently high AUC scores ($>$ 0.99) across ensemble methods demonstrate robust discriminative capability across classification thresholds.

TABLE III
SCHOLARSHIP TIER PREDICTION PERFORMANCE

Model	Train Accuracy	Test Accuracy	Test F1	Test AUC
XGBoost	1.000	0.985	0.985	0.997
Gradient Boosting	1.000	0.978	0.978	0.996
CatBoost	0.996	0.973	0.972	0.997
Random Forest	0.997	0.955	0.953	0.996
Logistic Regression	0.881	0.855	0.840	0.919
SVC	0.848	0.825	0.763	0.925

B. Model Deployment and Production Performance

Based on the evaluation results, we selected XGBoost Classifier as the primary model with Gradient Boosting as a backup model. A training pipeline was implemented incorporating data preprocessing, feature engineering, model training, and evaluation stages. The pipeline includes an automated model update mechanism: the production model is replaced only when a newly trained model achieves greater than 7% improvement in average F1-score across both tasks. This threshold balances the computational cost of model retraining against the benefit of incremental performance improvements.

The deployed model configuration achieved an average F1-score of 0.978 across both prediction tasks. Table IV presents detailed performance metrics for the deployed model on the test set.

TABLE IV
DEPLOYED MODEL PERFORMANCE METRICS

Target	F1 Score	Precision	Recall
Admission Decision	0.985	0.985	0.985
Scholarship Tier	0.973	0.973	0.973

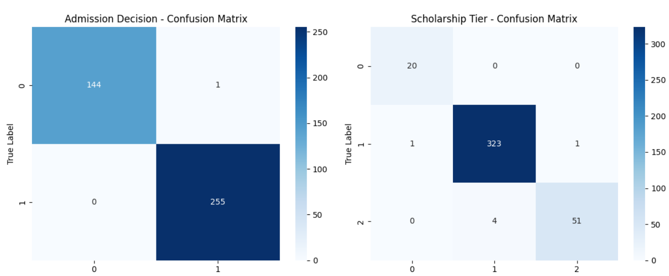

Fig. 3. Confusion matrices for admission decision and scholarship tier prediction tasks. Near-perfect classification for admission, high accuracy for scholarship tier with minor confusion between adjacent tiers.

C. Feature Importance Analysis

To understand which applicant attributes most strongly influence admission and scholarship decisions, we conducted feature importance analysis using the deployed XGBoost model. Figure 4 presents the top 10 features ranked by their contribution to model predictions.

The weighted composite score consistently contributes most strongly to both admission and scholarship decisions, validating our feature engineering approach. This finding aligns

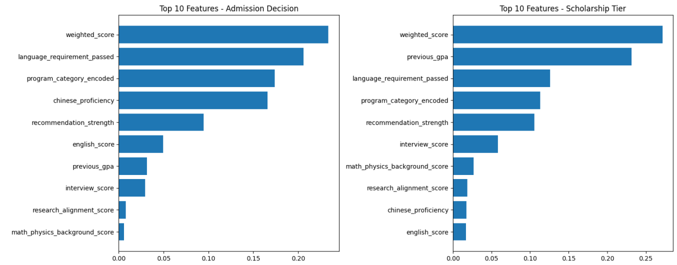

Fig. 4. Top 10 features contributing to model prediction. The weighted composite score contributes most strongly, followed by language requirement flag, previous GPA, interview score, recommendation strength, and research alignment score. Categorical features such as program category and degree language show moderate importance. Country of origin shows relatively low importance.

with domain expectations, as the weighted score integrates multiple relevant academic indicators into a single comprehensive metric. The language requirement flag ranks as the second most important feature, reflecting the critical role of language proficiency in program admission criteria. Previous GPA ranks third, confirming its fundamental importance in academic evaluation.

Other notable features include interview score, recommendation strength, and research alignment score, each contributing meaningfully to predictions. Categorical features such as program category and degree language show moderate importance, suggesting that admission criteria exhibit some program-specific variation. Country of origin shows relatively low importance, indicating that the model primarily focuses on academic merit rather than geographical factors, which supports the fairness objectives established in this study.

IV. DISCUSSION

A. Summary of Findings

This work introduces a machine learning-based framework for analyzing and automating international student admission and scholarship decisions in Chinese universities. The results demonstrate that machine learning models can provide accurate, consistent, and data-driven predictions for both admission outcomes and scholarship allocation. Our multi-stage pipeline encompasses synthetic data validation, preprocessing and feature engineering, comparative evaluation of six classical machine learning algorithms, and deployment of our prediction system.

The deployed XGBoost model achieves F1-scores exceeding 0.97 for both admission and scholarship prediction tasks, with consistent performance across training and test datasets indicating strong generalization capability. These findings support our hypothesis that machine learning can substantially improve the efficiency, consistency, and fairness of international student evaluation processes.

B. Implications for Practice

The findings highlight several practical implications for international student recruitment in Chinese universities. First, the proposed ML-based framework directly addresses critical challenges inherent in manual admission processing: it reduces time requirements, minimizes resource consumption, and mitigates evaluator bias, thereby improving fairness and consistency. The framework enables universities to process large volumes of applications systematically while maintaining high decision quality.

Second, the feature importance analysis provides actionable insights into which applicant characteristics most strongly predict success. The prominence of the weighted composite score, language proficiency, and prior academic performance suggests that holistic evaluation incorporating multiple dimensions yields more reliable predictions than single-metric assessments. This finding supports the value of comprehensive application review processes while demonstrating that such processes can be automated effectively.

C. Limitations

Although our trained models show promising outcomes, certain limitations should be acknowledged. First, the analysis relies entirely on synthetic data generated based on Beijing Institute of Technology's application form structure. Although we rigorously validated feature distributions and incorporated domain knowledge, the synthetic data may not fully capture the complexity and nuance of real application data. The actual performance on genuine applications remains to be validated.

Second, the current framework implements a static model that does not update or improve based on predictions and their outcomes. This design limits the model's ability to adapt to evolving applicant pools, changing program requirements, or shifting institutional priorities over time. A static model may gradually lose accuracy as the underlying data distribution shifts.

Third, the dataset size (2,000 samples) is relatively modest for deep learning approaches, though adequate for the classical ML methods employed. Larger datasets would likely further reduce variance, enhance stability, and improve generalization to unseen applicant profiles.

Fourth, the model focuses exclusively on structured numerical and categorical features, excluding potentially valuable unstructured information sources such as personal statements, recommendation letter text, research proposals, and interview transcripts. These qualitative materials often provide crucial insights into applicant potential that structured features cannot fully capture.

D. Future Work

Future research should address these limitations through several directions. First, validation with anonymized real-world application data is essential to assess the framework's practical effectiveness and identify potential gaps between synthetic and actual data distributions. Collaboration with university admissions offices to obtain historical application data (with appropriate privacy protections) would enable robust empirical validation.

Second, expanding the dataset to include multiple years of applications across different programs and institutions would enhance generalizability and enable the model to learn stable, transferable patterns while accounting for temporal and institutional variation.

Finally, investigating fairness metrics and bias detection mechanisms specifically tailored to international student admissions would ensure the framework promotes equitable outcomes across demographic groups. Systematic fairness auditing should examine potential disparities across countries of origin, gender, socioeconomic background, and other protected attributes.

V. Conclusion

This paper presents BIT-ADMIT, a machine learning framework for automating admission and scholarship decisions for international applicants at Chinese universities. Using 2,000 synthetic student records modeled on Beijing Institute of Technology's application form, multiple classifiers were developed and evaluated. XGBoost and Gradient Boosting achieved F1-scores above 0.97 for both admission and scholarship predictions. The framework is deployed as a FastAPI web service enabling real-time inference.

Results show ML can effectively address the limitations of manual admission processes, delivering accurate, consistent, and data-driven decisions while reducing time, resource use, and evaluator bias. Feature importance analysis highlights composite academic metrics, language proficiency, and prior performance as key predictors, supporting a holistic evaluation approach.

Though trained on synthetic data, BIT-ADMIT provides a scalable proof-of-concept pipeline. Future work will validate with real applications, implement continual learning, and integrate large language models for unstructured materials. By complementing initiatives like the CSCA, this framework advances fair, transparent, and efficient international student admission.

References

[1] A. Priyadarshini, B. Martinez-Neda, and S. Gago-Masague, "Admission Prediction in Undergraduate Applications: an Interpretable Deep Learning Approach," in *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*, IEEE, Sep. 2023, pp. 135–140. [Online]. Available: http://dx.doi.org/10.1109/TransAI60598.2023.00040

[2] T. Liu, C. Schenk, S. Braun, and A. Frey, "A Machine-Learning-Based Approach to Informing Student Admission Decisions," *Behavioral Sciences (Basel)*, vol. 15, no. 3, p. 330, Mar. 2025. doi: 10.3390/bs15030330. [Online]. Available: https://doi.org/10.3390/bs15030330

[3] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, 2022. [Online]. Available: https://doi.org/10.1186/s40561-022-00192-z

[4] G. Raftopoulos, G. Davrazos, and S. Kotsiantis, "Fair and Transparent Student Admission Prediction Using Machine Learning Models," *Algorithms*, vol. 17, no. 12, p. 572, 2024. [Online]. Available: https://www.mdpi.com/1999-4893/17/12/572

[5] China Ministry of Education, "Statistics on International Students in China," 2021. [Online]. Available: http://en.moe.gov.cn/documents/statistics/2021/national/202301/t20230104_1038055.html

[6] China Ministry of Education, "National Education Statistics," 2021. [Online]. Available: http://en.moe.gov.cn/documents/statistics/2021/national/202301/t20230104_1038055.html
[7] B. Wächter and F. Maiworm, Eds., *English-Taught Programmes in European Higher Education: The State of Play in 2014*. Bonn, Germany: Lemmens Medien GmbH, 2014.
[8] R. Huang, "The improvement of the selection mechanism of international students coming to China: Take countries along the Belt and Road as an example," *Transactions on Social Science, Education and Humanities Research*, vol. 5, pp. 508–515, 2024. [Online]. Available: https://doi.org/10.62051/d5ppq352
[9] Y. Chen, M. Jiang, and O. Kesten, "An empirical evaluation of Chinese college admissions reforms through a natural experiment," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 117, no. 50, pp. 31696–31705, 2020. [Online]. Available: https://doi.org/10.1073/pnas.2009282117

APPENDIX

The complete data generation, preprocessing, and model training code is available at: https://github.com/yeabwang/bit-admit

The synthetic dataset used in this study is available in the GitHub repository for reproducibility purposes.