A MACHINE LEARNING APPROACH FOR EVALUATING AIR TRAFFIC CONTROLLER CANDIDATES: A COMPARISON WITH AHP AND ANP METHODS

Mustafa Özdemir[1]
Department of Air Traffic Control
Erzincan Binali Yıldırım University
Turkey
mozdemir@erzincan.edu.tr

Müjgan Sağır Özdemir
Department of Industrial Engineering
Eskisehir Osmangazi University
Turkey
mujgan.sagir@gmail.com

ABSTRACT

The safety and efficiency of air traffic operations depends heavily on air traffic controllers, making the accurate identification of suitable candidates a critical challenge for air navigation service providers. The selection process is traditionally based on standardized aptitude testing combined with structured interviews and expert evaluation, which may be affected by subjectivity and require substantial time and resources. This study introduces a machine learning based decision support approach for air traffic controller (ATCO) candidate selection. The candidate evaluation task is formulated as a binary classification problem to predict interview outcomes using historical data from an actual selection process involving 194 candidates, including 30 who passed the interview phase. Three machine learning models, Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT), are trained using candidate background variables such as prior examination scores, type of high school attended, and high school grade point average. The performance of the models is evaluated in comparison with existing approaches based on the Analytic Hierarchy Process and the Analytic Network Process. The results of this study provide preliminary evidence that, in the context of ATCO selection, machine learning models may approximate the outcomes of traditional interviews, with LR achieving an accuracy of 93%, and both DT and SVM achieving accuracies of 90%. Using the best-performing model, 46 eliminated candidates were correctly classified as eliminated (true negatives) and 9 selected candidates were correctly identified as selected (true positives), while 4 eliminated candidates were incorrectly classified as selected (false positives). These results suggest that the proposed approach may support air navigation service providers and related institutions by assisting early-stage screening through data driven prioritization of ATCO candidates.

[1] Corresponding author

Keywords: air traffic controller selection; personnel selection; interview; machine learning.

1. Introduction

Air traffic controllers (ATCOs) are responsible for maintaining the safe, orderly, and efficient movement of aircraft in airspace and at airports, where even small errors may lead to significant safety risks (Federal Aviation Administration, n.d.). Recent high-profile incidents (National Transportation Safety Board, 2025) have underscored the importance of robust air traffic management systems and drawn attention to broader issues such as air traffic controller selection, training, and staffing (Pan et al., 2025). In this context, selecting qualified ATCO candidates is essential given the critical demands of the role, including rapid decision-making and effective stress management. Moreover, ATCO training is highly structured and resource-intensive (Dönmez et al., 2020), requiring substantial human, material, and organizational resources (Eurocontrol, 2010). Therefore, given the associated costs and the substantial involvement of instructors, a robust candidate selection process becomes even more critical.

Traditionally, the ATCO candidate selection process relies on multistage evaluations (Pecena et al., 2013), including interviews designed to assess a range of non-technical and cognitive competencies. However, the interview process can be time-consuming and subject to limitations such as subjectivity and lack of standardization, which may affect fairness and consistency in selection. These challenges can make it more difficult to identify the most suitable candidates, potentially increasing training costs and affecting operational performance. In response, this study explores machine learning as a complementary alternative to the traditional interview phase. Specifically, it investigates whether interview outcomes can be approximated using a binary classification approach based on candidate-related features such as previous exam results, high school background, and grade point average (GPA). Accordingly, the study addresses the following research question: Can machine learning models approximate interview outcomes with reasonable accuracy and support a more systematic and efficient early-stage screening process for ATCO candidates using readily obtainable candidate information?

2. Background

Multi-criteria decision-making (MCDM) methods have been widely applied in personnel selection across various domains. For example, Smith and Bayazit (2024) present a comprehensive application of the Analytic Hierarchy Process (AHP) in the context of faculty recruitment at a regional U.S. university. Traditionally reliant on institutionally defined processes, the selection committee adopted the AHP to address limitations in quantifying and prioritizing candidate evaluation criteria. Nabeeh et al. (2019) address the challenge of personnel selection under uncertainty by integrating neutrosophic AHP and TOPSIS methods to support multi-criteria decision-making and apply the approach to the selection of a customer service manager.

However, despite the critical importance of accuracy and objectivity in air traffic controller selection, research on structured and systematic selection approaches remains limited. Existing studies have primarily explored multi-criteria decision-making (MCDM) methods to support selection and assessment processes. Notably, the AHP and the ANP have been applied to address the limitations of traditional interview-based evaluations in ATCO selection.

Özdemir and Sağır Özdemir (2018) proposed an AHP approach for the interview phase of the ATCO selection process. Their methodology consisted of two main stages. In the first stage, the evaluation criteria and their corresponding intensity levels were identified, and the weights of the criteria were determined through pairwise comparisons using the AHP. In the second stage, the alternatives were assessed using the rating method implemented in the Expert Choice software. The model included the overall goal, the evaluation criteria with their calculated weights, and the intensity levels assigned to each criterion. Rating categories were prioritized through pairwise comparisons, and the same procedure was followed for all criteria. They determined the relative importance of the criteria through a survey completed by 13 experts, each with 3 to 25 years of experience in air traffic control. These experts had varied professional backgrounds, including roles such as academics, licensed air traffic controllers, ATC instructors, and members of interview panels (Özdemir & Sağır Özdemir, 2018). The hierarchical structure of their model is illustrated in Figure 1.

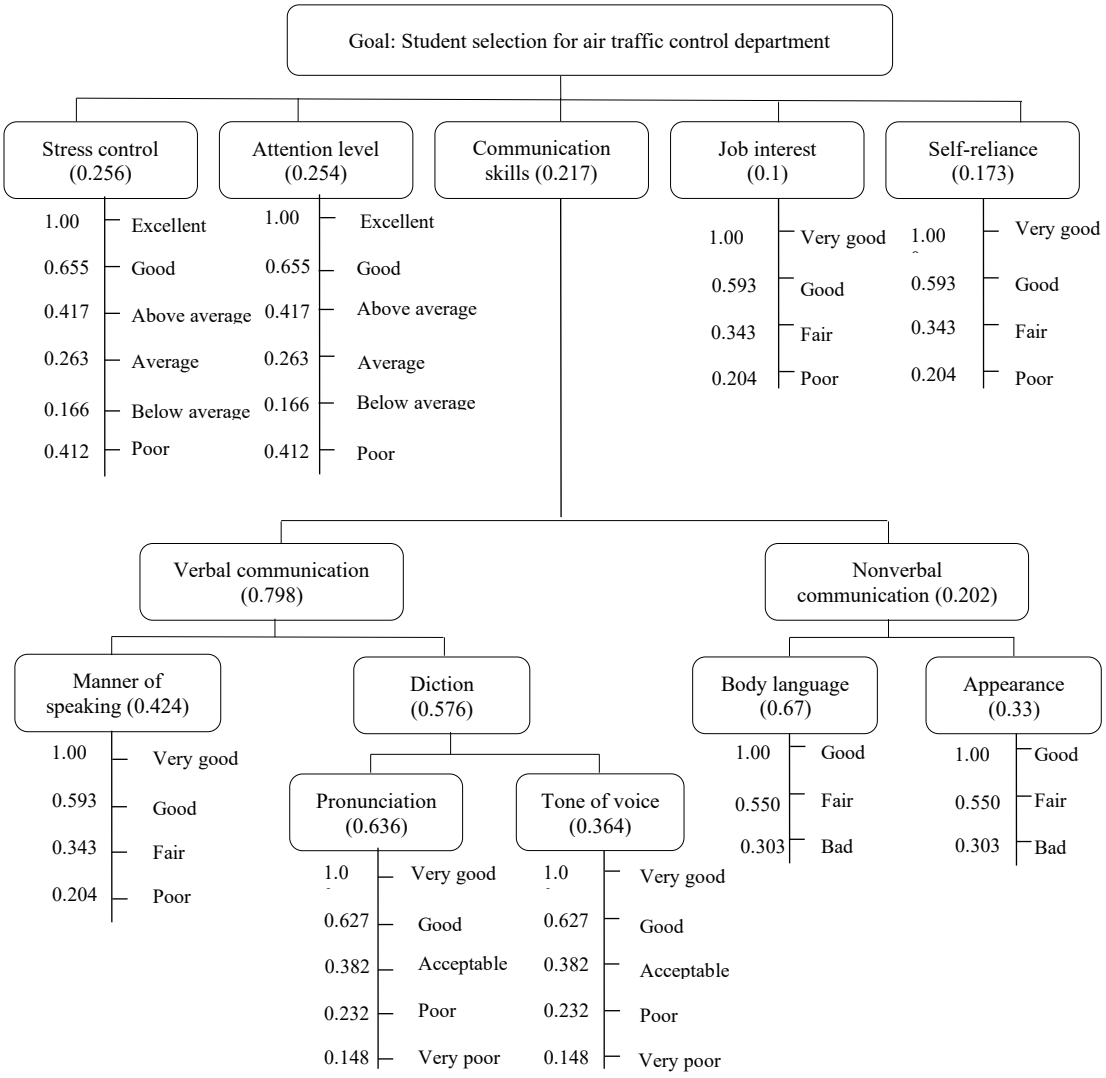

Figure 1 AHP model for the interview phase of ATCO selection (Özdemir & Sağır Özdemir, 2018).

In another study, Özdemir and Sağır Özdemir (2021) proposed an ANP model for evaluating the interview stage of the ATCO selection process. The rationale behind adopting the ANP was its ability to account for interdependencies among the evaluation criteria, which are typically overlooked in hierarchical models like the AHP. In their model, two types of relationships between clusters were identified: one-way dependencies, represented by directed arrows, and two-way dependencies, indicated by bidirectional arrows, as illustrated in Figure 2. Notably, two-way interdependencies were observed between stress control and self-reliance, self-reliance and attention level, as well as self-reliance and communication skills. Furthermore, inner dependencies existed among the sub-criteria of communication skills, which include manner of speaking, pronunciation, tone of voice, body language, and appearance.

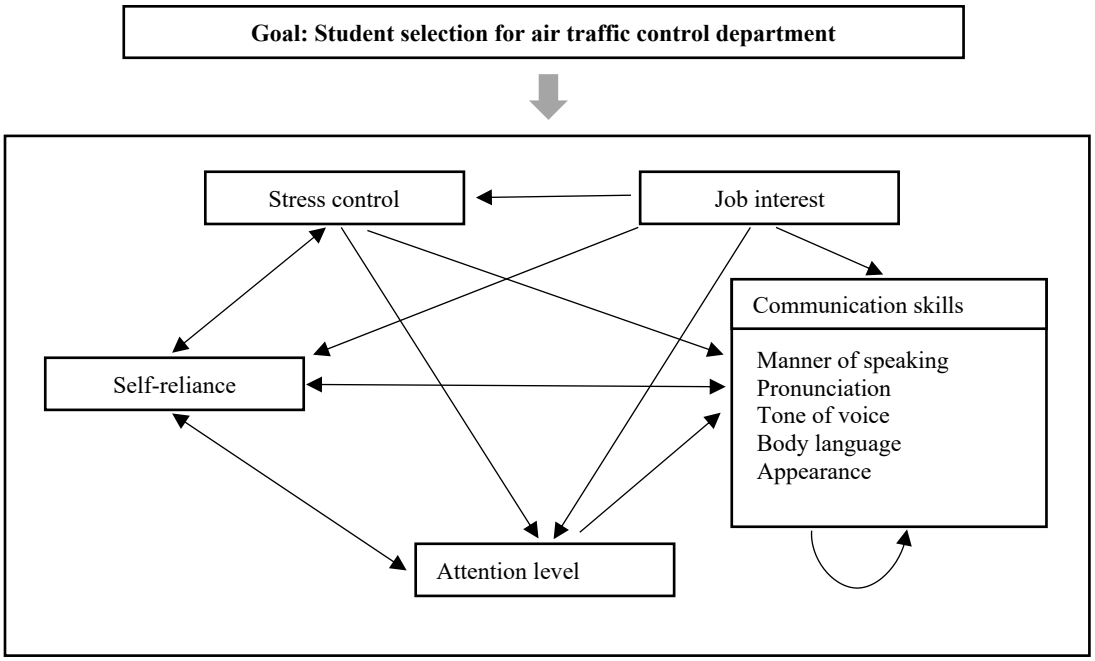

Figure 2 ANP model for the interview phase of ATCO selection (Özdemir & Sağır Özdemir, 2021).

Several others have applied MCDM approaches to ATCO selection. Iradukunda (2024) developed a hybrid MCDM approach for ATCO selection, integrating the AHP with Fuzzy TOPSIS and MOORA to evaluate candidates across health, educational, cognitive, and interpersonal criteria. Bongo and Seva (2023) applied a fuzzy DEMATEL–Best–Worst Method to analyze nine interrelated human factors using data from a control tower in the Philippines. Their findings identified communication as the most influential factor, with situation awareness, teamwork, vigilance, and attention acting as primary drivers of ATCO performance, while workload, trust, fatigue, and stress emerged mainly as reactive outcomes. Taylan et al. (2014) framed ATCO selection as a multi-criteria decision-making problem aimed at selecting the best candidate among ten alternatives based on five criteria: communication, mathematical skills, awareness, personal skills, and motivation. They employed a fuzzy decision tree to weight these criteria using evaluations from five decision-makers, followed by fuzzy TOPSIS to rank the candidates. However, these studies did not fully eliminate the interview phase; instead, they used these techniques as supplementary tools, still requiring the interview to be conducted.

One of the primary objectives of ATCO selection is to ensure candidates' success during training. Several studies have explored this link. Conzelmann et al. (2011) employed correlational and regression analyses to assess the predictive validity of various ATCO selection tools, including cognitive ability tests, work samples, team exercises, and semi-structured interviews. Basic cognitive abilities such as spatial orientation, working memory, and attention were identified as strong predictors of training success. However, the authors note that some trainees still fail during

training despite having passed the selection process, highlighting the limitations of the current tools in terms of predictive accuracy. This underscores the need for further refinement and improvement of ATCO selection methods. Fox (2014) conducted a study to support improvements in FAA's ATCO selection and training processes by assessing the extent to which general mental ability and biodata predict job performance, measured by the likelihood of meeting training requirements. Langr et al. (2019) argued that current standard aptitude assessments in ATCO selection are reaching their limits, and recommend a shift toward evaluating biodata profiles, learning potential, and instinctive decision-making under practical training conditions.

Recent studies have explored the application of machine learning models in human resource management; for example, Garg et al. (2022) reviewed 105 studies and highlighted the growing use of machine learning, particularly in recruitment and candidate evaluation processes. In another study, Koenig et al. (2023) investigated six real-world applications of machine learning in personnel selection and reported performance comparable to human raters in evaluating constructed responses. Hickman et al. (2024) applied closed-vocabulary text mining and elastic net regression using mock video interviews from 490 undergraduate psychology students to predict Big Five personality traits. Pham et al. (2025) evaluated logistic regression, support vector machines, decision trees, and other supervised models to predict university enrollment decisions using large-scale administrative admissions data. They found that program preference order, admission cutoff score, and high school graduation year were the most influential predictors of enrollment behavior.

However, different sectors require tailored methodologies reflecting their specific operational and safety requirements. In this regard, the aviation domain is particularly relevant, as it involves safety critical roles that demand high standards and rigorous selection procedures to ensure reliable service delivery. Although not focused directly on selection, a closely related study by Jenkins (2022) developed and compared several supervised classifiers, including logistic regression, single decision trees, random forests, and Extra Trees, to predict whether candidates would successfully complete the U.S. Air Force Specialized Undergraduate Pilot Training program. The Extra Trees achieved the highest performance (~94% accuracy), outperforming other models, and feature-importance analysis revealed that degree type and commissioning source were the strongest predictors of success.

The current study represents an early effort to apply machine learning as a tool for automating and standardizing ATCO candidate assessments, aiming to reduce variability and improve consistency. It seeks to address this gap by exploring the use of machine learning models to approximate interview evaluations, with the aim of supporting more scalable and consistent assessment processes. An earlier version of this work was presented as a conference paper (Özdemir & Sağır Özdemir, 2024), and the present article substantially extends that work.

3. Methodology

This study adopts a machine learning approach to predict the success of ATCO candidates in the interview phase based on their profile characteristics. The methodology consists of three main components: data preparation, model development, and performance evaluation. First, the dataset was preprocessed to handle class imbalance and prepare the features for training. Then, multiple

classification models were developed and tuned. Finally, the models were evaluated using several performance metrics to determine their predictive capability and reliability.

3.1 Data preparation and feature definition

The dataset includes background information for 194 ATCO candidates, 30 of whom passed the interview phase. Accordingly, the target variable was encoded as 1 for successful candidates and 0 for unsuccessful candidates. This distribution creates a class imbalance, with an approximate ratio of 1:5 between selected and eliminated candidates. To address this issue, class-weighted learning was employed in the machine learning models. Specifically, class weights were assigned inversely proportional to class frequencies in the training data, such that misclassification of the minority class (selected candidates) incurred a greater penalty during model fitting. This approach was intended to reduce bias toward the majority class while preserving the original class distribution.

To prevent data leakage, missing-value imputation was performed only after the train-test split. Missing values occurred only in the aptitude variable, since candidates who did not pass the mathematics exam were not eligible to take the aptitude test. These missing values were imputed using the k-nearest neighbors method based on related academic and background variables, with the imputer fitted only on the training set and then applied to the test set. In addition, during cross-validation, the same procedure was repeated separately within each training fold, and the learned imputation structure was applied only to the corresponding validation fold. Thus, no information from the test or validation data was used during preprocessing or model training.

Identifying the input variables that accurately predict a candidate's future success in ATCO training is a complex task. Candidate performance is influenced by a wide range of factors, and no single characteristic guarantees success. For instance, Fox (2014) found that both socioeconomic status and overall high school grade point average were significant predictors of ATCO job performance. Additionally, demographic and educational factors such as age, sex, and university background have been shown to influence selection outcomes. For example, Langr et al. (2019) reported that male candidates aged 26–32, holding technical university degrees and demonstrating above-average intelligence, achieved the highest success rates. In our study, we included similar candidate features to those mentioned above in the development of our machine learning models.

In this study, we identified 15 exploratory features, including gender; prior national exam results (six scores reflecting different academic domains such as science, social science, literature, and their combinations); aptitude test results; high school domain and type; high school GPA; mathematics score; previous application record; university acceptance status; and city of birth. Some features, such as city of birth, were included for exploratory analysis, whereas others were selected due to their relevance to core competencies and academic performance, which may reflect a candidate's potential for success in ATCO training programs. The city of birth is encoded as a binary variable, with a value of 1 indicating that the candidate was born in the city where the training institution is located. This feature may hold predictive value, as a significant portion of applicants are from this city. Limited recognition of the institution in other regions, combined with the financial and logistical challenges of relocation, may discourage candidates from more distant areas from applying. Including this variable allows the model to capture the influence of geographic accessibility and local familiarity, alongside academic performance and other personal background indicators.

3.2 Model development

According to prior research, machine learning applications in human resource management are predominantly formulated as classification tasks, with decision tree-based algorithms being among the most commonly used (Garg et al., 2022). In addition, the limited size of the dataset motivated the selection of models known to perform reliably in small-sample settings, including Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT) (Kokol et al., 2022).

LR is suitable for binary classification problems and provides interpretable coefficients, which help in understanding the relationship between candidate features and their predicted suitability (Hastie, 2009). The SVM was selected for its robustness in handling high-dimensional data and its effectiveness in capturing complex, non-linear patterns (Salcedo-Sanz et al., 2014). The DT was included for its simplicity and transparency, offering a clear visual representation of the decision-making process, which is especially useful for interpreting how individual candidate classifications are made (Navada et al., 2011).

3.3 Evaluation

Model performance was evaluated using accuracy, precision, recall, and the F1 score to provide a comprehensive assessment of classification performance. While accuracy reflects overall correctness, it can be misleading in imbalanced datasets with a predominance of unsuitable candidates. Therefore, precision and recall were used to assess the model's ability to correctly identify suitable candidates, and the F1 score was employed to balance these measures. All metrics were computed using stratified five-fold cross-validation to ensure robustness across data subsets.

4. Results

4.1 Predictor analysis and final feature set

Earlier, 15 exploratory predictors were identified and classified according to their measurement properties as continuous, binary, or categorical. Performance-related measures, including six examination scores, grade point average, mathematics score, and aptitude test results, were treated as continuous variables. Gender, prior application record, and university acceptance status were considered binary variables, while city of birth, high school type, and high school domain were treated as categorical variables.

Correlation analysis among the continuous variables revealed strong associations between the individual examination scores, grade point average, and mathematics score, indicating substantial overlap among these measures. Although such correlation does not necessarily reduce predictive performance, it may affect coefficient interpretability and model stability. Accordingly, aptitude test scores were retained as the most comprehensive among the overlapping performance-related measures, whereas the other highly correlated academic performance variables were excluded to achieve a more parsimonious and stable predictor structure. The resulting predictor set comprised seven variables: gender, prior application record, aptitude, university acceptance status, city of birth, high school type, and high school domain.

4.2 Model performance and prediction results

The three machine learning models were evaluated for their effectiveness in replicating ATCO interview assessments, and the corresponding results are presented in Table 1. LR achieved an

overall accuracy of 93%, with strong performance across both classes (weighted F1-score of 0.94), demonstrating its ability to distinguish between suitable and unsuitable candidates. The DT model had a slightly lower accuracy of 90% and a weighted F1-score of 0.90, with reduced precision for the positive class (0.67), indicating some limitations in classifying suitable candidates. SVM produced results similar to DT, with an accuracy of 90% and a weighted F1-score of 0.91. Overall, LR presented the highest accuracy and a more balanced performance, while DT and SVM also showed reasonably good classification performance.

Table 1
Performance metrics for machine learning models

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	Weighted Avg F1-score
LR	0.93	0.69	1.00	0.94
DT	0.90	0.67	0.67	0.90
SVM	0.90	0.60	1.00	0.91

Further analysis was conducted on the LR model to assess its ability to correctly classify selected and eliminated candidates, as it was the best-performing model. The performance of the logistic regression model was evaluated using a confusion matrix computed on the test dataset, as shown in Figure 3. The matrix provides a detailed breakdown of the model's predictions across the two classes: selected and eliminated candidates.

- The model correctly identified 46 eliminated candidates as eliminated (true negatives) and correctly classified 9 selected candidates as selected (true positives).
- However, the model incorrectly predicted 4 eliminated candidates as selected (false positives).
- No selected candidates were incorrectly classified as eliminated (false negatives).

These results indicate that the model performs strongly in identifying eliminated candidates, with a high number of true negatives and a limited number of false positives. All selected candidates in the test set were correctly classified, suggesting effective identification of the selected class despite its smaller representation in the dataset. The presence of a small number of false positives indicates a modest tendency to overpredict selection, which is a common characteristic when modeling imbalanced outcome classes.

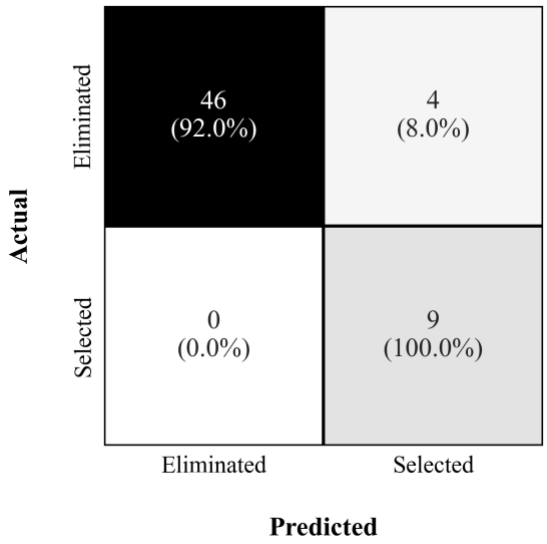

Figure 3 Confusion matrix for logistic regression model showing actual vs. predicted classifications

We evaluated the generalizability of the LR model across different sub-datasets using five-fold cross-validation. The average precision for Class 1 was 0.7486 ± 0.0980, indicating moderate to good precision in identifying selected candidates, albeit with some variability across folds. Recall for Class 1 remained high at 0.9333 ± 0.0816, demonstrating strong sensitivity to selected candidates. The weighted average F1-score reached 0.9412 ± 0.0284, reflecting a strong balance between precision and recall across classes. Overall classification accuracy was 0.9381 ± 0.0310, indicating robust model performance. While the model performs consistently well across folds, variability in precision suggests occasional misclassification of eliminated candidates. Figure 4 presents the confusion matrix for each fold of the five-fold cross-validation. As observed, the model correctly identified all selected candidates in three of the five folds, whereas variation across folds was primarily associated with occasional misclassification of eliminated candidates as selected, particularly in fold 4 (Figure 4d).

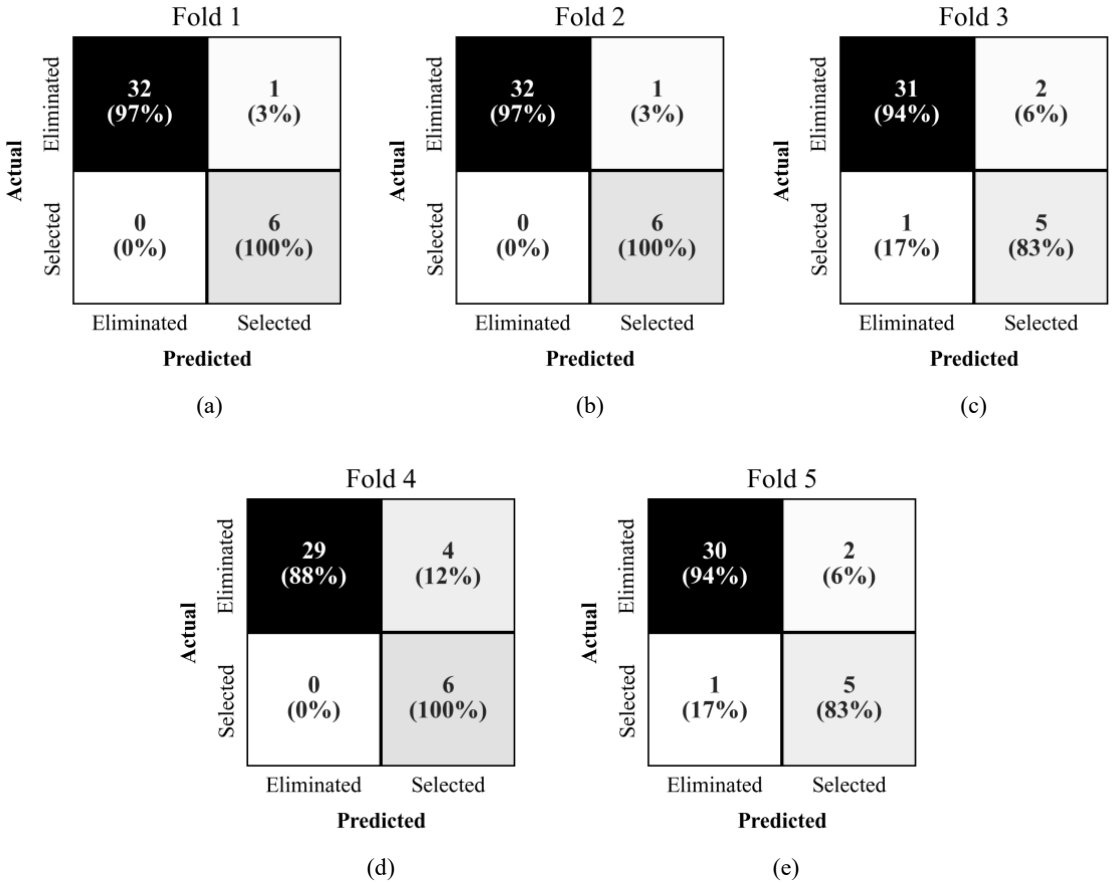

Figure 4 Confusion matrices for each fold of 5-fold cross-validation using logistic regression

Figure 5 illustrates the performance of the LR model in identifying selected and eliminated candidates on the test dataset. All eliminated candidates were correctly classified, while four selected candidates were incorrectly predicted as eliminated.

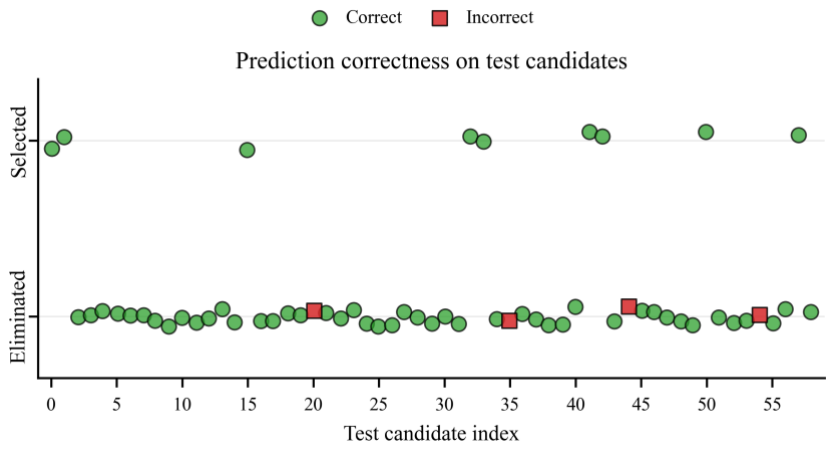

Figure 5 Prediction correctness of LR on test data

In comparison, Özdemir and Sağır Özdemir (2021) applied an ANP approach to the same dataset, and their model achieved accuracies of 77.8% for eliminated candidates (7/9) and 93.3% for selected candidates (28/30). In the current study, accuracy increased to 100% for selected candidates (9/9) and 92.0% for eliminated candidates (46/50). While direct comparison is limited by differences in sample size and potential variations in misclassification for candidates not shown in the figure, the results indicate improved identification of selected candidates and comparable performance in identifying eliminated candidates. Importantly, these results were achieved without the need for an interview, demonstrating the potential effectiveness of the proposed approach.

The feature importance results in Figures 6, 7, and 8 highlight the key attributes for predicting selected candidates across the LR, DT, and SVM models. In LR, the three most influential features are city of birth, gender, and aptitude, indicating that city of birth is the strongest contributor in this model, followed by gender and aptitude. In DT, aptitude emerged as the most important feature by far, with only limited contributions from previous record, city of birth, and university acceptance, while the remaining variables showed no meaningful importance. In SVM, the three most influential features were aptitude, previous record, and gender, suggesting that aptitude is the strongest differentiator in this model, while previous application history and gender also contributed to the classification. Overall, aptitude appears as a consistently influential predictor across models, whereas the relative importance of the other features varies depending on the modeling approach.

The emphasis on city of birth indicates that location may be a factor in candidate selection, possibly due to local familiarity with the school or accessibility. While this may initially seem counterintuitive, closer examination of the data reveals that 19.6% of candidates from the university's city were selected, compared to just 9.8% of candidates from outside the university's city. Another contributing factor could be that the city is a hub for aviation activities, hosting a pilot and ATC training school, a military jet base, and several aviation manufacturing companies. As a result, many residents are familiar with the aviation field, and a significant portion of the local population is employed in related sectors. This environment may foster greater interest and motivation among candidates to pursue a career in aviation. Additionally, local candidates may

already possess basic aviation knowledge, which could give them an advantage during the selection process compared to applicants from other cities.

This disparity highlights the potential role that geographic factors may play in the selection process. The high school domain, which may relate to candidates' educational backgrounds, and aptitude test scores further provide a strong basis for assessing suitability.

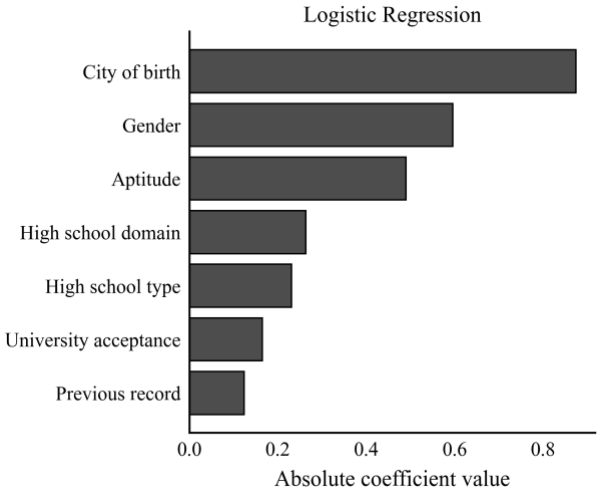

Figure 6 Feature importance based on LR

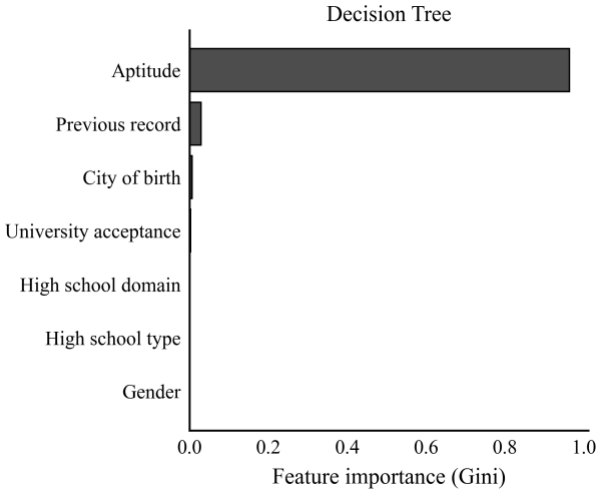

Figure 7 Feature importance based on DT

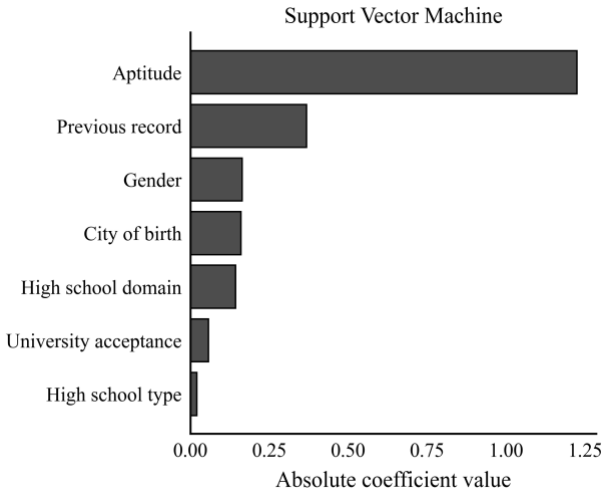

Figure 8 Feature importance based on SVM

5. Conclusions and limitations

To conclude, this study highlights the potential of machine learning as a practical alternative for improving the ATCO candidate selection process. In particular, the findings suggest that the proposed approach may offer a scalable, consistent, and data-driven method that reflects key elements of traditional interviews. Although some classification errors were observed, the models showed a reasonable ability to distinguish between successful and unsuccessful candidates. These findings indicate that machine learning can play a valuable role in enhancing candidate evaluation and decreasing dependence on time-consuming interviews. The study offers a foundation for further research and refinement toward a more effective and dependable ATCO selection system.

While this study introduces the potential of machine learning models to assess ATCO candidates, several limitations must be acknowledged. First, some misclassifications were observed; specifically, the LR model incorrectly classified four eliminated candidates as selected. However, a certain degree of inconsistency is also acceptable in classical multi-criteria decision-making methods such as the AHP and ANP. Therefore, small differences in results, such as a small number of eliminated candidates being classified as selected by the machine learning model, or vice versa, may be considered tolerable. This is expected in decision processes involving human judgment. Second, the dataset is limited to 194 candidates, which restricts model robustness and generalizability. A larger and more diverse dataset would enhance the reliability of the results and support broader application of the findings. Third, the study relies on interview grades as the target variable, which introduces the risk that the models may replicate any subjectivity or implicit bias present in the original assessments. Since the models were trained on historical interview outcomes, their predictions may also reflect patterns embedded in past human evaluations, with possible implications for fairness in candidate selection. This concern is consistent with challenges highlighted in previous research regarding the fairness and accountability of machine learning algorithms in human resource management (Garg et al., 2022). One way to mitigate this issue is to incorporate additional features that more accurately reflect candidate aptitude. Specifically, integrating cognitive assessment data could improve both the objectivity and predictive validity of

the models. As identified in a recent meta-analysis, cognitive factors such as quantitative reasoning, processing speed, working memory, and visuo-spatial ability strongly predict ATCO training success, whereas personality traits have limited predictive value (Mouratille, 2022). Including such cognitive indicators would allow future models to better capture the competencies essential for ATCO roles.

Despite these limitations, the study offers meaningful insights into the application of machine learning in ATCO selection and provides a foundation for future improvements in candidate evaluation practices. Continued research should focus on expanding feature sets, improving data quality, and fostering collaboration between domain experts and machine learning practitioners to ensure both accuracy and fairness in automated decision-making.

REFERENCES

Bongo, M. F., & Seva, R. R. (2023). Evaluating the performance-shaping factors of air traffic controllers using fuzzy DEMATEL and fuzzy BWM approach. *Aerospace*, *10*(3), 252. https://doi.org/10.3390/aerospace10030252

Conzelmann, K., Heintz, A., & Eißfeldt, H. (2011). A large scale validation study on air traffic controller selection and training–design, challenges and results. In *Proceedings 16th International Symposium on Aviation Psychology* (p. 375–380). 16th International Symposium on Aviation Psychology, May 2–5, 2011, Dayton, OH, USA. https://doi.org/10.1027/2192-0923/a000039

Eurocontrol. (2010). *Guidance for developing ATCO basic training plans* (Edition 2.0, ID 11/03/23-14). EUROCONTROL.

Dönmez, K., Demirel, S., & Özdemir, M. (2020). Handling the pseudo pilot assignment problem in air traffic control training by using NASA TLX. *Journal of Air Transport Management*, *89*, 101934. https://doi.org/10.1016/j.jairtraman.2020.101934

Federal Aviation Administration (FAA). (n.d.). *Section 5. Pilot/controller roles and responsibilities*. FAA. Retrieved November 23, 2024, from https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap5_section_5.html.

Fox, K. D. (2014). *Prediction of air traffic controller trainee selection and training success using cognitive ability and biodata (Publication No. 3614818)*. [Doctoral dissertation, Walden University].

Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2022). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, *71*(5), 1590-1610. https://doi.org/10.1108/ijppm-08-2020-0427

Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, *162*, 107268. https://doi.org/10.1016/j.infsof.2023.107268

Hastie, T., Tibshirani, R., & Friedman J. (2009). *The elements of statistical learning: data mining, inference, and prediction.* SpringNature Link.

Hickman, L., Saef, R., Ng, V., Woo, S. E., Tay, L., & Bosch, N. (2024). Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews. *Human Resource Management Journal*, *34*(2), 255–274. https://doi.org/10.1111/1748-8583.12356

Jenkins, P. R., Caballero, W. N., & Hill, R. R. (2022). Predicting success in United States Air Force pilot training using machine learning techniques. *Socio-Economic Planning Sciences*, 79, 101121. https://doi.org/10.1016/j.seps.2021.101121

Koenig, N., Tonidandel, S., Thompson, I., Albritton, B., Koohifar, F., Yankov, G., ... & Newton, C. (2023). Improving measurement and prediction in personnel selection through the application

of machine learning. *Personnel Psychology*, *76*(4), 1061–1123. https://doi.org/10.1111/peps.12608

Kokol P, Kokol M, Zagoranski S. (2022). Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress*, *105*(1), 368504211029777. https://doi.org/10.1177/00368504211029777.

Iradukunda, J. (2024). *A hybrid multi-criteria decision-making model for air traffic controller recruitment, case study: Rwanda.* [Master's thesis, University of Turkish Aeronautical Association].

Langr, D., Ploch, J., Rudolf, P., Szabo, S., & Vajdova, I. (2019). Selection process as a key human aspect in air traffic control. *Transport Problems*, *14*(3), 75–84. https://doi.org/10.20858/tp.2019.14.3.7

Mouratille, D., Amadieu, F., & Matton, N. (2022). A meta-analysis on air traffic controllers selection: cognitive and non-cognitive predictors. *Journal of Vocational Behavior*, *138*, 103769. https://doi.org/10.1016/j.jvb.2022.103769

Nabeeh, N. A., Smarandache, F., Abdel-Basset, M., El-Ghareeb, H. A., & Aboelfetouh, A. (2019). An integrated neutrosophic-topsis approach and its application to personnel selection: A new trend in brain processing and analysis. *IEEE access*, 7, 29734–29744. https://doi.org/10.1109/access.2019.2899841

National Transportation Safety Board (NTSB). (2025). Preliminary report on the January 29, 2025 mid-air collision near Washington, D.C. NTSB. Retrieved March 2025, from https://www.ntsb.gov/investigations/Documents/DCA25MA108%20Prelim.pdf.

Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. In *2011 IEEE control and system graduate research colloquium* (pp. 37-42). IEEE. https://doi.org/10.1109/icsgrc.2011.5991826

Özdemir, M., & Sağır Özdemir, M. (2024). *A machine learning approach to replicate interview assessments for selecting air traffic controller candidates.* International Symposium on the Analytic Hierarchy Process. Web Conference, December 13–15. https://doi.org/10.13033/isahp.y2024.002.

Özdemir, M., & Sağır Özdemir, M. (2021). Application of the ANP in the interview phase of air traffic controller candidate selection process. *Aircraft Engineering and Aerospace Technology*, *93*(10), 1616–1623. https://doi.org/10.1108/aeat-08-2020-0166

Özdemir, M., & Sağır Özdemir, M. (2018). The use of the analytic hierarchy process for admission to an air traffic control department. *International Journal of the Analytic Hierarchy Process*, *10*(1), 64–82. https://doi.org/10.13033/ijahp.v10i1.503

Pham, T. N. D., Nguyen, H. D. N., & Tran, T. D. (2025). Analysis and prediction of enrollment decisions of university-admitted candidates. *Neural Computing and Applications*, *37*(32), 26485–26499. https://doi.org/10.1007/s00521-025-11632-5

Pan, W., Li, Y., Jiang, Y., Wang, R., Feng, Y., & Xv, G. (2025). Analysis of controller-caused aviation accidents based on association rule algorithm and Bayesian network. *Applied Sciences*, *15*(17), 9690.

Pecena, Yvonne & Keye, Doris & Conzelmann, Kristin & Grasshoff, Dietrich & Maschke, Peter & Heintz, Alexander & Eißfeldt, Hinnerk. (2013). Predictive validity of a selection procedure for air traffic controller trainees. *Aviation Psychology and Applied Human Factors*, *3*(1), 19–27. https://doi.org/10.1027/2192-0923/a000039.

Salcedo-Sanz, S., Rojo-Álvarez, J. L., Martínez-Ramón, M., & Camps-Valls, G. (2014). Support vector machines in engineering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4*(3), 234–267. https://doi.org/10.1002/widm.1125

Sibarani, I. J. B., & Gunawan, F. E. (2024). The relationships between the binary correlation and accuracy in logistic regression for predicting employee attrition. In *2024 10th International Conference on Computing, Engineering and Design (ICCED)* (pp. 1-7). IEEE. https://doi.org/10.1109/icced64257.2024.10983133

Smith, C., & Bayazit, O. (2024). Transitioning from an institutionally defined candidate search to the Analytic Hierarchy Process: improving faculty selection. *International Journal of the Analytic Hierarchy Process*, *16*(2), 1–19. https://doi.org/10.13033/ijahp.v16i2.1231

Taylan, O., Alidrisi, H., & Kabli, M. (2014). A multi-criteria decision-making approach that combines fuzzy topsis and DEA methodologies. *South African Journal of Industrial Engineering*, *25*(3), 39–56. https://doi.org/10.7166/25-3-667