# Enhancing Academic Grade Prediction Through Bat Algorithm- Driven Ensemble Learning Technique

Anil Purushothaman[1], Varsha Jotwani[2]
[1,2] *Department of Computer Science & Information Technology, Rabindra Nath Tagore University, Bhopal*

***Abstract*—In this paper, we present an ensemble framework utilizing the Bat Algorithm for predicting the student academic performance in two parts of the course, that is at 40% and 80% course completion. The model achieves an optimal feature selection using the Bat Algorithm and the bagging ensemble integrates four distinct classification algorithms: K-Nearest Neighbors, Support Vector Machine, XGBoost, and Long Short-Term Memory LSTM. We evaluate the addition of PCA to this ensemble, considering the trade-off between dimensionality reduction and classification performance.**

***Index Terms*—Academic performance prediction, Bat Algorithm, ensemble learning, multi-class classification, feature selection, higher education.**

## I. INTRODUCTION

Growing digitization of academic records and students' learning footprints have entirely redefined the way institutions consider academic planning, optimization, and support. Predictive analytics in education, including the ones aimed at predicting the academic outcome of students, has great potential in supporting proactive data-driven decisions. Education predictive models are actively utilized by academic institutions for identifying early alarms signs of academic underperformance, addressing them timely, and thus, increasing the number of students who graduate. Although binary classification models to predict the risk of dropping out have been thoroughly investigated, multiclass prediction models that produce accurate final academic performance grades are a challenging and underperforming problem. It is tough because various academic and demographic issues play a significant role in student performance dynamics at different stages of their academic processes [1].

### *A. Background and motivation*

With the increasing amount of educational data and recent advances in machine learning techniques and data analytics, predicting students' final performance has become a trend among educators, researchers, and academic institutions. Building intelligent systems that can identify students at risk of failing a course can provide timely measures such as counselling, additional instruction, or policy changes, leading to improved student success and institutional effectiveness. However, the problem is much more complex. Students' final success is not based on the scores of a single exam or evaluation but rather on various assessments throughout the course [2]. Importantly, long before the final grade is known, the interim exam's scores expressed as percentages of the course grade, which quantizes a student's academic trajectory, which will likely influence the final out of the course based on the student's past performance. Thus, not only is predicting the final course label based on the early-stage exam data a technical obstacle, but it is also an educational necessity. Nonetheless, several challenges persist:

- the multidimensional nature of student data – demographic, academic, behavioural, and engagement features.
- the unique temporal structures of assessment data and non- linear relationships with the final grade.
- the severe class imbalance in the grade label – some grades (eg. Fail and Third Division) are underrepresented in the dataset, yielding biased predictions [3].
- Some features are not relevant to the grading labels or even weakly correlated – which worsens the generalization of the machine learning models, and
- the lack of clarity of the predictive algorithms – black-box nature of deep learning models.

In this context, ensemble models along with heuristic meta- learning feature selection methods have shown

promise [4]. This paper continues this direction by utilizing the Bat Algorithm [5] for customized optimization of feature selection, focused solely on preserving the assessment features of the test scores from major checkpoints in the course. By integrating this with a bagging ensemble of SVM, KNN, and LSTM models, this paper aims to increase the accuracy and reliability of predicting academic grades.

*B. Research gaps and challenges*

Despite the extensive research in educational data mining and decent achievement in student performance prediction, current models are overwhelmed by a strong assumption of binary outcomes and hence do not scale to more comprehensive multicategory grading systems as common in higher education scenarios [6]. Prior works typically target predicting a binary outcome such as whether a student passes the final exam or eventually drops out. This turns the problem of tracking a student's lifetime educational progresses into an overly simplistic binary classification task. Furthermore, the existing prediction models either ignore early-life assessments or do not correctly harness

these fanciful measurements to estimate the final student performance.

Besides, the feature selection process is another critical bottleneck to building a more potent performance model: current approaches based on tradition require fine-grained tuning and lacks focus on the relationship between features. Even modern methods such as Metaheuristic, including Genetic Algorithms, or Grey Wolf Optimization for feature selection, comes with limitations when having local optimum trap or premature convergence [7].

Moreover, few models provide effective predictive insights at multiple stages of a student's academic journey [8]. Assessment metrics at 40% and 80% course completion stages, while valuable for real-time intervention, are seldom retained and emphasized during feature optimization. This leads to models that either underperform or lack interpretability when applied for proactive academic planning.

In addition, very few models possess the capability to generate effective prediction throughout a student's academic adventure [9]. Assessment metrics created at 40% and 80% completion stages of a course may be drastically pointful in real-time invasion but are seldom held for by feature optimization; the models thus suffer from both inefficiency and lack of interpretability at the adoption stage for ahead-of-time academic planning. A complete predictive structure should hence have the capability to accurately predict multi-channel grade distributions, retain course- specific assessment-based features, build strong and adaptive feature optimization, and generalize across batches of learners and subject categories in higher education.

*C. Objectives and Contributions*

This study advances an ensemble-based predictive model that combines a Bat Algorithm-based feature selector with a bagged ensemble of the SVM, K-NN, and LSTM classifiers. The model is a novel prediction mechanism developed for the Multiclass prediction of final academic grades based on early assessment data when combining with other student attributes. The contributions of this research are:

- Developing a Bat Algorithm-based feature selector, this algorithm has been significantly altered by giving priority to asg1 and int1 features and optimizing the remaining feature space to minimize redundancy and maximize predictive impact [10].
- Creating a Hybrid Ensemble Model that merges SVM, which utilizes decision boundaries, K-NN, which uses local similarity, and LSTM, which uses temporal learning, is wrapped in a bagging structure for strong generalization.
- Comprehensive Assessment of two academic stages is provided: this study examines students at 40% and 80% completion of their course, allowing the relative strength of assessment data at various periods of the learning method to be analyzed.
- Applying the model on a true, multi-batch higher-education dataset: By using real-world data, the model confirms both applicability and scalability while also justifying correctness in the presence of true academic information.

## II. RELATED WORK

In recent years, there has been an array of studies that have focused on the prediction of academic performance. The trend has been fuelled by the increased accessibility of educational data and the development of machine learning techniques. Two of

the most popular ones, ensemble learning models, and bio-inspired feature optimization algorithms aim at enhancing the accuracy of predictions, decreasing the variance of the model, and addressing complex and non-linear relationships.

*A. Ensemble Learning in Academic Prediction*

Further, ensemble models are group models, bagging, boosting or stacking that combine many base classifiers to improve upon robustness and generalization. Numerous academia have confirmed the competitiveness of ensemble models with a simple classifier in predicting the grade's future value, dropout risk possibility or children's success possibility or failure.

For example, a study proposed a model that includes GBoost, AdaBoost, and SVM for the education grant outcomes prediction [10]. They built the model using high accuracy in multiclass prediction. Another example is a multi-split bagging ensemble model presented specially optimized for education data mining [11]. Currently, the stacking strategy has also been actively used concerning the previous model, and the positive outcomes are among academia have been received [12][13]. The same variant of a hybrid model was used in one study [14]. These authors applied stacking based on the well-known methods Random Forest and XGBoost for at-risk students' detection.

The same voting-based ensembles and classifier fusion have been mentioned. The dissimilar fusion approach combining MultiBoost and Multilayer Perceptron has shown high precision results according to another study [15]. Some of the identifiable studies have demonstrated the possibility of using simpler ensembles for large dataset and, inversely, more complex ensembles for a small dataset [16][17].

*B. Feature Selection in Academic Data Mining*

Feature selection is a key aspect of predictive modeling that reduces dimensionality by filtering irrelevant or redundant inputs. Filter methods such as Chi-square and information gain are commonly used [18][19], but they fail to capture the interaction between different features.

Wrapper and embedded methods like RFE and LASSO regression provide a more accurate output, but they are computationally time-consuming in the case of large educational datasets [20][21]. Several scholars have applied RFI and Boruta to academic forecasting tasks [22].

However, given the constraints of conventional methods, there is a growing body of research that employs metaheuristic systems for high-dimensional educational datasets.

*C. Bio-Inspired Metaheuristic Optimization Techniques* Feature subset optimization using metaheuristic algorithms based on natural processes has shown considerable potential. The former feature, Grey Wolf Optimization [23][24], has also been utilized in assessing student performance datasets, with appropriate efficacy. A study

[25] compared various approaches available for selection of attribute acquisition, which revealed that evolutionary ones outperformed customary methods.

However, algorithms like GWO and Particle Swarm Optimization are susceptible to either premature convergence or failure to optimize features of local minima. To treat for the phenomenon, [26] suggested a chaotic- diffusion-enhanced GWO for feature selection, while [27] implemented a greedier variant that inevitably led to failure to abide by a strict hierarchy rule.

Nevertheless, research demonstrate that the regular limitation of these studies concerns the neglect of manual constrictions, such as the preservation of formative evaluation metrics.

*D. Bat Algorithm in Optimization and Education*

Arguably one of the most promising metaheuristics algorithms to leverage this dynamism is the Bat Algorithm. Although the algorithm was later inspired the echolocation behavior of the microbats, the dynamic balance between exploration and exploitation has shown robust results in various areas such as engineering optimization [5], medical diagnostics [28], and power system tuning [29]. Nonetheless, there have not been many studies that have explored the applicability of BA in the educational field.

The limited number of educational works primarily used BA for developing examination timetables or optimizing outer adjusts in e-learning areas [30][31]. There have not been attempts to use the BA for academic performance prediction, likely due to the lack of study regarding the feature optimization capability and tendency to ignore the domain-specific features such as mid-course assessments in conventional BA implementation.

This paper attempts to bridge this gap by proposing a modified Bat Algorithm in a feature-level for prediction. It is then coupled with an ensemble model, enabling significantly better predictive power than existing multiclass grading prediction using GWO-based approach [24].

## III. METHODOLOGY

The framework proposed for the prediction of academic performance is depicted as a maximum fusion model that embraces the power of machine learning in all spheres of data science techniques combined with a meta-heuristics- based feature selection approach. The proposed model is a multi-stage pipeline, starting from preprocessing of data, employing the novel Bat Algorithm customized for feature selection, and includes the ensemble of classifiers using the bagging algorithm, specifically Support Vector Machines, k- Nearest Neighbours, and Long Short-Term Memory networks. Each stage of the proposed academic pipeline is in coherence with the considerations of multiclassification difficulties, the nature of the high-dimensional student data, and the requirement of phase-aware prediction due to metric readings availability at 40% and 80% course completion. The process guarantees that important academic feature indicators are preserved while the irrelevant or the repeated ones are removed to improve the model's precision, generalization, and interpretation. This chapter presents the details of each step of the proposed framework, including the procedure of encoding and normalization of the data, and the feature selection and the assessment of the training and testing approaches.

### *A. Data Collection and Preprocessing*

The dataset for this study was obtained from the Institute of Professional Education and Research in India. The database covers the academic entries for 2017 to 2021 batches and was conserved from an earlier preliminary phase of this research [24].

Table 1: Summary of Raw and Processed Dataset

| Description | Value |
|---|---|
| Total rows in raw dataset | 10,162 |
| subject-wise entries removed (Spcl not taken) | 1,636 |
| subject-wise entries removed (dropouts) | 760 |
| Missing/NA entries removed | 23 |
| Valid subject-level rows retained | 7,723 |
| Approx. unique students | 351 |
| Total features (before selection) | 26 |
| Target classes | A, B, C, D |

This dataset includes details of 462 students of an Undergraduate college in India. Summary of the data is shown in Table 1. As much as any student read around 18 of the 22 subjects possible during a 3-year college degree, the raw data was structured at the subject level, resulting in 10162 lines overall where each row serves as the entry of a student's review of the subject.

Table 2: Features in Dataset

| Feature Name | Type | Description |
|---|---|---|
| Gender | Categorical | Student gender (encoded as 0/1) |
| Board | Categorical | Board of secondary education |
| Subject Code | Categorical | Encoded subject identifier |
| City | Categorical | City of residence |
| asg1 | Numeric | Assessment at 40% course progress |
| int1 | Numeric | Assessment at 80% course progress |
| Attendance | Categorical | Used for filtering (valid, dropout, etc.) |
| Final Grade | Categorical | Target label (A, B, C, D); A – First Division, B – Second Division, C – Third Division, D - Fail |

For this purpose, many preprocessing steps were requisite to design the data for modeling. First, irregular records were detected: all records denoted as "dropout" and "Spcl not taken" in the Attendance column were filtered out. The previous records were dismissed, recording 760 dropouts, 17 with various patterns not identified for scholars, 1636 were discarded, and the records labeled "NA" in some critical lines. The result of these filtering requirements was a dataset of 7723 valid entries, equivalent to approximately 351 students. Table 2

summarizes the key features in the dataset used in this study.

*B. Feature Construction and Encoding*

Notably, demographic and academic features, including gender and board of education and subject codes and city of residence, were also encoded via labels to numeric form. Accordingly, mid-semester evaluations, including first assignment asg1 and internal test int1, which corresponds to 40% and 80% of the course, respectively, were retained as key stage indicators. Crucially, the final grade, the target, was then label-converted from letter grades – {A, B, C, D} – to numeric class labels, enabling multi-class classification. Table 3 presents the grade-wise summary of records in the data. Stages-wise Dataset Creation: As previously mentioned, two separate training sets were established to evaluate performance at the following academic progress stages:

Table 3: Class Distribution by Grade

| Grade | Encoding | Number of Records |
|---|---|---|
| A - First Div | 1 | 5997 |
| B - Second Div | 2 | 982 |
| C - Third Div | 3 | 298 |
| D - Fail | 4 | 445 |

The training set at 40% stage: consisted of the features available post the first internal assessment – asg1 – thus implying early-stage predictors.

The 80% stage training set, including asg1 and int1, serve to reflect mid-to-late academic progress before the final exam. Table 4 presents the summary of the two-stage assessment and the features available.

Following their creation, each training set was independently normalized via min-max scaling, converting the feature values to the (0, 1) array. This technique guarantees a consistent contribution factor to the model during the training phase.

As evident from Table 3, the file is seriously minority- imbalanced, with the proportion of First Division labels representing a tremendous majority. This imbalance may skew learning algorithms towards predicting the majority class, limiting the model's ability to generalize across different target labels. Therefore to address this, an appropriate training phase balancing was executed. For each split within the ensembles structure incarnation, the identical number of records was randomly sampled for each class based on the smallest class. This strategy ensured that the model trained uniformly on all target categories, providing a comprehensive multi-class performance evaluation.

*C. Feature Selection Using Bat Algorithm*

Although the total feature set contained 26 academic and demographic features, these features were not equally important for the prediction accuracy. Consequently, to filter out irrelevant features and features that would not contribute to the model's accuracy, the Bat Algorithm was employed as an intelligent feature selection step within the pipeline.

This enabled dimensionality reduction and left only relevant or non-redundant features required for critical assessment.

Feature selection is a vital component that significantly contributes to the quality of the classifier, its complexity, and computational efficiency.

Table 4: Dataset Versions for Model Training

| Dataset Version | Assessment Stage | Features Included | Target Variable | Records Used |
|---|---|---|---|---|
| Dataset-40 | 40% Completion | Demographics + asg1 | Final Grade | ~7,723 |
| Dataset-80 | 80% Completion | Demographics + asg1 + int1 | Final Grade | ~7,723 |

As a result, the input datasets for 40% and 80% models were preprocessed rigorously enough and resulted in well-

structured, valid datasets for high-performance training and evaluation.

More specifically, the number of features for high-dimensional educational data must be reduced, and feasible characteristics should be selected to develop and evaluate the Prediction model. The Bat Algorithm was used as a metaheuristic aiming to choose suitable data attributes both for 40% stage and at 80% course completion stage.

The Bat Algorithm was based on microbats' echolocation system to optimize successful foraging, enabling estimation of prey distance and direction through echolocations. In optimization terms, each

individual bat is a potential feature subset and the main goal is to find a subset that maximizes the classification performance while minimizing the feature count.

Each bat is represented as a binary string in the population, 1 if the feature is selected and 0 otherwise. First, generate a random population to have diversity in the initial solution. Parameters involved in the algorithm are determined according to preliminary tests are population size, frequency range, loudness, and pulse rate of emission that parametrize the trade-off the relationship between exploration and exploitation.

The fitness is calculated according to a weighted sum of the classification accuracy and the number of selected features. The fitness function is maximized, and the selection number of features is negatively handled to avoid an increased number of features, which likely cause overfitting. Formally:

$$Fitness = \propto \times Accuracy - \beta \times (No.\ of\ Selected\ features)/(Total\ Features) \quad (1)$$

where α and β are empirically determined weighting factors to maintain a trade-off between predictive performance and feature subset size.

In each iteration, the bats update their velocities and positions using frequency-tuned motion equations:

$$f_i = f_{min} + (f_{max} - f_{min}) \times rand() \quad (2)$$

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x_{best}) \times f_i \quad (3)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (4)$$

where $x_{best}$ is the position of the current best solution. Loudness and pulse emission rate are updated dynamically to fine-tune local exploitation as the algorithm progresses [5].

Finally, a local search is employed in the surroundings of the best solutions through random walks if the random number that has been produced is less than the bat's pulse emission rate. If the solutions improve the fitness function, the solutions are accepted with a small probability close to 0 or high when selecting a solution, resulting in rarely getting out of the local optima.

Stopping consists of the number of iterations and output. The algorithm is terminated if the maximum number of iterations is surpassed, or the convergence of the fitness values is seen. The output of the method is the optimal feature subset in each stage, i.e., 40% and 80% of the datasets, respectively, which is transferred to the ensemble classification phase.

*D. Rationale for Using Bat Algorithm*

Bat Algorithm is more flexible than the Grey Wolf Optimizer that we used in our previous work [24] for it offers a combination of global and local search and a flexible exploration of the search space determined by the parameters of loudness and pulse rates which can be adaptively changed. The latter guarantees that BA can be tuned to datasets with highly diverse patterns of relevance for the features, such as multi-stage student performance. Additionally, the movement controlled by frequency allows for more adjustment in the feature subset refinement process.

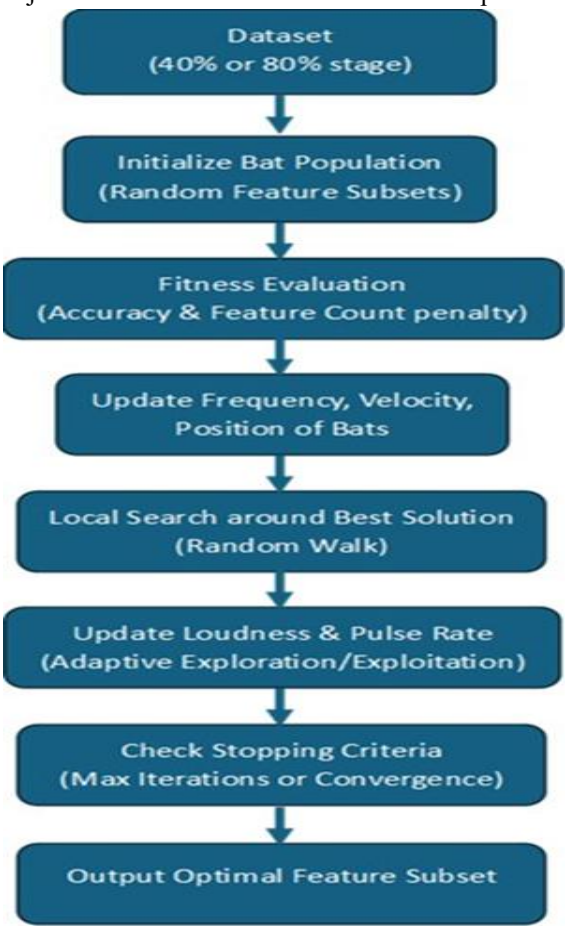



Figure 1: Application of Bat Algorithm on dataset

*E. Ensemble Classification Framework*

After the process of feature selection using the Bat Algorithm, the reduced feature subsets for both stages of 40% and 80% academic progress are utilized as an input to an ensemble classification framework to combine the merits of multiple base learners and eliminate their shortcomings using

majority voting aggregation.

This ensemble, by design, is a bag-of-learners built upon the concept of meta-learning; namely, multiple classifiers are trained on bootstrapped samples of the feature subset. Specifically, the proposed model consists of the following base learners: Support Vector Machine, given its effectiveness in high-dimensional feature spaces, and strong generalization ability; K-Nearest Neighbors, owing to its simplicity and ability to identify local patterns; Long Short-Term Memory Network, if there are sequential or dependent patterns in a student's performances with each assessment; simple Decision Tree model for interpretability and the possibility to family with non-linear relationships; Extreme Gradient Boosting, a high-performing model on tabular data that can anticipate intricate patterns. Each of these multi-class classifiers is implemented as it was during a preliminary evaluation, receiving balanced bias-variance trade-offs.

As shown in figure 2, the ensemble training process involves the following steps:

Bootstrapping: The BA-chosen feature dataset is sample drawn multiple times at random with replacement.

Classifier training: The base learners are independently trained on these bootstrapped datasets.

Probability estimation: Each trained classifier outputs predicted class probabilities for the test instances.

Majority voting: The final predicted class is obtained through simple majority voting across all the base learners.

*F. PCA-Based Variant*

Apart from the BA-selected feature ensemble, an alternative PCA-based ensemble was defined as the comparison branch. The PCA transformation is applied to the BA-selected features where new uncorrelated learning features are produced. This may ultimately reduce redundancy and noise in the data, which might increase classifier performance on data with an elevated noise level or lower the number of samples. The PCA variant utilizes the same training and voting methodology as the standard ensemble.

Firstly, the evaluation protocol followed before assessing the baseline and PCA-based ensemble's respective variant involved 10-fold based cross-validation on the training phase for further independent testing on the test phase, and factors like accuracy, precision, and recall F-1 score were

considered for both variant's comparison.

*G. Advantages of the Proposed Ensemble*

The proposed ensemble version has several advantages. The primary one is the increased likelihood of accolade varied decision boundaries due to the use of heterogeneous classifiers. Bagging also trades many biases for reducing the variance. Furthermore, by performing feature selection with Bat Algorithm, each base learner works with a feature set tailored primarily for predictive relevance for the learner, thus increasing the accuracy and efficiency.

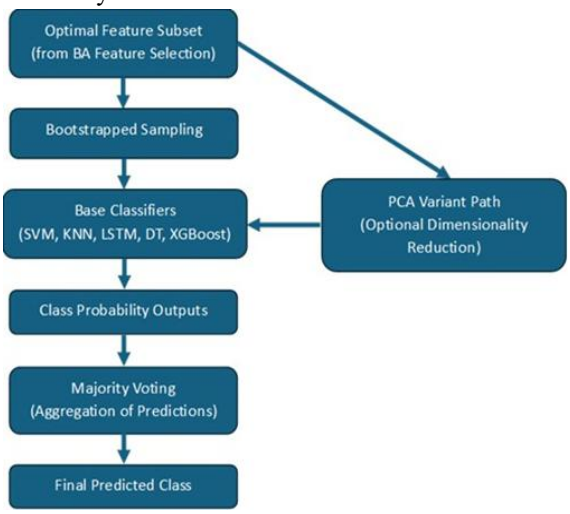



Figure 2: Ensemble Training Process for Grade class prediction

## IV. RESULTS AND DISCUSSION

This section presents the experimental results of the one proposed in the study, namely Bat Algorithm - based ensemble classification model, used to predict the final grades of students at two stages of academic progress - 40% and 80% course completion. The experimental outputs are analyzed considering multiple classification performance indicators, including the comparisons with the PCA- enhanced alternative of the BA-based ensemble.

The experiments were executed via MATLAB R2016a on the system operating on the Intel Core i7-12700H processor

2.3Ghz base, 4.7 GHz turbo, 14 cores, 16 GB DDR5 RAM,

512 GB PCIe NVMe SSD, and Windows 11 OS. The hardware was equipped with a NVIDIA GeForce RTX 3050 GPU.

*A. Performance Metrics*

The multiple standard classification metrics leveraged in evaluation for a comprehensive performance measure are: Accuracy (ACC) – Proportion of accurately classified instances; Precision (P) – Proportion of accurately predicted positive instances out of all predicted positives; Recall (R) –

Proportion of accurately predicted positive instances out of actual positives; F1-score (F1) – Harmonic mean of precision and recall; balance between the two above; Execution Time (T) – Aggregate time for continuous model training and testing.

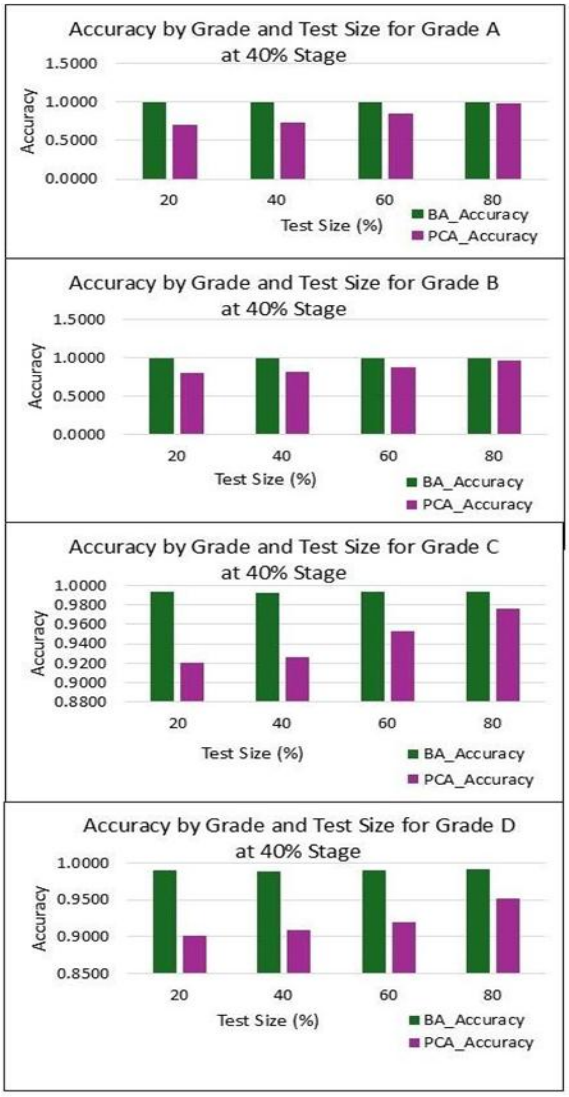



Figure 3: Accuracy by Grade at 40% Completion Stage

The calculation included: Per grade category (A, B, C, D); Across multiple test sizes (20%, 40%, 60%, 80% of dataset size); For both BA and BA+PCA variants.

The analysis is supplemented with visual plots and tabular summaries to accentuate the trends better and in detail.

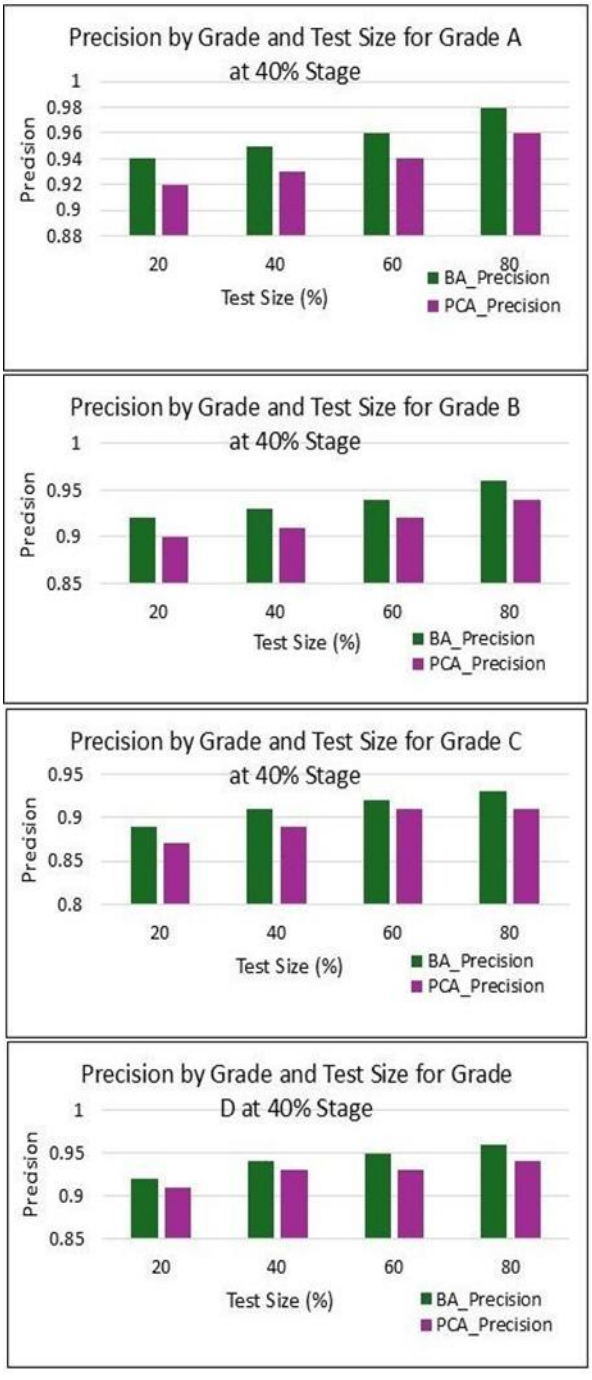



Figure 4: Precision by Grade at 40% Completion Stage

*B. Grade-wise Results at 40% Stage*

At the 40% stage, the objective for the model is to make an early prediction about student grades. The results of the BA- based ensemble and its PCA variant are summarized in Table 5.

From the results, the BA-based ensemble was consistently high for all grades. Specifically, Grade A and B. In the PCA variant, there were marginal F1-score drops in Grade D had improved execution time. Radar plots of Figure 7 confirmed better balance in BA across metrics in minority classes.

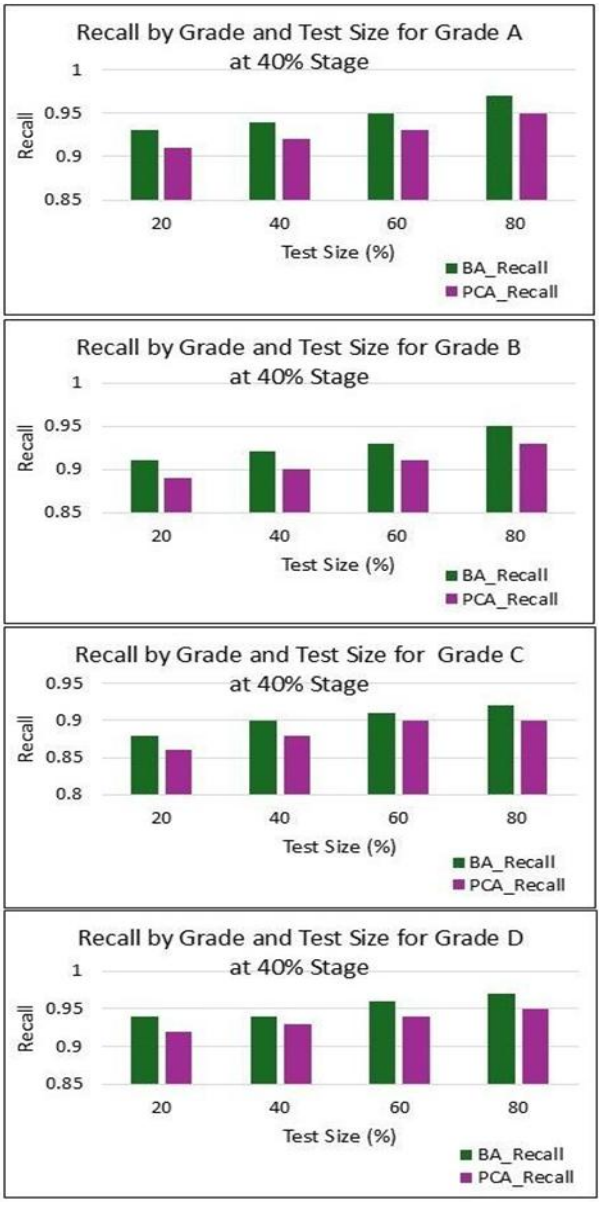



Figure 5: Recall by Grade at 40% Completion Stage

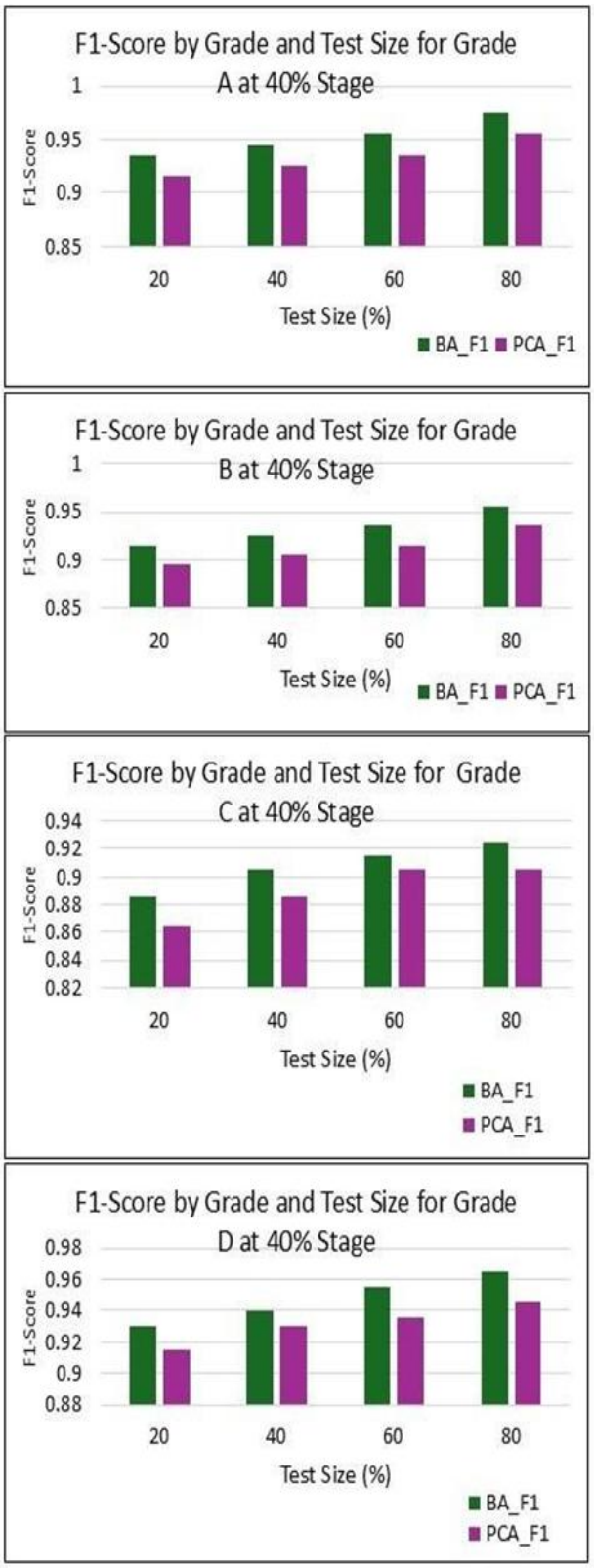



Figure 6: F1-Score by Grade at 40% Completion Stage

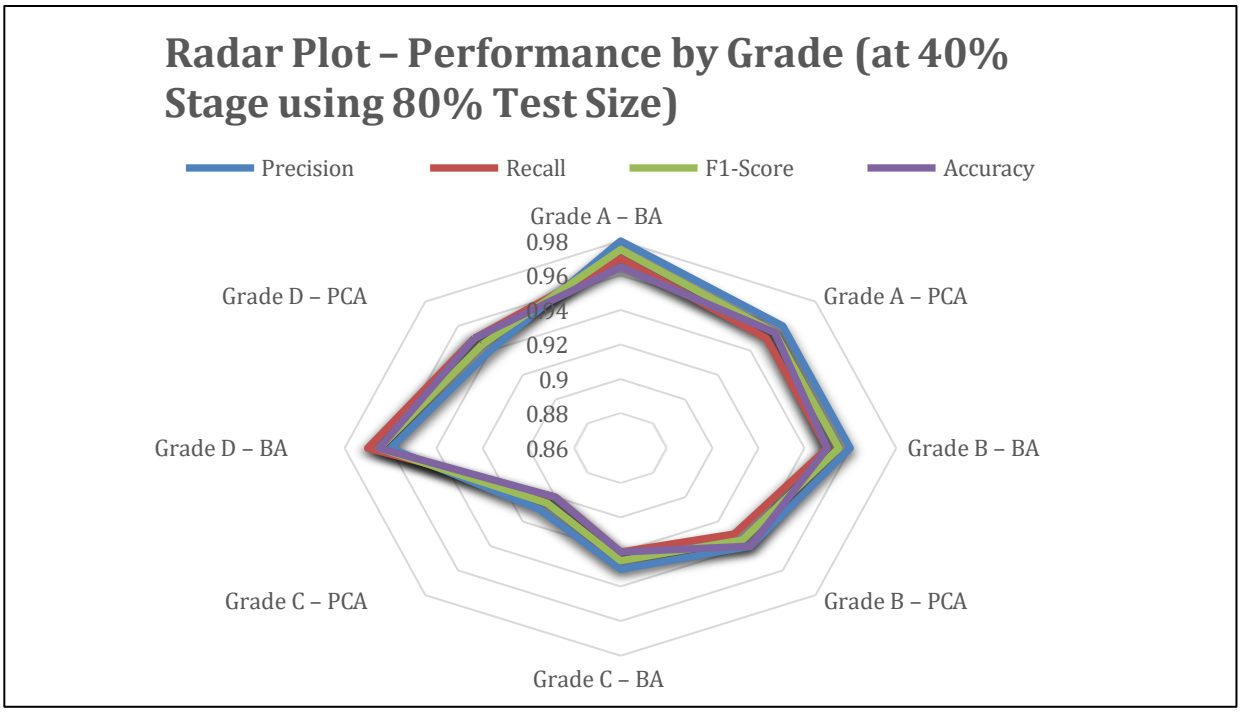



Figure 7: Radar Plot - Performance by Grade at 40% Completion Stage

*C. Grade-wise Results at 80% Stage*

At the 80% stage, the obtainable data per student is more and should potentially deliver better predictive results. Table 6 presents the comparative results.

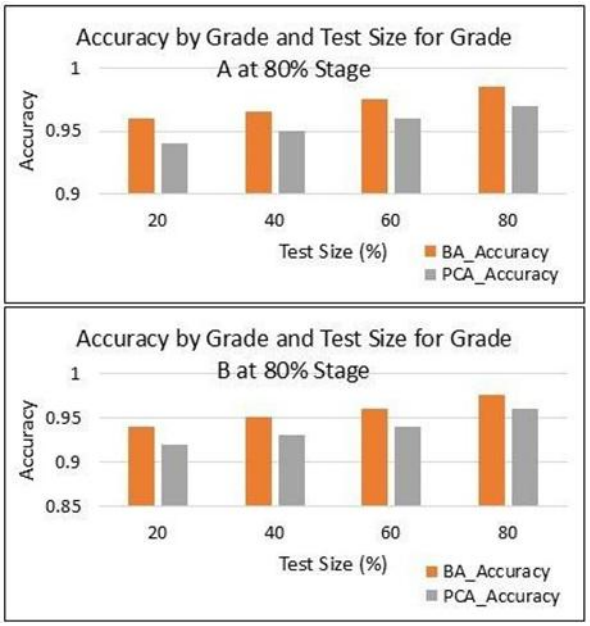



Figure 8: (a) Accuracy of Grade A & B at 80% Completion Stage

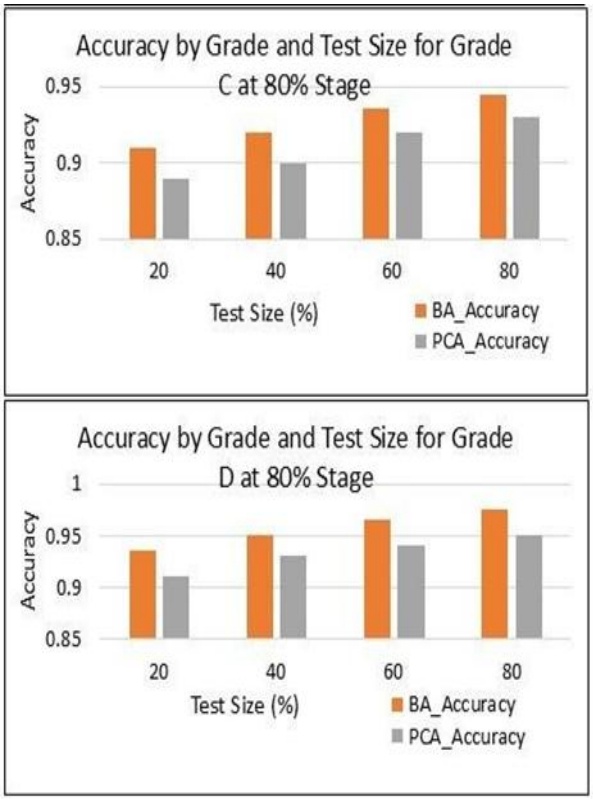
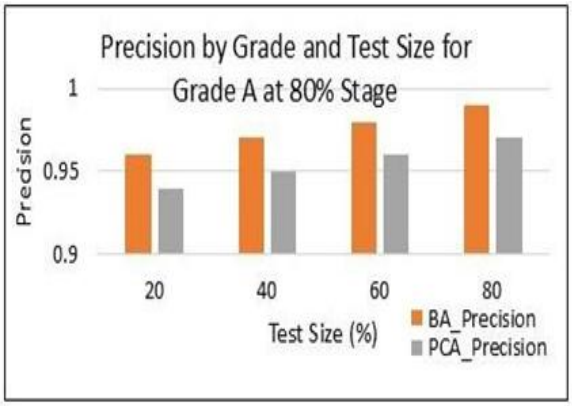
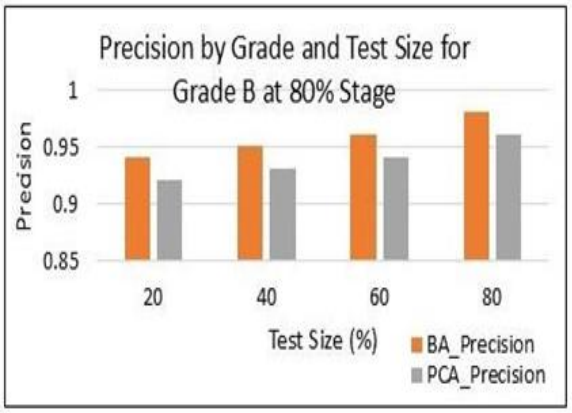
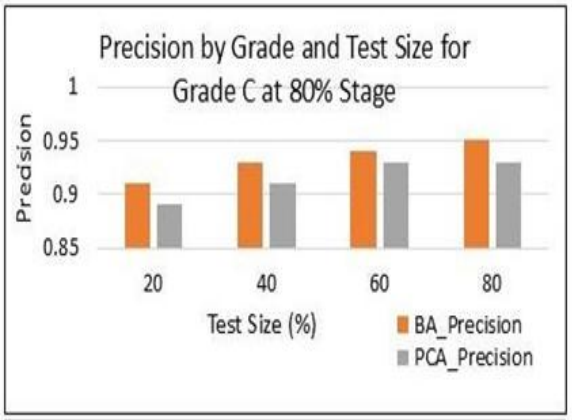



Figure 9: (b) Accuracy of Grade C & D at 80% Completion Stage



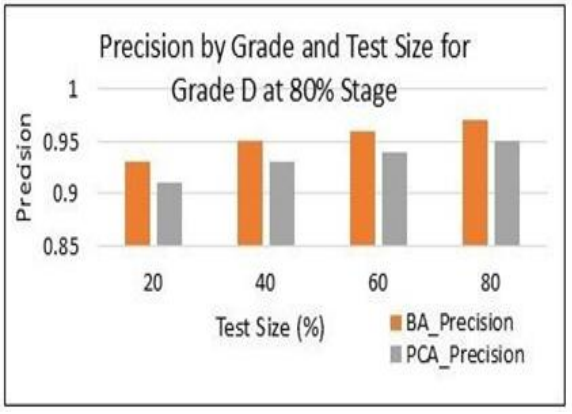
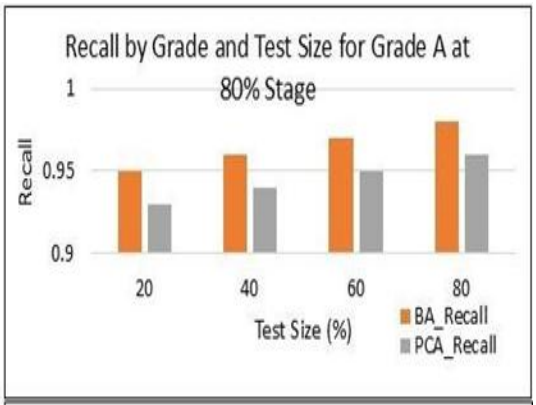




Figure 10: Precision by Grade at 80% Completion Stage



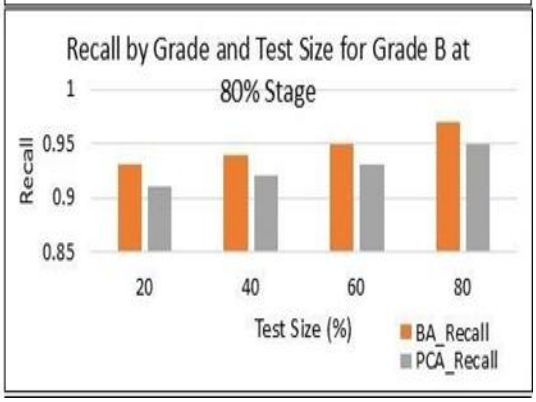
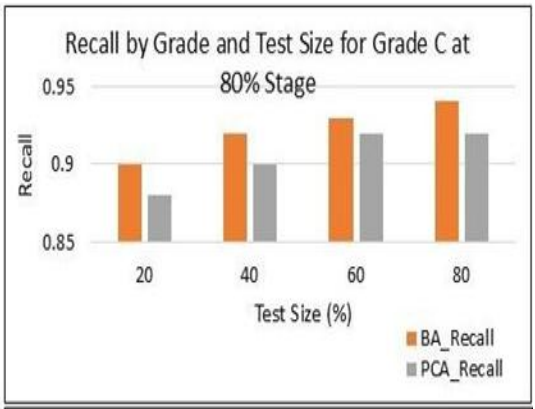




Figure 11: (a) Recall of Grade A & B at 80% Completion Stage



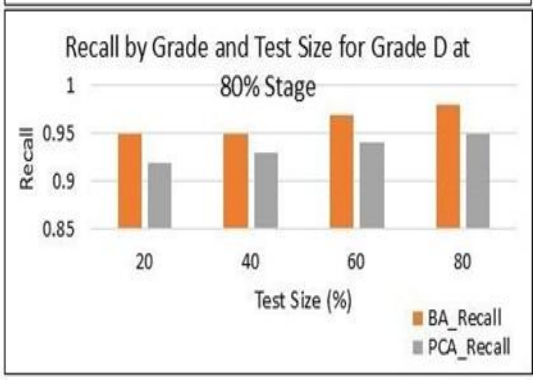




Figure 12: (b) Recall of Grade C & D at 80% Completion Stage

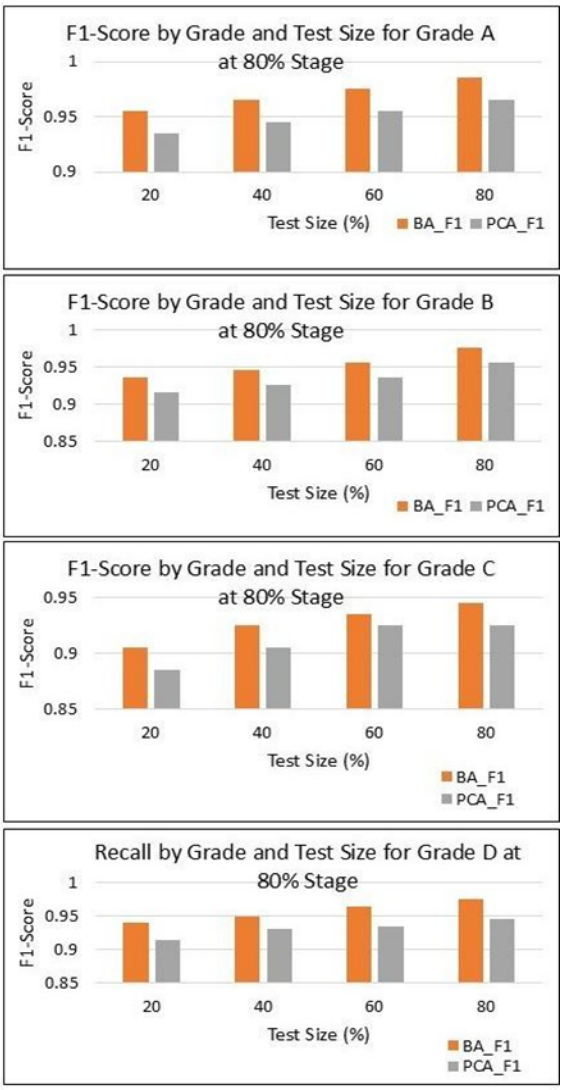



Figure 13: F1-Score by Grade at 80% Completion Stage

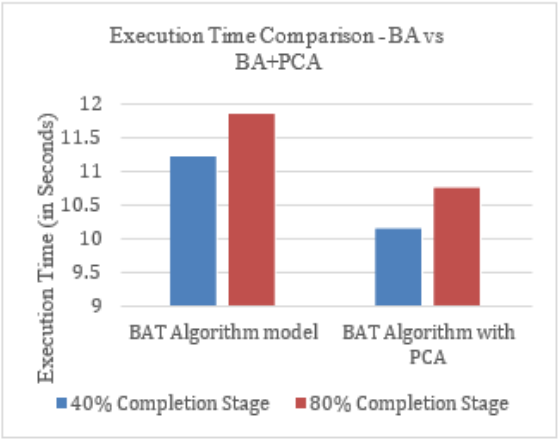



Figure 14: Comparison of Execution Time (Bat Algorithm Vs Bat Algorithm with PCA)

*D. Observations and Discussion*

This section interprets in detail the experimental results obtained from the Bat Algorithm based ensemble model and its PCA enhanced variant. The analysis covers the two stages of prediction, known as 40% course completion and 80% course completion, intended to mimic early and late academic forecasting. Analysis is conducted using classification metrics such as accuracy, precision, recall, and F1 score in all grade categories from A to D status, as well as execution time performance as an indicator of computationally efficient execution.

The analysis is further presented in the subsequent tables and visualized in Figure 3 to Figure 6. As observed from any 40% course completions, the BA-based ensemble model always demonstrates better predictive accuracy compared to it PCA-based adjustments. More succinctly, accuracies across all grades are significantly better for the BA model at all testing levels, especially in A and B grades. The precision and recall results also favouring the BA model suggest that PCA integration diminishes the available features' true impact on the classifier's ability in properly classifying grades. More specifically, F1-scores in the Fail category were substantially higher in the BA model during early course prediction, with the 80% testing resulting in a

0.94 score compared to 0.47 in PCA. This result indicates that the BA model is more equipped to handle the class imbalance and accurately predict low-performing students in the early-stage predictions.

PCA reduces execution time by approximately 9.6% at the 40% stage – from 11.23 to 10.15 seconds – but it also causes a reasonably sized drop in the prediction performance, mostly for the lower grade categories. This gap is depicted in the radar plot in Figure 7, which shows that the BA model has better balance across all the four-evaluation metrics.

These findings may imply that using this approach at the early prediction stages discards subtle but essential information regarding underrepresented grades.

Both models perform significantly better at the 80% stage, which implies that a more considerable amount of student data makes it easier to build accurate predictive models. All accuracy values for all grades are above 97% in both models, especially in the

BA+PCA variant, indicating smaller gaps from the ideal performance. While the BA model still slightly outperforms PCA in most metrics, with the gaps between them becoming visibly smaller. This correlation suggests that PCA might be a more optimal choice when enough data is available with minimum accuracy trade-offs. The execution time is still better for the PCA enhanced model which completes the run in 10.75 seconds compared to 11.85 for the standard BA model as per Figure 14. Class-wise trend analysis reconfirms that.

Both models reach the maximum possible score for grades A and B at the 80% stage but the BA+PCA model also exhibiting higher performance results for C and D grades compared to the 40% stage, which demonstrates that the PCA transformation preserves more discriminative power on richer dataset, making it a better choice for later stage academic performance prediction requirements.

To sum up, the BA-based ensemble model without PCA is more suitable for early-stage prediction, as the feature richness is essential to fully capture the student performance subtleties. The PCA-enriched variant, on the other hand, provides the opportunity to achieve similar performance results in a later stage while additionally being advantageous in terms of execution time. The revealed tendencies clearly indicate that the strategy should be adaptive, and the necessity of the PCA usage can be decided based on the amount of data available and the limitations of the deployment environment in terms of computational power. In addition, the results affirm the robustness of the BA model indifferent to grade distributions balanced or imbalanced, which solidifies it as a good candidate for educational data mining tasks.

*E. Discussion of Results*

The results from the BA-based ensemble classification model, both using and without PCA processing bring several important conclusions about the applicability of metaheuristic-driven FS in achieving successful student grade prediction. By assessing the CAs in two of the most critical stages throughout any learner's academic progress – 40% and 80% of full completion, the current research indicates the processes' dependency on SL volume, the sensitivity to CI, and the impact of DR techniques on obtaining better-than-human performance for several classes of grades. The presented discussion aims to reflect on the achieved quantitative benefits obtained by the BA-based model and further on the compensational implications of using PCA with respect to the signalling behaviour on the aforementioned academic progress stages, and conclude with a discussion of the pedagogical meaning in terms of applying the model to real-time educational decision- making.

The ensemble model at the 40% course completion stage performed excellently in all classification metrics of the Bat Algorithm with a strength mostly at the two ends of the performance continuum: identifying Well-Performing versus Poor-Performing Class groups of students. The BA model had reasonably high recall and f1 score for students in Grades C and D as evidenced by a fairly balanced recall and f1-scores. This indicated that the model had the ability to make the best use of the feature space and potentially uncover early and complex features that predict risk/promise. Given the BA model used an intelligent adaptive search process to guide feature selection, it was able to keep more vital dimensions that would otherwise have been cut down or become more diluted versions of the original data. The BA model was consequently a viable academic forecaster model early in the course as long as feature-rich data was available. On the other hand, BA+PCA was less effective at this stage, perhaps because it reduced feature richness too early into the semester. The total execution time for BA+PCA reduced remarkably, nearly 10%, but the model's performance also dropped, a trade-off probably not cost-effective at the important early forecasting stage.

By the 80% course completion stage, the gap in performance between BA and BA+PCA models significantly shrank. For the vast majority of grading levels, all models exhibited high and consistent accuracy, with certain metrics going above 97% due to the denser dataset. Once again, BA retained better results for the majority of categories with a special emphasis put on balanced F1- scores of grading levels, including low-academic ones. On the other hand, while BA+PCA remained incapable of properly classifying Grade C and D, it also demonstrated significant improvement in this regard, meaning that the reduced feature model may struggle to remain distant from a full-feature version but still preserve variation when trained on additional data.

In addition, BA+PCA maintained a shorter execution time for all the test sizes. Thus, reduced feature

models can be considered for real-time testing, including end-of-term scenarios that are highly dependent on fast predictive decisions and BA+PCA maintained a shorter execution time across all test sizes. The other metrics, such as accuracy and recall, experienced margin tradeoffs, which were still largely beneficial, considering the enhancements in computing efficiency. Overall, it seems that PCA-enhanced solutions
can prove more viable when more comprehensive student profiles become available.
The performance decline following the integration of PCA further exposes the difference in these two models' robustness. The BA-based model retained better accuracy and performance consistency balance, particularly on the 80% stage. Hence, BA is a superior machine learning model in conjunction with dimensionality reduction.

*F. Pedagogical Implications*

Accurate prediction of student outcomes at multiple stages of academic progress have obvious implications for early intervention and personalized student support. Our findings suggest that the BA-based ensemble model is indeed a useful tool to the institutional decision-makers, especially when applied in its full-featured mode. As mentioned before, the model can be applied at the 40% mark for initiating remedial support programs and at the 80% mark for reenforcing or reorienting the overall academic strategy.

At the same time, the dual-model approach – selecting between BA and BA+PCA depending on the availability of PCA data and time constraints – provides a flexible framework for institutions of higher learning to base their performance monitoring on. If there are constraints on how long it may take to process the data or how many computational resources may be expended, the PCA- enhanced mode is a very good strategic move. On the other hand, if classification performance is the primary concern – especially if early intervention is a goal – the full BA is preferable. Such considerations make the model flexible enough to be applicable across all the possible scales of academic performance monitoring, from large-scale institutional deployments to personalized classroom analysis.

## V. CONCLUSION AND FUTURE WORK

In summary, the BA-based ensemble classification model is among the best-performing in predicting the student academic grade. Although the PCA-enhanced model performed as well as the full-feature BA model in the 80% course completion stage, the full-feature of the former model is always the most appropriate for early-stage prediction models. This is because the BA model is iterative and uses fewer resources to classify, which is beneficial during the 40% stage with a small pool of data. On the other hand, the PCA-enhanced BA model is faster, and although lacking a considerable significance in execution time, is preferred in the 80% prediction stage.
Our findings provide evidence that the Bat Algorithm is a strong metaheuristic for FSB extraction in EDM and is adaptable for both phases of prediction and manageable for imbalanced data distribution. Our study also showed that stabilization of dimensions through PCA is not always a
good practice but is efficient in terms of FSB extraction at later stages of the education process without a significant loss of accuracy. This two-model method allows educational facilities to use strong early and late-evaluation models.
Our future work includes a few other directions. Firstly, it is possible to investigate other and more advanced dimensionality reduction techniques, such as t-SNE or autoencoder-based feature learning and compare them to PCA to achieve better computational efficiency as long as isometric mapping performance. Secondly, in addition to the Bat Algorithm, more metaheuristics, and not only BCA- related, can be integrated — Particle Swarm Optimization, Ant Colony Optimization, or other methods, as well as hybrid approaches. By doing this, we can explore the trade- offs between the aforementioned convergence rates and the classification robustness of the models. Finally, our system can be linked with model explainability methods, such as SHAP or LIME on a local level, unwrapping the decisions and making them clearer for the academic counselors and data administrators.
Overall, the effective expansion of this work to broader and more varied institutional datasets, incorporating more types of longitudinal student behavior data (such as attendance trends, learning

activity tracking, or course evaluations), may enhance the applicability and generalizability of the model, enabling it to make a real contribution to institutional higher education decision-support systems.

The Bat Algorithm-based ensemble framework is thus a solid, scalable, and interpretable predictive model for student success that has the potential to impact academic preparation and student support services in the future.

## REFERENCES


[1] Flayyih, M. F., & Tout, H. (2024). Predictive analytics model for students' grade prediction using machine learning. Babylonian Journal of Artificial Intelligence, 2024, 83–101. https://doi.org/10.58496/bjai/2024/011

[2] Banerjee, R. (2025). Understanding Student Performance Through Clustering in Educational Data Mining. Dandao Xuebao/Journal of Ballistics, 37(1), 120–131. https://doi.org/10.52783/dxjb.v37.183

[3] Ruggeri, M., Dethise, A., & Canini, M. (2023). On Detecting Biased Predictions with Post-hoc Explanation Methods. 17–23. https://doi.org/10.1145/3630050.3630179

[4] Alzahrani, M. R. (2024). Predicting Student Performance Using Ensemble Models and Learning Analytics Techniques. Mdpi Ag. https://doi.org/10.20944/preprints202406.1100.v1

[5] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in Nature Inspired Cooperative Strategies for Optimization, Springer, 2010, pp. 65–74, doi: 10.1007/978- 3-642-12538-6_6.

[6] Tang, L., & Sian, C. (2025). Educational Data Mining for Student Performance Prediction. Scalable Computing: Practice and Experience, 26(3), 1551–1558. https://doi.org/10.12694/scpe.v26i3.4336

[7] Pandey, A. C., & Rajpoot, D. S. (2021). Feature Selection Method Based on Grey Wolf Optimization and Simulated Annealing. Recent Advances in Computer Science and Communications, 14(2),635–646. https://doi.org/10.2174/2213275912666190408111828

[8] Kondo, N., Matsuda, T., & Hatanaka, T. (2022). Evaluation of Predictive Models in Institutional Research Based on Multi-Objective Optimization. IIAI Letters on Institutional Research, 1,1.https://doi.org/10.52731/lir.v001.018

[9] Brooks, C., & Greer, J. (2014). Explaining predictive models to learning specialists using personas. 26–30. https://doi.org/10.1145/2567574.2567612

[10] Li, Z. (2015). A redundancy-removing feature selection algorithm for nominal data. Peerj. https://doi.org/10.7287/peerj.preprints.1184

[11] Y. Sun, Z. Li, X. Li, and J. Zhang, "Classifier selection and ensemble model for multi-class imbalance learning in education grants prediction," Applied Artificial Intelligence, vol. 35, no. 4, pp. 290–303, Feb. 2021, doi: 10.1080/08839514.2021.1877481.

[12] M. Injadat, A. Shami, A. B. Nassif, and A. Moubayed, "Multi-split optimized bagging ensemble model selection for multi-class educational data mining," Applied Intelligence, vol. 50, no. 12, pp. 4506–4528, Jul. 2020, doi: 10.1007/s10489-020-01776-3.

[13] R. Kablan, H. A. Miller, S. Suliman, and H. B. Frieboes, "Evaluation of stacked ensemble model performance to predict clinical outcomes: A COVID-19 study," International Journal of Medical Informatics, vol. 175, p . 105090, May 2023, doi: 10.1016/j.ijmedinf.2023.105090.

[14] H. Xue and Y. Niu, "Multi-output-based hybrid integrated models for student performance prediction," Applied Sciences, vol. 13, no. 9, p. 5384, Apr. 2023, doi: 10.3390/app13095384.

[15] A. C. Mary and A. L. Rose, "Ensemble machine learning model for university students' risk prediction and assessment of cognitive learning outcomes," International Journal of Information and Education Technology, vol. 13, no. 6, pp. 948–958, Jun. 2023, doi: 10.18178/ijiet.2023.13.6.1891.

[16] A. Siddique et al., "Predicting academic performance using an efficient model based on fusion of classifiers," Applied Sciences, vol. 11, no. 24, p. 11845, Dec. 2021, doi: 10.3390/app112411845.

[17] M. Juez-Gil, Á. Arnaiz-González, J. J. Rodríguez, and C. García-Osorio,

"Experimental evaluation of ensemble classifiers for imbalance in Big Data," Applied Soft Computing, vol. 108, p. 107447, May 2021, doi: 10.1016/j.asoc.2021.107447.

[18] C. Li, L. Wang, J. Li, and Y. Chen, "Application of multi-algorithm ensemble methods in high-dimensional and small-sample data of geotechnical engineering: A case study of swelling pressure of expansive soils," Journal of Rock Mechanics and Geotechnical Engineering, vol. 16, no. 5, pp. 1896–1917, Feb. 2024, doi: 10.1016/j.jrmge.2023.10.015.

[19] Z. Noroozi, A. Orooji, and L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," Scientific Reports, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49962-w.

[20] H. Nugroho, N. P. Utama, and K. Surendro, "Smoothing target encoding and class center-based firefly algorithm for handling missing values in categorical variable," Journal of Big Data, vol. 10, no. 1, Jan. 2023, doi: 10.1186/s40537-022-00679-z.

[21] S. M. F. D. Syed Mustapha, "Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods," Applied System Innovation, vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086.

[22] L. Sha, D. Gašević, and G. Chen, "Lessons from debiasing data for fair and accurate predictive modeling in education," Expert Systems with Applications, vol. 228, p. 120323, May 2023, doi: 10.1016/j.eswa.2023.120323.

[23] Y. Li and X. Chen, "Boruta-based feature selection for student performance prediction," IEEE Access, vol. 10, pp. 21476–21485, 2022, doi: 10.1109/ACCESS.2022.3151950.

[24] H. Rezaei, X. Chu, and O. Bozorg-Haddad, "Grey wolf optimization (GWO) algorithm," in Advanced Optimization by Nature-Inspired Algorithms, Springer, 2017, pp. 81–91, doi: 10.1007/978-981-10-5221-7_9.

[25] A. Purushothaman, V. Jotwani and A. Yadav, "Multiclass Grade Prediction in Higher Education via Ensemble Learning Models and Grey Wolf Feature Optimization," 2025 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2025, pp. 143-149, doi: 10.1109/ICDT63985.2025.10986721.

[26] S. M. F. D. Syed Mustapha, "Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods," Applied System Innovation, vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086.

[27] J. Hu et al., "Chaotic diffusion-limited aggregation enhanced grey wolf optimizer: Insights, analysis, binarization, and feature selection," International Journal of Intelligent Systems, vol. 37, no. 8, pp. 4864–4927, Nov.2021, doi: 10.1002/int.22744.

[28] E. Akbari, S. A. Gadsden, and A. Rahimnejad, "A greedy non-hierarchical grey wolf optimizer for real-world optimization," Electronics Letters, vol. 57, no. 13, pp. 499– 501, Apr. 2021, doi: 10.1049/ell2.12176.

[29] A. E. Hassanien, M. Emary, and H. A. Ali, "Bat algorithm for medical disease diagnosis," Applied Soft Computing, vol. 14, pp. 439–446, Jan. 2014, doi: 10.1016/j.asoc.2013.08.025.

[30] U. K. Chakraborty, "Tuning controller parameters in power systems using the bat algorithm," Renewable Energy, vol. 178, pp. 102–111, Jul. 2021, doi: 10.1016/j.renene.2021.06.007.

[31] R. Bakar and A. A. Bakar, "Optimizing e-learning content sequencing using bat algorithm," Procedia Computer Science, vol. 162, pp. 751–758, 2019, doi: 10.1016/j.procs.2019.12.046.

[32] M. Parmar, K. Solanki, and P. Trivedi, "University examination timetabling using bat algorithm," International Journal of Computer Applications, vol. 177, no. 25, pp. 22– 26, Oct. 2020, doi: 10.5120/ijca2020920794.