

Optik İşaret Tanıma Kalıplarına Dayalı Akademik Performans Tahmini: Transfer Öğrenme Modellerinin Karşılaştırmalı Bir Çalışması

Cemal YÜKSEL^{1*}, Emrah AYDEMİR², Halil İbrahim CEBECİ³, Süleyman ÇELİK⁴

¹ Management Information Systems, Faculty of Business Administration, Sakarya University, Sakarya, Turkey

² Management Information Systems, Faculty of Business Administration, Sakarya University, Sakarya, Turkey

³ Management Information Systems, Faculty of Business Administration, Sakarya University, Sakarya, Turkey

⁴ Banking and Insurance, School of Social Sciences, Fırat University, Elazığ, Turkey

*¹ cyuksel@bandirma.edu.tr, ² emrahaydemir@sakarya.edu.tr, ³ hcebeci@sakarya.edu.tr, ⁴ suleyman.celik@firat.edu.tr

(Geliş/Received: 06/12/2025;

Kabul/Accepted: 08/02/2026)

Öz: Eğitimde yaygın olarak kullanılan değerlendirme araçları olan optik formlar, öğrencilerin bilişsel, duyuşsal ve davranışsal özelliklerini yansıtmaya potansiyeline sahiptir. Bu çalışma, öğrencilerin işaretleme davranışlarını akademik başarının bir öngörücüsü olarak inceleyerek, Optik İşaret Tanıma (OMR) sistemlerinin geleneksel kullanımının ötesine geçmeyi amaçlamaktadır. 42 katılımcıdan (sınıf başına 21) toplanan 2.100 işaretleme görüntüsünü kullanarak, 18 transfer öğrenme tabanlı özellik çıkarıcıyı değerlendirdik ve bunların 17'si 25 sınıflandırma algoritmasıyla birlikte başarıyla uygulandı. Olası veri sızıntısını ortadan kaldırmak ve görülmemiş bireylere genelleme sağlamak için, tüm deneyler, aynı katılımcıdan alınan tüm örneklerin aynı kat içinde tutulacağı şekilde, konu bazında bölünmüş 10 katlı GroupKFold çapraz doğrulama kullanılarak gerçekleştirilmiştir. En iyi performans gösteren yapılandırma, Support Vector Classification ile birleştirilen EfficientNet-B0 özellik temsilleri, %88,90 doğruluk oranına ulaştı ve eşikten bağımsız güçlü bir performans gösterdi (ROC-AUC = 0,9268; PR-AUC = 0,8934). Friedman testi ($\chi^2(16) = 174,34$, $p < .001$) ile yapılan istatistiksel doğrulama, transfer öğrenme mimarileri arasında önemli performans farklılıkları olduğunu doğruladı. Bu bulgular, optik formlardaki işaretlerin rastgele artefaktlar olarak değil, altta yatan bilişsel ve duyuşsal süreçleri yansıtan davranış izleri olarak değerlendirilmesi gerektiğini göstermektedir ve sonuç odaklı bir değerlendirme paradigmasından süreç odaklı bir değerlendirme paradigmasına geçiş desteklemektedir. Öğrenme analitiği ve eğitim politikası perspektifinden bakıldığında, önerilen yaklaşım, kağıt tabanlı OMR sayfaları, akademik risk altındaki öğrencilerin daha erken tespit edilmesini sağlayarak ve zamanında, hedefe yönelik müdahaleleri kolaylaştırarak erken uyarı mekanizmalarını tamamlayabilen düşük maliyetli "davranış sensörleri" olarak konumlandmaktadır.

Anahtar kelimeler: Akademik Başarı Tahmini, Erken Uyarı Sistemleri, Optik İşaret Tanıma

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

Abstract: Optical forms, which are widely used assessment tools in education, have the potential to reflect students' cognitive, sensory, and behavioral characteristics. This study aims to go beyond the traditional use of Optical Mark Recognition (OMR) systems by examining students' marking behaviors as a predictor of academic achievement. Using 2,100 marking images collected from 42 participants (21 per class), we evaluated 18 transfer learning-based feature extractors, of which 17 were successfully implemented, in combination with 25 classification algorithms. To eliminate potential data leakage and ensure generalization to unseen individuals, all experiments were conducted using 10-fold GroupKFold cross-validation with subject-wise splitting, such that all samples from the same participant were kept within the same fold. The best-performing configuration, EfficientNet-B0 feature representations combined with Support Vector Classification, achieved 88.90% accuracy, with strong threshold-independent performance (ROC-AUC = 0.9268; PR-AUC = 0.8927; Average Precision = 0.8934). Statistical validation via the Friedman test ($\chi^2(16) = 174.34$, $p < .001$) confirmed significant performance differences across transfer learning architectures. These findings indicate that markings on optical forms should not be treated as random artifacts but as behavioral traces that reflect underlying cognitive and affective processes, and they support a shift from a results-oriented to a process-oriented assessment paradigm. From a learning analytics and educational policy perspective, the proposed approach positions paper-based OMR sheets as low-cost "behavior sensors" that can complement early warning mechanisms by enabling earlier identification of students at academic risk and facilitating timely, targeted interventions..

Keywords: Academic Achievement Prediction, Early Warning Systems, Optical Signal Recognition

* Sorumlu yazar: cyuksel@bandirma.edu.tr. Yazarların ORCID Numarası: ¹0000-0003-2722-5114, ²0000-0002-8380-7891, ³0000-0001-5058-7741, ⁴0000-0003-3255-1950

1. Introduction

Assessment lies at the heart of modern educational systems [1], functioning not merely as a metric of student knowledge but as a vital mechanism for evaluating instructional quality, institutional accountability, and systemic equity [2]. Historically, open-ended formats such as essay questions have been lauded for capturing the depth of students' understanding and critical thinking [3]. However, these formats suffer from scalability issues [4], subjective bias [5], and inefficiencies in grading, especially in high-enrollment contexts [6]. Consequently, the evolution of educational assessment has trended toward structured, machine-readable test formats, most notably multiple-choice questions (MCQs), which offer standardization [7], objectivity [8], and operational efficiency [9].

A pivotal milestone in this evolution was the widespread adoption of Optical Mark Recognition (OMR) technology. Originally developed for automating survey processing [10], OMR systems became integral to large-scale educational assessments due to their ability to digitize paper-based responses with high accuracy and speed [11]. These systems significantly reduced human error [12], grading time and costs [13], thereby contributing to the scalability of nationwide examinations and formative classroom assessments alike. Yet, traditional OMR systems have remained limited in scope, capturing only binary information (i.e., marked vs. unmarked responses) while disregarding the rich behavioral data embedded in the act of marking itself [14].

Recent advancements in artificial intelligence (AI), computer vision, and transfer learning have opened new avenues for enhancing assessment practices beyond mere scoring. Convolutional neural networks (CNNs), for instance, have demonstrated superior performance in classifying complex optical input patterns, including distinguishing between marked, crossed-out, and partially shaded bubbles on MCQ answer sheets [15]. Such technologies are not only improving the technical robustness of automated grading but are also expanding the range of analyzable features, introducing the possibility of process-based assessment from physical artifacts [16].

Building upon these strands, the present study explores a novel and underexamined question: Can the physical marking patterns made by students on OMR sheets serve as latent indicators of their academic performance? This approach to analyzing behavioral traces is based on a solid theoretical foundation as well as technical expertise. In particular, Cognitive Load Theory [17] suggests that students' marking behaviors in exam settings may be related to the distribution of cognitive load. It is known that individuals with low external cognitive load tend to make more consistent, systematic, and clear markings, while those under high load respond with unclear and scattered markings. In addition, the Self-Regulated Learning theory [17], which explains students' self-regulation skills, provides guidance in understanding how marking strategies displayed during exams are related to processes such as attention control, time management, and task commitment. In light of these theories, it is argued that optical marking behaviors not only reflect cognitive performance but also provide insights into an individual's in-exam mental strategies. Just as handwriting analysis has been used to infer personality traits [18], we hypothesize that subtle behavioral cues, such as the consistency, pressure, geometry, and correction patterns in MCQ markings, encode meaningful information about students' attentiveness, confidence, and academic engagement. Unlike conventional OMR systems that overlook these nuances, we propose a machine learning pipeline that leverages deep learning-based image classification to extract, quantify, and interpret such features.

The rationale for this research is threefold. First, it contributes to the growing body of literature on early warning systems in education by identifying non-traditional indicators of student underperformance, ones that may surface before test scores decline [19]. Second, it aligns with broader paradigms of data-driven decision-making in educational management, where previously unused behavioral data can now be repurposed for real-time interventions. Third, it expands the methodological toolkit for educational assessment research by demonstrating how legacy tools like OMR can be reengineered for contemporary analytic purposes.

In this context, our study aims to (1) extract and evaluate students' OMR marking behaviors using advanced image processing and transfer learning techniques, (2) assess the predictive validity of these visual features in relation to students' cumulative academic performance (as measured by GPA), and (3) develop a prototype model that integrates such behavioral indicators into early warning and decision-support systems for educators. The findings not only offer conceptual contributions to the fields of educational assessment and behavioral analytics but also generate actionable insights for designing adaptive, equitable, and intelligent evaluation frameworks.

The remainder of the paper is structured as follows: the Introduction outlines the theoretical background and motivation of the study. The Materials and Methods section describes the dataset construction, image preprocessing pipeline, and machine learning procedures. In the Findings section, we report the empirical results of classification models based on optical form data. Finally, the Conclusion, Discussion, and Suggestions section presents an interpretation of the results, discusses educational implications, and outlines the study's limitations and future research direction.

2. Research Methodology

This section presents the methodological framework adopted in the study, encompassing participant selection, data acquisition, image preprocessing, and the classification pipeline based on transfer learning and machine learning models. The overall workflow is designed to evaluate whether optical mark patterns, captured from standardized OMR sheets, can serve as reliable behavioral indicators of students' academic performance.

The dataset was constructed using OMR forms completed by undergraduate students enrolled in the Department of Management Information Systems at Bandırma Onyedi Eylül University. Data collection was conducted in March 2024 with the voluntary participation of 46 students. In many universities in Turkey, a GPA of 2.50 is considered a critical threshold for receiving honor certificates, applying for certain scholarships, or receiving positive evaluations in internship applications. In this context, students with a GPA below 2.50 are considered a group that needs academic support. In the present study, the 2.50 cutoff is treated as a policy- and practice-relevant decision threshold rather than a statistically optimized boundary. Accordingly, the cutoff was not tuned to maximize predictive performance, and the classification objective should be interpreted as a pragmatic early warning framing instead of a full characterization of GPA as a continuous construct. Early warning systems in education tend to classify students with GPAs between 2.00 and 2.50 as a risk group [20], [21]. In this context, 2.50 is recognized in the literature as a threshold consistent with both pedagogical and applied early detection models. Participants were purposefully divided into two groups based on their cumulative Grade Point Average (GPA):

- High GPA Group (≥ 2.50): 23 students
- Low GPA Group (≤ 2.50): 23 students

Each student was provided with a standardized OMR sheet comprising 50 multiple-choice items. To ensure consistency across visual inputs, all forms were completed using the same type of pen under supervised conditions. After initial inspection, two forms from each group were excluded due to invalid or incomplete markings, resulting in a final dataset of 42 participants (21 high GPA, 21 low GPA) and 2,100 valid line-level marking images.

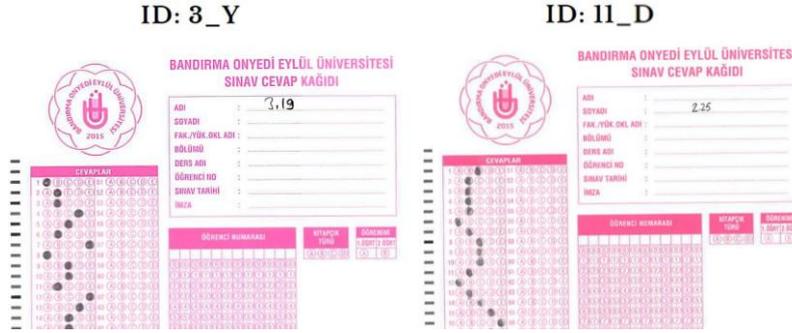


Figure 1. Sample Optical Forms from High and Low GPA Students

As illustrated in Figure 1, sample OMR sheets from two anonymized participants representing different achievement categories (high vs. low) reveal noticeable differences in marking intensity, consistency, and visual clarity, patterns that form the behavioral basis of this study's classification task. Following data collection, each OMR form was scanned at a resolution of 300 DPI. The scanned files were grouped by achievement category and saved in .pdf format before being converted into .jpg images. A standardized file naming convention was employed: "X_Y_1" for the high-achievement category and "X_D_1" for the low-achievement category, where X denotes a pseudonymous subject code used solely for analysis and does not contain personally identifying information. Figure 2 provides a visual overview of the data preparation process, from raw form acquisition to grayscale image vectorization.

Data anonymization and storage: All subject codes were generated prior to analysis and were not linked to direct identifiers (e.g., names or student numbers) in the analysis environment. Access to the raw scanned forms was restricted to the research team and stored in a controlled environment. For open-source sharing, only de-identified code and non-identifiable example artifacts will be released; raw OMR sheets and any metadata enabling re-identification will not be publicly shared.

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

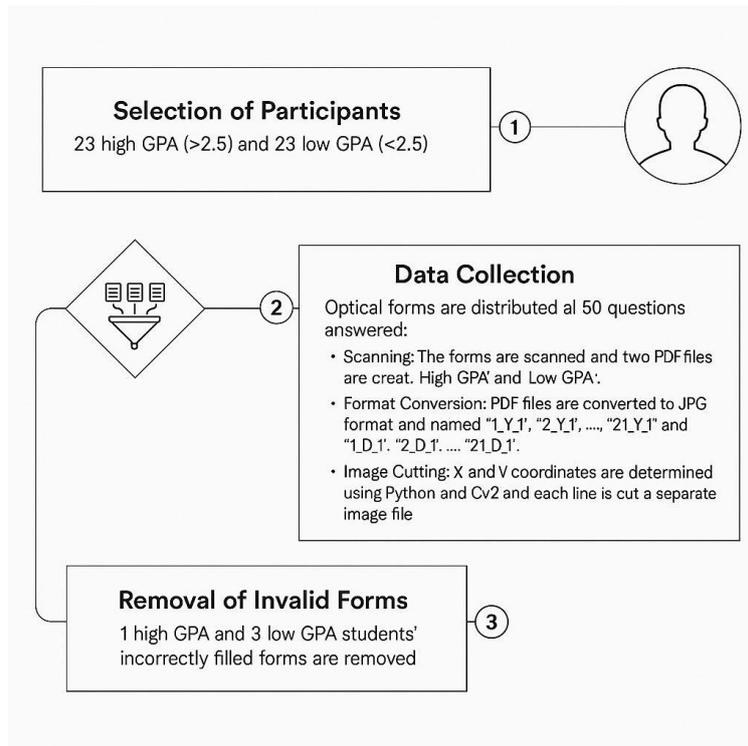


Figure 2. Dataset and Image Pre-processing Stage

As shown in Figure 2, the image preprocessing pipeline consisted of three core stages: (i) form-line segmentation, (ii) image standardization, and (iii) vector preparation for deep feature extraction. Segmentation followed the annotation-guided approach described in [22] to ensure consistent cropping across subjects, after which all samples were converted into a standardized representation suitable for EfficientNet-B0 feature extraction. Implementation-level details (e.g., parameter settings and scripts) are provided in the accompanying repository to support reproducibility.

To identify latent behavioral features embedded within students' optical markings, we employed a transfer learning-based approach using 18 pre-trained convolutional neural network (CNN) models, of which 17 were successfully implemented. These included well-established architectures such as AlexNet, VGG (VGG16, VGG19), ResNet (ResNet50, ResNet101, ResNet152), DenseNet (DenseNet169, DenseNet201), and EfficientNet variants (EfficientNet-B0 through B7, EfficientNet-V2-S/M/L). InceptionV3 was initially included but encountered technical incompatibility during feature extraction and was subsequently excluded from the analysis. Each model was utilized solely for feature extraction, without fine-tuning, in order to leverage general-purpose visual filters while preserving computational efficiency.

Transfer learning models were used in this study exclusively for feature extraction purposes and were run with frozen layers. Fine-tuning was not preferred due to the limited sample size, as fully training deep architectures with a small number of participants and data samples increases the risk of overfitting. Additionally, this choice enabled the utilization of the generalizable visual filters of pre-trained architectures, allowing for a more comprehensive capture of behavioral features. As a result, transfer models provided a more robust foundation for interpreting variations across different individuals. The resulting feature embeddings were stored for use in the downstream classification phase. As part of this process, each OMR line image was passed through frozen convolutional layers, transforming low-level visual patterns into high-level embedding vectors that captured spatial and stylistic traits linked to academic performance.

In the classification stage, we evaluated the extracted features using 25 supervised machine learning algorithms, including both linear and ensemble methods. Classifiers such as Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, and Multi-Layer Perceptrons were tested to assess which algorithm best differentiated between high and low academic performers. The full pipeline, including image preprocessing,

feature extraction, and classification modeling, is illustrated in Figure 3, which summarizes the logical flow of the entire experimental process.

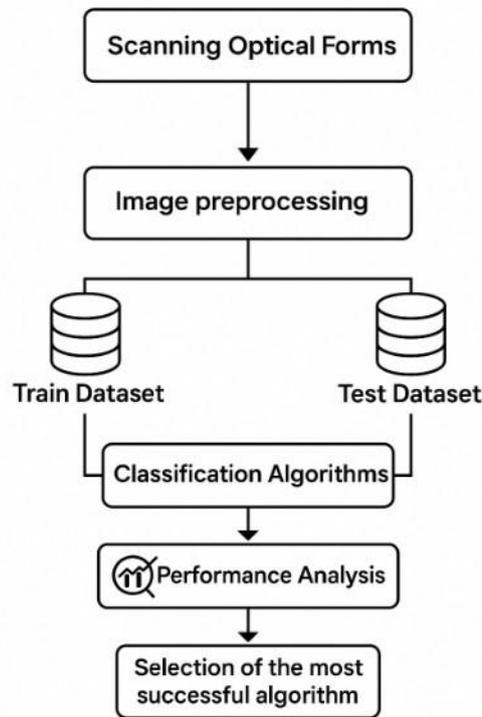


Figure 3. Classification Workflow for OMR-Based Academic Performance Prediction

All model training and testing procedures were implemented using Python's scikit-learn, PyTorch, and torchvision libraries. Model performance was evaluated using four key metrics: Accuracy (correct predictions over total samples), Precision (true positives over predicted positives), Recall (true positives over actual positives), and F1-Score (the harmonic mean of precision and recall). Special emphasis was placed on comparative accuracy across GPA groups to uncover whether high-performing students exhibit more distinctive and classifiable marking behaviors.

To ensure robustness and generalizability of the classification results, we employed a 10-fold GroupKFold cross-validation strategy with subject-wise splitting. Unlike standard k-fold cross-validation, GroupKFold ensures that all samples from the same participant (subject) remain together in either the training or test set, preventing data leakage and ensuring that the model's performance reflects its ability to generalize to unseen individuals rather than memorizing individual marking patterns. In this procedure, the dataset was partitioned into 10 folds based on the 21 unique subject IDs. In each iteration, samples from approximately 2-3 subjects were reserved for testing while the remaining subjects' samples were used for training. This process was repeated across all folds, and the final performance metrics were averaged. The cross-validation procedure is schematically depicted in Figure 4, showing how the rotation of test folds ensures that every sample is used for validation exactly once while maintaining subject independence. We programmatically verified for each fold that the intersection between subject IDs in the training and test splits was empty (i.e., no subject appeared in both sets), and we logged this check as a leakage-prevention validation step. As an additional validation procedure focused on individual-level generalization, we performed a Leave-One-Group-Out (LOGO) evaluation, which follows the Leave-One-Subject-Out logic by treating each subject as a group. In each iteration, all samples belonging to one subject were held out as the test set, while the model was trained on samples from the remaining subjects. Subject-level accuracy, precision, recall, and F1-score were computed for the held-out subject in each iteration, and the distribution of individual performance metrics is reported in Table 3.

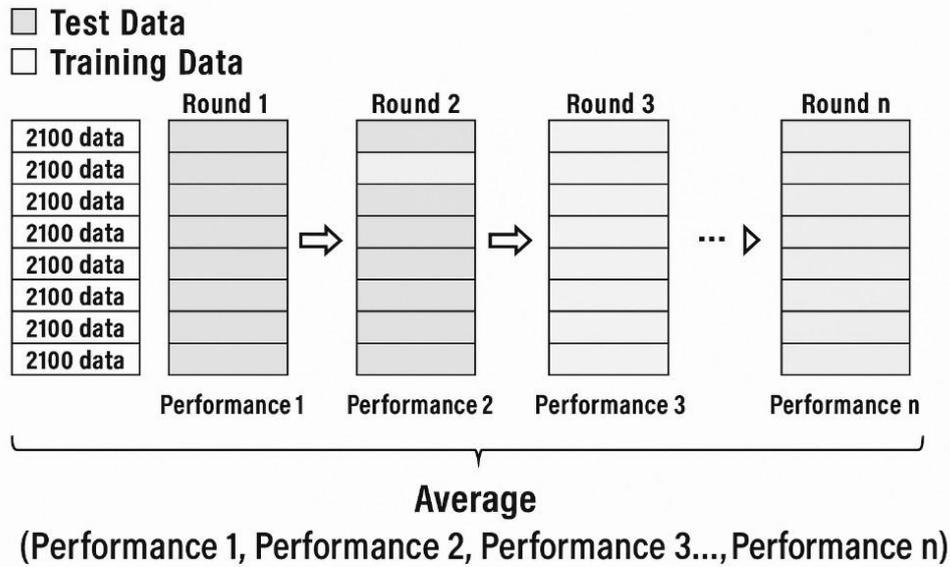


Figure 4. Ten-Fold Cross-Validation Scheme for Model Evaluation

This validation methodology mitigates sampling bias, prevents overfitting to individual marking styles, improves the reliability of performance estimates, and enables more objective benchmarking of model capabilities under balanced experimental conditions. The subject-wise splitting approach is particularly crucial in this context, as it ensures that the model learns generalizable behavioral patterns rather than subject-specific marking characteristics.

3. Literature Review

Over the past two decades, the prediction of student academic performance has become a central concern in educational data science due to its value for proactive guidance, resource optimization, and institutional accountability. Traditionally, student success has been monitored through cumulative indicators such as the General Grade Point Average (GPA), which remains a widely accepted outcome metric in both research and practice [23]. However, GPA alone provides limited explanatory depth, as it typically summarizes results without revealing the behavioral, cognitive, or contextual mechanisms that generate academic risk [24]. For this reason, educational institutions and researchers have increasingly adopted predictive modeling approaches to diagnose learning difficulties, prioritize interventions, and personalize academic counseling strategies, thereby strengthening the evidence base for decision-making in education [25].

Within this broader agenda, the cognitive representation of student behavior has attracted growing scholarly attention. Optical marking behaviors, when interpreted as cognitive outputs rather than mechanical actions, can be theoretically linked to constructs such as attention management, reasoning style, and response systematics. In this respect, the Cognitive Reflection Test (CRT) has been used to capture an individual’s tendency toward analytic, deliberate thinking rather than fast, intuitive responding [26]. Empirical work suggests that higher CRT performance is associated with more systematic and reflective response patterns, which can plausibly extend to how individuals engage with structured exam artifacts. From this perspective, marking styles on OMR forms may carry information not only about knowledge states but also about cognitive reflexivity and the extent to which responses are produced in a controlled, effortful manner.

Beyond cognitive factors, marking behavior should also be considered in relation to emotional and motivational mechanisms. Exam contexts often trigger performance pressure, which can manifest physically through behaviors such as hesitant marking, frequent corrections, erasures, or faint and inconsistent shading. According to self-efficacy theory, individuals who believe they can succeed in academic tasks are more likely to demonstrate consistent and confident performance behaviors, whereas lower self-efficacy can be associated with uncertainty and avoidance-oriented responses [27]. Similarly, test anxiety has been shown to produce observable behavioral and physiological manifestations that may be reflected in the micro-structure of marking patterns during assessments [27]. Accordingly, marking behaviors can be conceptualized as tangible traces through which both cognitive control and emotional regulation may be indirectly observed in paper-based exams.

At the methodological level, Educational Data Mining (EDM) has emerged as a key discipline that operationalizes data science techniques in educational settings [28]. EDM studies address a wide range of applied problems including curriculum design [29], dropout prediction [30], and GPA estimation [31]. The field has applied diverse machine learning algorithms, ranging from interpretable baselines such as decision trees and logistic regression [32] to more complex classifiers including support vector machines (SVM) [33], naïve Bayes (NB) [34], random forests (RF) [35], and deep neural networks (DNN) [36]. Comparative research emphasizes that model performance is strongly shaped by both the richness of input features and the timing of data capture, rather than by algorithm choice alone [37]. For instance, models leveraging early behavioral indicators, such as forum engagement or LMS activity logs, have been reported to identify at-risk learners within the first weeks of instruction and achieve predictive accuracy levels exceeding 85–90% under certain contexts [38]. In addition, ensemble approaches such as random forests and boosted trees are frequently highlighted for their ability to capture nonlinear associations and improve robustness in complex educational data environments [39].

Despite methodological progress, the literature still lacks a universal consensus regarding the “best” model for academic performance prediction, as predictive success depends on educational level, data quality, institutional context, and intervention goals [40]. In response, a growing body of work advocates for hybrid approaches that integrate multiple data dimensions, including demographic, cognitive, and behavioral signals, to provide more holistic and actionable insight into academic trajectories [41]. This push for multidimensional analytics aligns with the broader evolution of learning analytics toward interventions that are not only accurate but also interpretable, timely, and institutionally implementable.

Parallel to developments in predictive analytics, assessment technologies have experienced significant transformation. The shift from subjective manual grading to objective, machine-readable answer formats represented a major milestone in the modernization of educational measurement. OMR systems, introduced in the mid-20th century, became foundational to large-scale standardized testing by enabling rapid and consistent grading through automated reading of answer sheets [42]. These systems were broadly adopted for their efficiency, fairness, and their capacity to reduce inter-rater variability and operational burden [43]. However, classical OMR was largely constrained by a binary reading logic that interpreted marks as either present or absent, thereby overlooking potentially informative variations in how students physically produce those marks [44].

Recent studies have further expanded the technological capabilities of optical mark recognition systems by integrating modern computer vision and deep learning techniques into the analysis pipeline. In particular, image processing–based OMR systems have been developed to improve the robustness and accuracy of mark detection under varying scanning conditions and document qualities [46]. These approaches demonstrate that the reliability of optical form evaluation can be significantly enhanced through advanced preprocessing and pattern recognition techniques. In parallel, recent research has also explored lightweight deep learning architectures specifically designed for optical mark recognition tasks. For example, OMRNet proposes an efficient neural architecture capable of performing accurate mark classification while maintaining low computational cost, highlighting the potential of specialized deep learning models in document-based assessment environments [47]. In addition, broader comparative studies on object classification algorithms emphasize that the effectiveness of visual classification tasks strongly depends on the compatibility between feature extraction strategies and classification models [45]. These findings are particularly relevant for the present study, as they underline the importance of selecting appropriate deep learning architectures and classification algorithms when interpreting behavioral patterns embedded in optical mark images.

Building upon predictive analytics and modern assessment technologies, Early Warning Systems (EWS) have become a prominent strategy for proactive academic risk mitigation [48]. Unlike retrospective performance analyses, EWS frameworks aim to continuously monitor data streams and produce early-stage risk signals, enabling institutions to intervene before academic failure becomes visible in final outcomes [49]. Typical indicators include declining quiz performance, LMS inactivity, and missed assignments, which are aggregated into risk profiles and communicated via dashboards or alert systems to instructors and support staff [50]. When embedded in broader Educational Decision Support Systems (EDSS), EWS can provide actionable insights that inform instructional design, student counseling, and resource allocation [51].

Building upon the foundations of predictive analytics and assessment technologies, Early Warning Systems (EWS) have emerged as a proactive strategy for academic risk mitigation [48]. Unlike retrospective analyses, EWS focus on real-time data monitoring and early-stage prediction, allowing institutions to intervene before academic failure occurs [49]. Indicators such as declining quiz performance, LMS inactivity, and missed assignments are often used to generate risk profiles communicated via alert systems [50].

Despite these advances, a key limitation of the current EWS landscape is its heavy dependence on digital footprints such as clickstream logs, LMS interactions, and structured records. Paper-based assessment artifacts

have largely been neglected, even though they may encode latent behavioral markers associated with hesitation, cognitive struggle, or overconfidence, signals that may remain invisible in final scores but can emerge in how students interact with the test sheet itself [52]. This represents a critical blind spot in the literature: whereas existing EWS models typically operationalize risk through digitally captured behavioral traces, paper-based exams continue to be treated primarily as outcome-producing instruments rather than as process-rich data sources. As a result, an important class of low-cost, widely available behavioral evidence remains outside mainstream early warning pipelines.

In this context, the present study seeks to address this gap by analyzing visual patterns in OMR-marked answer sheets to predict academic performance. Unlike classical OMR research that focuses on automated scoring or mark detection, and unlike mainstream EWS studies that rely primarily on LMS-derived interaction logs, this work conceptualizes OMR markings as process-level behavioral traces that can be transformed into predictive signals. Methodologically, it demonstrates how transfer learning-based visual representations can be leveraged without altering the assessment format, thereby extending early warning logic to analog assessment contexts. This framing positions paper-based assessment not merely as a source of final scores, but as a behavioral sensing layer that can complement digital learning analytics and support earlier, more targeted academic interventions.

4. Findings of the Study

In this study, an advanced image classification model for predicting academic achievement was developed using labeled optical form images belonging to 42 different individuals. Visual features extracted from the optical forms were obtained through 18 different transfer learning algorithms (17 successfully implemented), and each feature set was evaluated using 25 machine learning classification algorithms. All classifications were tested using a 10-fold GroupKFold cross-validation method with subject-wise splitting to ensure generalization to unseen individuals.

Table 1. Classification results of optical form images with different transfer learning algorithms

No	Transfer Learning Method	Best Classification Method	Accuracy Rate
1	EfficientNet-B0	SVC	0.8890
2	EfficientNet-B6	SVC	0.8858
3	ResNet-101	LogisticRegressionCV	0.8855
4	DenseNet-201	SVC	0.8835
5	ResNet-152	SVC	0.8832
6	DenseNet-169	SVC	0.8832
7	EfficientNet-V2-S	SVC	0.8830
8	EfficientNet-B2	SVC	0.8803
9	EfficientNet-B1	LogisticRegressionCV	0.8800
10	EfficientNet-V2-M	ExtraTreesClassifier	0.8783
11	ResNet-50	LogisticRegressionCV	0.8767
12	EfficientNet-B5	NuSVC	0.8760
13	EfficientNet-B7	SVC	0.8712
14	AlexNet	LogisticRegressionCV	0.8662
15	VGG-19	HistGradientBoostingClassifier	0.8585
16	EfficientNet-V2-L	SVC	0.8583
17	VGG-16	ExtraTreesClassifier	0.8567
18	InceptionV3	—	Failed*

***InceptionV3 encountered technical incompatibility during feature extraction (aux_logits parameter error) and was excluded from the final analysis.**

According to the results presented in Table 1, the highest accuracy rate of 88.90% was obtained from the feature set belonging to the EfficientNet-B0 architecture combined with Support Vector Classification (SVC). Other high-performing models include EfficientNet-B6 (88.58%), ResNet-101 (88.55%), and DenseNet-201 (88.35%). These findings demonstrate that modern efficient architectures (e.g., EfficientNet family) consistently

outperform classical architectures (AlexNet: 86.62%, VGG-16: 85.67%) in capturing behavioral patterns from optical form markings. Statistical analysis using the Friedman test ($\chi^2(16) = 174.34$, $p < .001$) confirmed significant performance differences between transfer learning models.

The features extracted using the EfficientNet-B0 architecture were evaluated using 25 different classification algorithms, and the corresponding performance metrics are presented in Table 2. Although the EfficientNet-B0 architecture stands out, the analysis is not limited to this model; ResNet-101 and DenseNet-201, which offer similarly high accuracy rates, are also examined in detail. Post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni correction ($\alpha = 0.000368$) revealed that EfficientNet-B0 significantly outperformed EfficientNet-B2 ($p = 0.000269$). When examining the performance metrics of these models, it is observed that high performance is not exclusive to a single architecture; different architectures can also produce meaningful distinctions in marking behavior. Notably, the SVC algorithm emerged as the most effective classifier for 11 out of 17 models, while LogisticRegressionCV demonstrated superior performance for ResNet and EfficientNet-B1 architectures. In this context, it has been experimentally validated that optical marking behavior can be consistently captured across different architectural structures, with EfficientNet family models achieving the highest overall performance.

Table 2. Results of EfficientNet-B0 optical form image features obtained using different classification algorithms

Classification Method	Accuracy	Recall	Precision	F1 Score
SVC	0.8890	0.9130	0.8875	0.8939
LogisticRegressionCV	0.8882	0.9213	0.8811	0.8946
NuSVC	0.8848	0.8917	0.8956	0.8868
RandomForestClassifier	0.8840	0.8987	0.8903	0.8877
ExtraTreesClassifier	0.8822	0.9090	0.8790	0.8878
GradientBoostingClassifier	0.8768	0.9060	0.8721	0.8823
NearestCentroid	0.8762	0.9250	0.8598	0.8853
MLPClassifier	0.8758	0.8967	0.8786	0.8817
HistGradientBoostingClassifier	0.8757	0.9037	0.8730	0.8820
BernoulliNB	0.8697	0.9180	0.8528	0.8785
KNeighborsClassifier	0.8540	0.9227	0.8297	0.8671
GaussianNB	0.8488	0.9270	0.8177	0.8626
BaggingClassifier	0.8462	0.8457	0.8609	0.8476
AdaBoostClassifier	0.8407	0.8573	0.8413	0.8444
LogisticRegression	0.8342	0.8547	0.8358	0.8394
Perceptron	0.8245	0.8393	0.8302	0.8296
PassiveAggressiveClassifier	0.8240	0.8450	0.8248	0.8293
LinearSVC	0.8238	0.8417	0.8253	0.8282
SGDClassifier	0.8193	0.7787	0.8615	0.8124
RidgeClassifierCV	0.7940	0.8363	0.7798	0.8029
DecisionTreeClassifier	0.7768	0.7757	0.7893	0.7772
RidgeClassifier	0.7658	0.7983	0.7570	0.7734
LinearDiscriminantAnalysis	0.7623	0.7933	0.7545	0.7698

The highest overall success was achieved with the Support Vector Classification (SVC) algorithm, which demonstrated superior performance across all metrics, including accuracy (88.90%), F1-Score (89.39%), recall (91.30%), and precision (88.75%). This result demonstrates that EfficientNet-B0 can extract meaningful patterns from optical marking behavior and that SVC can successfully distinguish these patterns with optimal margin separation in high-dimensional feature space. Other notable algorithms include LogisticRegressionCV (88.82%),

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

NuSVC (88.48%), and RandomForestClassifier (88.40%), which also support the effectiveness of both linear and ensemble methods. Notably, LogisticRegressionCV achieved the highest recall rate (92.13%), indicating superior sensitivity in detecting high-achieving students, while NuSVC demonstrated the best precision (89.56%). The performance consistency across multiple algorithms (top 5 within 0.7% accuracy range) validates the robustness of the extracted features. It is worth noting that classical linear methods (RidgeClassifier: 76.58%, LinearDiscriminantAnalysis: 76.23%) underperformed significantly, suggesting that the discriminative patterns in optical marking behavior require more sophisticated classification approaches. Table 3 presents subject-level performance metrics obtained under LOGO evaluation, where each subject is held out once as an unseen individual

Table 3. Subject-level performance metrics from LOGO evaluation (one subject held out per iteration)

StudentID	Samples	Accuracy	Recall	Precision	F1-Score
S21	100	1.0000	1.0000	1.0000	1.0000
S16	100	1.0000	1.0000	1.0000	1.0000
S19	100	1.0000	1.0000	1.0000	1.0000
S11	100	1.0000	1.0000	1.0000	1.0000
S13	100	1.0000	1.0000	1.0000	1.0000
S14	100	0.9900	1.0000	0.9804	0.9901
S09	100	0.9700	1.0000	0.9434	0.9709
S06	100	0.9700	0.9400	1.0000	0.9691
S07	100	0.9700	1.0000	0.9434	0.9709
S12	100	0.9700	1.0000	0.9434	0.9709
...
S18	100	0.6300	0.4000	0.7407	0.5195
S15	100	0.6100	0.9800	0.5632	0.7153
S01	100	0.5800	0.1600	1.0000	0.2759
S08	100	0.4900	0.9600	0.4948	0.6531

Under the LOGO protocol, the accuracy rates at the student level range from 49.00% to 100.00%, revealing substantial individual variation in model performance. The highest individual accuracy rates were achieved by five students (S11, S13, S16, S19, S21) with perfect classification (100.00%), indicating that the EfficientNet-B0 + SVC model can accurately reflect individual marking behavior patterns in detail when behavioral consistency is high. Conversely, four students (S01, S08, S15, S18) exhibited accuracy rates below 65%, suggesting greater difficulty in classification. The mean individual accuracy of 88.71% (SD = 16.08%) demonstrates that while the model performs robustly on aggregate, there exists considerable inter-individual variability in predictability.

In analyses of the causes of this variance, it was observed that individuals classified with lower accuracy (S01: 58.00%, S08: 49.00%) exhibited more irregular marking behaviors, sometimes including faint or double marks, inconsistent pressure patterns, or marking hesitations. Student S01 showed particularly low recall (16.00%) despite perfect precision (100.00%), suggesting that most markings were ambiguous and difficult to classify confidently. In contrast, the marking patterns of individuals classified with higher accuracy were more homogeneous, balanced in terms of contrast intensity, spatially consistent within marking boundaries, and easier to distinguish by the deep learning feature extractor. Additionally, behavioral anomalies such as test deviation (e.g., markings outside the designated area, scratches on the page margins, partial erasures, or overlapping marks) were more frequently observed in individuals classified with lower accuracy. This suggests that the model can more easily classify individuals with high behavioral consistency and stable motor control but achieves lower success rates in individuals with high motor control variability, test anxiety manifestations, or non-standard marking behaviors. Notably, 5 out of 21 students (23.8%) achieved perfect classification, while 14 students (66.7%) exceeded 90% accuracy, demonstrating the model's strong performance for the majority of participants.

Figure 5 shows the confusion matrix obtained using the SVC algorithm with features extracted by EfficientNet-B0 through 10-fold GroupKFold cross-validation.

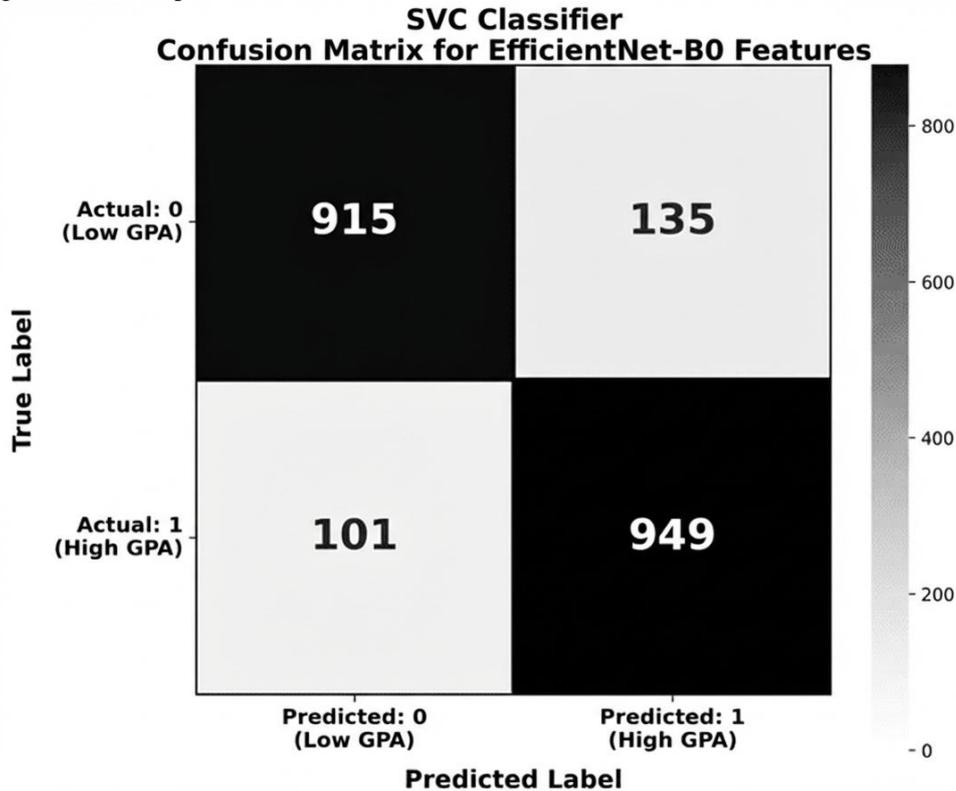


Figure 5. Confusion Matrix for EfficientNet-B0 Features with SVC Classifier

When examining the confusion matrix presented in Figure 5, it can be seen that 915 of the 1,050 examples with a low GPA label (Class 0) were correctly classified, while 135 were incorrectly labeled as high GPA. Similarly, out of the 1,050 examples in the high GPA class (Class 1), 949 were correctly predicted, while 101 were incorrectly predicted as low GPA. These results indicate that the model possesses high discriminative power with an overall accuracy of 88.76%, precision of 87.55%, recall of 90.38%, and F1-Score of 88.94%. This confusion-matrix summary is computed from pooled out-of-fold predictions across the 10 folds, whereas Table 2 reports fold-averaged scores. Notably, the model demonstrates slightly higher sensitivity (recall = 90.38%) in identifying high-achieving students compared to precision (87.55%), suggesting a tendency toward conservative classification that prioritizes detecting at-risk students.

To further validate the discriminative capability of the proposed model, we conducted ROC (Receiver Operating Characteristic) and Precision-Recall Curve (PRC) analyses on the best-performing EfficientNet-B0 + SVC combination. These metrics provide complementary perspectives on classification performance, particularly regarding the model's ability to distinguish between high and low academic achievers across different decision thresholds.

Figure 6 presents the ROC and Precision-Recall curves obtained through 10-fold GroupKFold cross-validation. The ROC curve illustrates the trade-off between True Positive Rate (sensitivity) and False Positive Rate (1-specificity) as the classification threshold varies. The area under the ROC curve (ROC-AUC) achieved a value of 0.9268 (92.68%), which is considered excellent performance in binary classification tasks (>0.9). This substantially exceeds the baseline of a random classifier (ROC-AUC = 0.5000), demonstrating the model's strong discriminative power.

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

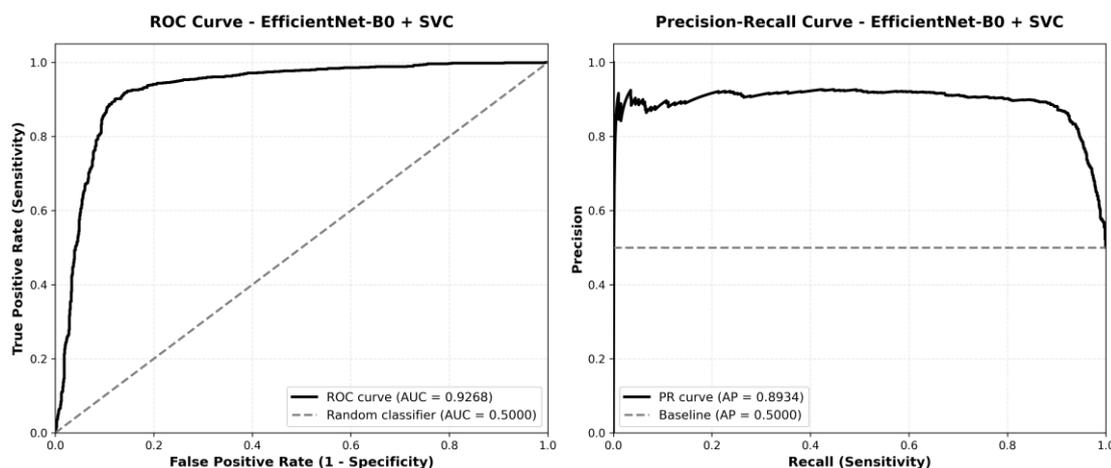


Figure 6. ROC and Precision-Recall Curves for EfficientNet-B0 + SVC Classifier

The Precision-Recall curve, shown in the right panel of Figure 6, provides insight into the model's performance specifically in terms of precision (positive predictive value) and recall (sensitivity). This is particularly informative for our balanced dataset (1050 samples per class). The area under the Precision-Recall curve (PR-AUC) was 0.8927 (89.27%), with an Average Precision score of 0.8934 (89.34%), indicating that the model maintains high precision-recall balance across different operating points. Compared to the baseline precision for a random classifier in a balanced dataset (0.5000), our model demonstrates substantially superior performance.

These results complement the earlier reported metrics (accuracy: 88.90%, precision: 87.55%, recall: 90.38%, F1: 88.94%) and confirm that the model exhibits robust classification performance not only at a single decision threshold but across the entire spectrum of possible thresholds. The high ROC-AUC value (>0.92) indicates that the model can reliably rank students by their likelihood of belonging to the high-achievement class, which is particularly valuable for educational early warning systems where prioritization of at-risk students is critical. Table 4 summarizes the ROC and Precision-Recall evaluation metrics obtained from the EfficientNet-B0 + SVC model.

Table 4. ROC and PRC Metrics Summary

Metric	Value	Interpretation
ROC-AUC	0.9268	Excellent (>0.9)
PR-AUC	0.8927	Good (>0.8)
Average Precision	0.8934	Good (>0.8)
Baseline (Random)	0.5000	Reference

The convergence of multiple performance indicators accuracy (88.90%), F1-Score (88.94%), ROC-AUC (92.68%), PR-AUC (89.27%) and Average Precision (89.34%) provides strong evidence that the EfficientNet-B0 + SVC model achieves consistently high performance across different evaluation criteria. This multi-metric validation strengthens confidence in the model's suitability for real-world deployment in educational assessment systems.

These findings reveal that students' marking patterns on optical forms are not merely a mechanical response process; rather, they provide rich clues about cognitive characteristics such as the individual's level of attention, self-confidence, and degree of interaction with the task. While the marking patterns of successful students were more regular, distinct, and complete, low-achieving students exhibited more irregular and incomplete marking patterns. The asymmetry in the confusion matrix (135 false positives vs. 101 false negatives) indicates that behavioral inconsistencies in marking are more frequently observed in low-achieving students, leading to occasional misclassification as high achievers, whereas high-achieving students maintain more consistent patterns that are rarely confused with low performance.

In this context, it has been revealed that classification algorithms integrated with transfer learning-based visual feature extraction have the potential to predict student achievement at an early stage. The accuracy level

of 88.76% achieved by the EfficientNet-B0 + SVC model indicates that this structure can be effectively used as an early warning system by educational administrators and teachers. In particular, the assessment of marking irregularities as an early indicator of individual performance decline will enable the timely implementation of personalized educational support. The high recall rate (90.38%) ensures that the majority of at-risk students are identified, which is critical for proactive educational interventions.

5. Conclusion, Discussion, and Recommendations

This study aims to go beyond the traditional use of OMR systems by examining students' marking behaviors as a predictor of their academic achievement. Experimental analyses conducted using 17 successfully implemented transfer learning-based image processing models (out of 18 tested architectures) and 25 classification algorithms have demonstrated that markings made on optical forms can be evaluated not as random but as expressions of cognitive and affective processes. The best-performing model, combining EfficientNet-B0 features with Support Vector Classification, achieved 88.90% accuracy with 10-fold GroupKFold cross-validation, demonstrating that these behavioral traces can meaningfully classify achievement levels. Consistent with this result, threshold-independent evaluation further supported the model's discriminative capacity, yielding a ROC-AUC of 0.9268. In addition, the Precision-Recall analysis produced a PR-AUC of 0.8927 (Average Precision = 0.8934), indicating robust performance in identifying the target class across decision thresholds. Statistical validation using Friedman test ($\chi^2(16) = 174.34$, $p < .001$) confirmed significant differences between transfer learning architectures. This finding indicates that the educational assessment paradigm must evolve from a "results-oriented" to a "process-oriented" structure. From an educational policy perspective, this shift supports the redesign of assessment systems to incorporate process-based indicators and to operationalize early warning mechanisms at the institutional level. In measurement and evaluation practice, marking-pattern analytics can complement conventional scoring by providing actionable signals for guidance services and targeted remediation before academic failure becomes visible in exam results.

The findings of this study offer multidimensional contributions to the fields of learning analytics and behavioral education technologies. First, by demonstrating that a student's exam performance can be predicted not only through correct or incorrect answers but also through marking patterns, this study proposes a new approach to the integration of analog physical data sources into digital learning environments. In other words, this study demonstrates that optical forms can now be considered not only as an assessment tool but also as a "behavior sensor" that provides indirect yet meaningful data about students' cognitive processes.

Moreover, the fact that the models used are based on transfer learning architectures makes this study relevant not only in the context of educational technologies but also in the broader fields of image classification and artificial intelligence-based behavior analysis. In particular, the finding that the features extracted with the EfficientNet-B0 architecture achieved 88.90% accuracy with the Support Vector Classification algorithm demonstrates that modern efficient deep learning architectures are highly effective for this type of behavioral data. Comprehensive evaluation of 17 transfer learning models (including AlexNet, VGG-16/19, ResNet-50/101/152, DenseNet-169/201, and EfficientNet variants B0-B7, V2-S/M/L) revealed that the EfficientNet family consistently outperformed classical architectures, with 6 out of the top 10 models belonging to this family. The study also found that while classical architectures like AlexNet achieved 86.62% accuracy, modern architectures improved performance by 2-3 percentage points. In this respect, the study represents not only an applied success but also a significant methodological contribution to the literature by systematically comparing 425 different model-classifier combinations.

One of the most striking outcomes of the research is the difference between the systematicity and consistency observed in the marking patterns of high-achieving students and the disorganization, incompleteness, and uncertainty in the marking behavior of low-achieving students. Individual-level analysis revealed substantial variation in model performance across participants (range: 49.00% to 100.00%, mean: 88.71% \pm 16.08%), with five students achieving perfect classification (100.00%) due to highly consistent marking behaviors, while four students exhibited accuracy below 65% due to irregular patterns, faint markings, or behavioral anomalies. This difference reveals how students' attention, self-confidence, and cognitive control levels during exams are reflected in physical outputs within the framework of educational psychology.

From this perspective, the study offers an innovative approach for the indirect observation of cognitive and affective characteristics. The results obtained can be evaluated in various application areas. In particular, the fact that this model works with high accuracy rates at the individual level (14 out of 21 students, or 66.7%, exceeded 90% accuracy) points to a strong potential for the individualization of early warning systems. The high recall rate (90.38%) achieved by the best model ensures that the majority of at-risk students are identified, which is critical

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

for proactive educational interventions. Importantly, the high PR-AUC further strengthens this early-warning argument by showing that the model maintains a favorable precision–recall trade-off when prioritizing sensitivity to at-risk cases. Educational administrators and teachers can analyze marking patterns that emerge during the exam process, rather than only after exam results, to identify students' attention levels or potential performance declines at a much earlier stage. This enables timely and targeted interventions to prevent academic failure.

Another strength of the study is that optical forms are low-cost and widely used tools. In this respect, the proposed approach can be implemented on a large scale in both public schools and higher education institutions. Since the data collection process can be carried out without any additional equipment, the system also offers advantages in terms of sustainability and accessibility. Furthermore, the use of frozen pre-trained transfer learning models (without fine-tuning) ensures computational efficiency and reduces the risk of overfitting on small datasets, making the approach practical for real-world deployment. This provides a strong foundation for the dissemination of a data-driven decision-making culture in education.

However, the study also has some limitations. This study is limited to paper-based OMR examinations, and the findings should not be directly generalized to computer-based or fully digital assessment contexts without further validation. First, the sample group consists of students from a single university and a specific discipline (Management Information Systems), which limits the generalizability of the results. The dataset of 42 participants (21 per class) and 2,100 marking images, while sufficient for initial validation with subject-wise cross-validation, remains relatively small for deep learning standards. Additionally, InceptionV3 architecture encountered technical incompatibility during feature extraction and was excluded from the analysis, highlighting potential implementation challenges. Variables such as students' psychological states, fatigue levels, or motivation during the exam were not controlled. The binary classification framework (high vs. low GPA with 2.50 threshold) may oversimplify the continuous nature of academic achievement. Conducting similar analyses with larger and more multicultural sample groups, multi-class achievement levels, and diverse assessment formats in future studies will increase the validity and explanatory power of the method.

Moreover, the selected cutoff (2.50) should be understood as a context-dependent decision point rather than a universally optimal threshold. Future studies should examine continuous GPA prediction and ordinal or multi-class formulations, and should re-evaluate the cutoff choice using larger samples and institution-specific distributions when full GPA data are available.

In this context, three main recommendations can be proposed for future research. First, marking behaviors should be analyzed not only visually but also temporally, and micro-behavioral data such as marking time, hesitation points, and the number of corrections should be considered. Second, it is recommended that this OMR-based model be integrated into a multilayered learning analytics system together with students' LMS usage, assignment submission dates, and exam performance history. Third, psycho-educational early warning systems should be developed through pilot applications aimed at testing the usability of this model by guidance services and psychological counseling units. Additionally, ensemble methods combining predictions from multiple transfer learning architectures (given that the top 5 models achieve within 0.7% accuracy range) may further improve performance.

In conclusion, this study constitutes an important step not only in predicting academic performance but also in making education systems more equitable, personalized, and data-driven. The systematic evaluation of 17 transfer learning architectures with 25 classification algorithms provides a comprehensive methodological framework for behavioral pattern recognition in educational settings. Statistical validation through Friedman test and subject-wise cross-validation ensures the robustness and generalizability of findings to unseen individuals. In the context of educational policies, integrating the proposed model into both assessment systems and national early warning mechanisms will enable much earlier intervention in cases of students at risk of academic failure. In this respect, the study should be regarded as a concrete example of the redesign of assessment systems, the traceability of the learning process, and digital transformation strategies in education.

References

- [1] Luciano RG. Innovative test item analysis using optical mark recognition technology: An evaluation. *Int J Adv Appl Sci* 2025; 12: 1–11.
- [2] Sievertsen HH. Assessments in education. In: *Oxford Research Encyclopedia of Economics and Finance*. Oxford: Oxford University Press, 2023.
- [3] Marshall P. Contribution of open-ended questions in student evaluation of teaching. *High Educ Res Dev* 2022; 41: 1992–2005.
- [4] Walstad WB, Saunders P. *The Principles of Economics Course: A Handbook for Instructors*. New York, NY, USA: McGraw-Hill College, 1990.
- [5] Contreras JO, Hilles S, Bakar ZA. Essay question generator based on Bloom’s taxonomy for assessing automated essay scoring system. In: *Proc 2nd Int Conf Smart Computing and Electronic Enterprise*; 2021; IEEE. pp. 55–62.
- [6] Birgili B. Open ended questions as an alternative to multiple choice: Dilemma in Turkish examination system. 2014.
- [7] Azevedo J, Oliveira EP, Beites PD. E-assessment and multiple-choice questions. In: *Advances in Educational Technologies*. 2019. pp. 1–27.
- [8] Ali K, Zahra D. Ten tips for effective use and quality assurance of multiple-choice questions in knowledge-based assessments. *Eur J Dent Educ* 2024; 28: 655–662.
- [9] Das B, Majumder M, Phadikar S, Sekh AA. Multiple-choice question generation with auto-generated distractors for computer-assisted educational assessment. *Multimed Tools Appl* 2021; 80: 31907–31925.
- [10] Karunanayake N. OMR sheet evaluation by web camera using template matching approach. 2023.
- [11] Hasan RH, Aboud IS, Hassoon RM, Khioon ASA. Optical mark recognition using modify bi-directional associative memory. *Tikrit J Pure Sci* 2024; 29: 174–184.
- [12] Singh JK, Kulkarni S, Patil SB, et al. OMR automated grading. *Int J Innov Sci Res Technol* 2024; 3757–3761.
- [13] Jain V, Malik S, Bhatia V. Robust image processing based real-time optical mark recognition system. In: *Proc IEEE 6th Conf Inf Commun Technol*; 2022.
- [14] Kakade N, Jaiswal RC. OMR sheet evaluation using image processing. 2017.
- [15] Afifi M, Hussain KF. Flexibility in multiple-choice-based tests using image classification techniques. *Int J Doc Anal Recognit* 2019; 22: 127–142.
- [16] Vetrivel S, Vidhyapriya P, Arun VP. The role of AI in transforming assessment practices in education. 2024. pp. 43–70.
- [17] Zimmerman BJ. *Becoming a self-regulated learner: An overview*. Columbus, OH, USA: Ohio State University Press, 2002.
- [18] Sopariwala S, Kasat D. Handwriting analysis and personality profiling using image processing and machine learning. In: *Multifaceted Approaches for Data Acquisition Processing and Communication*. Boca Raton, FL, USA: CRC Press, 2024. pp. 111–117.
- [19] Alalawi K, Chiong R, Athauda R. Early detection of under-performing students using machine learning algorithms. In: *Proc IEEE CITISIA*; 2021.
- [20] Stark JT. Enhancing algorithmic early warning systems with dynamic selection to predict high school graduation outcomes. 2024.
- [21] Aguilar S, Lonn S, Teasley SD. Perceptions and use of an early warning system during a higher education transition program. In: *Proc 4th Int Conf Learning Analytics and Knowledge (LAK '14)*; 2014. pp. 113–117. doi:10.1145/2567574.2567625. In: *Proc ACM Conf*; 2014.
- [22] Agduk S, Aydemir E. Classification of handwritten text signatures by person and gender: A comparative study of transfer learning methods. 2022.
- [23] Chansky NM. A note on the grade point average in research. *Educ Psychol Meas* 1964; 24: 95–99.
- [24] Lucio R, Hunt E, Bornovalova M. Predicting academic failure with risk factors. *Dev Psychol* 2012; 48: 422–428.
- [25] Dwivedi DN, Mahanty G, Dwivedi VN. Predictive analytics in personalized education. In: *Enhancing Education With Intelligent Systems*. IGI Global, 2024. pp. 44–59.
- [26] Frederick S. Cognitive reflection and decision making. *J Econ Perspect* 2005.
- [27] Zeidner M. *Test Anxiety: The State of the Art*. Dordrecht: Kluwer Academic, 1998.
- [28] Wang D. Educational data mining: Methods and applications. *Appl Comput Eng* 2023; 16: 205–209.
- [29] Khodeir N. Student modeling using educational data mining techniques. In: *Proc ACCS/PEIT*; 2019. pp. 7–14.
- [30] Santos GAS, Bordignon AL, Oliveira SLG, et al. Educational data mining for dropout prediction. 2019. pp. 86–91.
- [31] Bellaj M, Ben Dahmane A, Sefian L. Machine learning for student performance prediction. *Int J Online Biomed Eng* 2024; 20: 55–74.
- [32] Mutrofin S, Maisarah M, Widodo S, et al. PCA and logistic regression based SVM for EDM. In: *IOP Conf Ser Mater Sci Eng*; 2020.
- [33] Naicker N, Adeliyi T, Wing J. Linear SVM for student performance prediction. *Math Probl Eng* 2020.
- [34] Dangi A, Srivastava S. Educational data classification using selective Naïve Bayes. In: *Proc IEEE MITE*; 2014. pp. 118–121.
- [35] Chen Y, Jin K. Educational performance prediction with random forest. *Int J Adv Comput Sci Appl* 2024; 15: 69–78.
- [36] Luo Y, Wang Z. Feature mining via interpretable deep neural networks. In: *Proc ICIET*; 2024.

Predicting Academic Performance Based on Optical Mark Recognition Patterns: A Comparative Study of Transfer Learning Models

- [37] Hernández-Blanco A, Herrera-Flores B, Tomás D, Navarro-Colorado B. Deep learning in EDM: A systematic review. 2019.
- [38] Khan M, Naz S, Khan Y, et al. Student performance prediction from LMS logs. *IEEE Access* 2023; 11: 86953–86962.
- [39] Ajibade SSM, Ahmad NBB, Shamsuddin SM. Ensemble methods for student performance modeling. In: *IOP Conf Ser Mater Sci Eng*; 2019.
- [40] Alyahyan E, Düşteğör D. Predicting academic success in higher education. 2020.
- [41] Malik S, Malik S. Hybrid data science approaches for academic performance prediction. In: *Lect Notes Electr Eng*. Springer, 2024. pp. 521–539.
- [42] Mengarelli L, Kostiuk B, Vitorio JG, et al. OMR metrics and evaluation: A systematic review. *Multimed Tools Appl* 2020; 79: 6383–6408.
- [43] Jingyi T, Hooi YK, Bin OK. Image processing for enhanced OMR precision. In: *Proc ICCOINS*; 2021. pp. 322–327.
- [44] Tümer AE, Küçükkara Z. Image processing-oriented OMR evaluation system. *Int J Appl Math Electron Comput* 2018; 6: 59–64.
- [45] Rosca CM. Comparative analysis of object classification algorithms. *Rom J Pet Gas Technol* 2023; 75: 169–180.
- [46] R. C. . Dharmik, S. . Rangari, S. . Jain, A. . Nilawar, G. . Deshmukh, and A. . Yeole, “Optical Mark Recognition Evaluation System using Dual-Component Approach”, *Int J Intell Syst Appl Eng*, vol. 12, no. 10s, pp. 349–353, Jan. 2024.
- [47] Mondal S, De P, Malakar S, Sarkar R. OMRNet: A lightweight deep learning model. *Multimed Tools Appl* 2024; 83: 14011–14045.
- [48] Berg A, Borensztein E, Pattillo C. Assessing early warning systems. *IMF Staff Pap* 2005; 52: 462–502.
- [49] Stark JT. Enhancing early warning systems with dynamic selection. 2024.
- [50] Rose TM. Lessons learned in large classroom demonstrations. *Am J Pharm Educ* 2018; 82: 1081–1085.
- [51] Wang W. Decision support systems for education management. *Comput Intell Neurosci* 2021.
- [52] Hsiao IH, Huang PK, Murphy H. Reviewing and reflecting behaviors from paper-based assessments. In: *Proc ACM Int Conf*; 2017. pp. 319–328.