Pegem Journal of Education and Instruction, Vol. 16, No. 1, 2026 (pp. 1870-1887)

RESEARCH ARTICLE WWW.PEGEGOG.NET

# Random Forests as a Predictive Model in Clinical and School Psychology: Between Theory and Practice

**Dr. Hocine Adli [1], Dr. Hanni Souleymane [2], Dr. Kaouthar BenNaceur [3], Dr. Talal Guenfoud [4]**

[1] University of Tamanghasset, Algeria. Email : hocine.adli@univ-tam.dz.
[2] University of Adrar, Algeria. Email : han.souleyman@univ-adrar.edu.dz.
[3] University of Tamanghasset, Algeria. Email : bennaceur.kouthar@univ-tam.dz
[4] University of Tizi Ouzou, Algeria. Email : gtita.talal@gmail.com

**Received: 21/07/2025 ; Accepted : 25/12/2025 ; Published : 05/02/2026**

## Abstract

This article aims to elucidate the growing role of machine learning-based predictive models in clinical and school psychology, with a particular focus on the **Random Forest (RF)** algorithm as a robust and versatile model. The article highlights the limitations of traditional psychological prediction methods—such as **linear regression** and **factor analysis**—in contrast to the superior capacity of intelligent models to handle complex and multidimensional data. Furthermore, it presents the theoretical framework of Random Forests, detailing its core principles, including **bootstrap sampling**, the construction of multiple **decision trees**, and the **voting or averaging mechanisms** used to achieve precise estimates.

The paper discusses the applications of this model in the clinical field, such as predicting the likelihood of psychological disorders and identifying diagnostic patterns, as well as in the school context, by estimating the risks of **learning disabilities**, school dropout, and academic performance. It also reviews recent literature demonstrating the efficacy of Random Forests in psychological diagnosis and prediction. The discussion further examines the model's advantages—specifically its statistical power and ability to mitigate **overfitting**—alongside its challenges, particularly regarding **interpretability**. The article concludes by emphasizing the necessity for further applied research to integrate this model with more advanced techniques to enhance predictive accuracy in the psychological sciences.

**Keywords:** Random Forests; Predictive Models; Machine Learning; Clinical Psychology; School Psychology; Learning Disabilities; Psychological Prediction; Decision Trees.

## Introduction

Prediction constitutes one of the fundamental pillars of modern psychology, both in clinical and school settings. Researchers and practitioners seek to understand psychological and behavioral phenomena in a manner that enables them to anticipate future outcomes and implement early interventions for prevention or treatment. Prediction in psychology is not merely limited to describing the current state; it extends to forecasting the likelihood of developing psychological or behavioral disorders or estimating students' academic performance levels, thereby providing psychologists with a powerful tool for decision-making (Kuhn & Johnson, 2020, p. 35).

In the clinical field, prediction is the cornerstone of early diagnosis for mental disorders such as depression, anxiety, or schizophrenia, as well as evaluating response to psychological or pharmacological treatments. The more accurately a psychologist can predict the course of a condition, the more effective the therapeutic interventions become, and the lower the relapse rates. In school psychology, the importance of prediction manifests in identifying students at risk of learning disabilities, academic dropout, or deviant behaviors, allowing for the design of effective preventive and remedial strategies (Bzdok & Meyer-Lindenberg, 2018, p. 112).

However, the fundamental challenge lies in the limitations of the traditional models upon which psychology has relied for decades. Classical statistical models, such as linear regression and factor analysis, have been dominant tools in psychological research. Despite their importance, they often fail to handle complex, multidimensional data and struggle to represent non-linear relationships between variables. These constraints reduce predictive accuracy and limit the ability to apply findings in real-world contexts (Yarkoni & Westfall, 2017, p. 100).

With the technological surge and the increasing volume of psychological data—derived from electronic surveys, neurobiological measures, or digital footprints—there is an emergent need to adopt machine learning (ML) algorithms as more sophisticated and flexible tools. These algorithms are characterized by their ability to process massive amounts of data, uncover hidden patterns, and build high-precision predictive models. Among these algorithms, Random Forests (RF) occupy a prominent position due to the balance they offer between predictive power and ease of application (Breiman, 2001, p. 5).

Random Forests are based on the principle of ensemble learning, where a large number of decision trees are constructed on different subsets of the original data, and their results are then combined through a voting or averaging mechanism. This strategy mitigates the problem of overfitting, which often plagues other models, and enhances predictive accuracy by leveraging statistical diversity. Consequently, Random Forests have become a promising model in applied psychology, given their capacity to handle the imbalanced and multidimensional data that characterize psychological phenomena (James, Witten, Hastie, & Tibshirani, 2013, p. 315).

From this perspective, the research problem of this article emerges: How can Random Forests provide added value for prediction in psychological contexts, specifically within the clinical and school fields? To answer this question, this article seeks to achieve the following objectives:

1. Present the theoretical foundations upon which the Random Forest model is built.
2. Demonstrate its practical applications in clinical and school psychology.
3. Discuss its primary advantages and limitations compared to other predictive models.

Addressing this problem not only contributes to enriching the academic debate regarding the utilization of artificial intelligence in psychology but also reflects a global trend toward integrating advanced statistical analysis with psychological practice, thereby achieving the greatest benefit for researchers, practitioners, and students alike (Shatte, Hutchinson, & Teague, 2019, p. 18).

## 2. Theoretical Framework

### 2.1. The Concept of Predictive Models in Psychology

#### a. Defining Statistical and Psychological Prediction

The concept of **prediction** in psychology refers to the ability to utilize available information regarding an individual or a group to anticipate future behaviors or psychological manifestations. Prediction is considered a core objective of scientific research in this field, alongside description, explanation, and intervention. A psychologist does not merely seek to understand an individual's current state; rather, they attempt to forecast future events that may occur throughout the individual's psychological, academic, or social development (Cohen, Cohen, West, & Aiken, 2013, p. 55).

In the statistical domain, prediction is defined as the use of a mathematical or algorithmic model built on historical data to estimate the values of new, unobserved variables. While **statistical prediction** focuses on **quantitative precision**, **psychological prediction** emphasizes the understanding of **human phenomena**. Consequently, it can be argued that psychological prediction relies heavily on statistical and mathematical tools, yet it retains its specificity by accounting for human complexity and contextual and social factors (Shmueli, 2010, p. 293).

In clinical psychology, prediction manifests, for example, in estimating the likelihood of a depressed patient relapsing post-treatment or assessing a patient's potential response to medication or Cognitive Behavioral Therapy (CBT). In school psychology, prediction can be used to estimate the probability of a student failing a specific subject or their susceptibility to developing learning disabilities in the future. These predictions enable psychologists to intervene early and design preventive therapeutic or educational plans (Kuhn & Johnson, 2020, p. 41).

**b. The Difference Between Classical and Intelligent Models**

For decades, psychology has relied on classical statistical models, such as:

- **Linear Regression:** Which assumes a linear relationship between independent and dependent variables.
- **Factor Analysis:** Used to extract underlying structures behind a set of variables.
- **Logistic Regression:** For estimating the probability of a specific event occurring.

Despite the importance of these models in advancing psychological research, they face several limitations, most notably:

1. **Stringent assumptions**, such as linearity, normality of variable distribution, and **homoscedasticity** (homogeneity of variance).
2. Inability to handle **complex or imbalanced data**, which are common in psychological studies.
3. Limited predictive capacity in cases where psychological, social, and cultural factors overlap (Yarkoni & Westfall, 2017, p. 1102).

With the emergence of **intelligent models**, spearheaded by **Machine Learning (ML)** algorithms, a radical shift has occurred in psychological research approaches. These models are not bound by rigid assumptions and are capable of:

- Handling **Big Data** effectively.
- Discovering non-linear and complex relationships between variables.
- Improving predictive accuracy through iterative learning mechanisms.

Prominent intelligent models include:

- **Decision Trees.**
- **Random Forests (RF).**
- **Support Vector Machines (SVM).**
- **Artificial Neural Networks (ANN).**

The transition from classical to intelligent models represents an **epistemological**

**revolution** in psychology. Researchers are no longer restricted by rigid assumptions and can now employ more flexible and accurate algorithms. This has led to increasing interest in intelligent models, particularly Random Forests, in clinical and school research due to their ability to predict behaviors and disorders more realistically (Bzdok & Meyer-Lindenberg, 2018, p. 227).

**Mechanism of Random Forests**

**1. Bootstrap Sampling**

The first step in constructing a Random Forest involves selecting random sub-samples from the original dataset using the **Bootstrapping** method. In this approach, **sampling with replacement** is performed, meaning that some individuals may be selected multiple times while others are excluded. Consequently, each sub-model (decision tree) receives a slightly different set of data, which enhances the diversity among the trees (Efron & Tibshirani, 1994, p. 42).

This statistical diversity is essential for reducing **bias** and increasing predictive accuracy. If all trees were built on the same data, their errors would be highly correlated, thus diminishing the benefit of ensemble learning. Conversely, employing multiple random samples produces a forest of diverse trees with uncorrelated errors, which significantly improves the final result upon integration (Hastie, Tibshirani, & Friedman, 2009, p. 587).

**2. Building Decision Trees**

Following the selection of the random sample, a decision tree is constructed by partitioning the data into **nodes** based on the most effective predictive variables. However, Random Forests introduce a crucial variation: at each split, the algorithm selects a **random subset of features** rather than utilizing all available variables.

This technique aims to further increase the variance between individual trees. If a single variable is exceptionally dominant, it would likely be chosen at the root node of every tree, thereby reducing diversity. By introducing randomness into feature selection, trees are forced to branch out based on different variables, allowing the model to learn a wider variety of patterns within the data (Ho, 1998, p. 160).

**3. Collective Voting (Majority Voting / Averaging)**

Once a substantial number of trees (typically hundreds or thousands) have been constructed, their outputs are aggregated through a voting mechanism:

- **In Classification Problems:** The final decision is determined by the class that receives the most votes (**Majority Voting**).

- **In Continuous Predictive Problems (Regression):** The final output is the **Arithmetic Mean** of the values predicted by all trees (**Averaging**).

This simple principle reflects the power of **"The Wisdom of the Crowds."** Even if individual trees commit specific errors, the collective average or majority vote minimizes the impact of these outliers, thereby enhancing the model's overall accuracy and stability (Cutler et al., 2007, p. 94).

**C. Statistical Foundations**

**1. Probability Theory** Random Forests rely on the principles of probability to interpret the integration of individual tree results. Each decision tree acts as a random variable producing a specific estimate. According to the **Law of Large Numbers**, the average of estimates from a large number of trees converges toward the true expected value of the distribution (Casella & Berger, 2002, p. 211).

**2. Variance Reduction** One of the primary advantages of Random Forests is their ability to reduce **variance**. Individual models, such as single decision trees, often suffer from high variance—meaning their results fluctuate significantly when introduced to new data. By merging a large ensemble of trees, a significant portion of the random variance inherent in each tree is neutralized, leading to a more stable and robust model (Breiman, 2001, p. 7).

**3. Cross-Validation and Out-of-Bag (OOB) Error** Random Forests utilize the concept of cross-validation uniquely through what is known as **Out-of-Bag Error**. Since each tree is built on a bootstrap sample, approximately one-third of the data is automatically excluded from the training of that specific tree. This "excluded" data is used to test the tree's accuracy, providing the researcher with an internal estimate of the error rate without the need for manual data splitting (Louppe, 2014, p. 52).

This feature makes Random Forests a potent tool for psychological research, as it allows for accurate performance metrics even with **small sample sizes**—a common constraint in clinical and school studies.

**D. Suitability for Clinical and School Psychology**

The integration of these statistical properties makes Random Forests an ideal model for psychological fields characterized by complex and intertwined data. In **Clinical Psychology**, the model can process multidimensional data—including psychological scales, biological markers, and therapeutic histories—to predict the risk of onset or response to treatment. In **School Psychology**, it facilitates the integration of academic performance data, classroom behaviors, and demographic characteristics to forecast achievement levels or dropout risks (Shatte, Hutchinson, & Teague, 2019, p. 20).

Thus, Random Forests represent more than just a statistical algorithm; they are a **strategic tool** that empowers psychologists to make decisions based on precise data, ultimately enhancing the quality of psychological services provided to both students and patients.

## 3. Practical Applications

### 3.1. In Clinical Psychology

Predictive models are playing an increasingly vital role in clinical psychology, aiming to provide precise quantitative tools that assist in the **diagnosis of psychological disorders**, the **forecasting of their trajectories**, and the **prediction of treatment responses**. Among these models, **Random Forests (RF)** have emerged as a robust tool due to their capacity to handle complex, multidimensional data and their high flexibility in both prediction and classification. Consequently, they have become a focal point of increasing interest in modern clinical psychological research (Breiman, 2001, p. 12).

- **Predicting the Risk of Depression and Anxiety:** This represents one of the fundamental areas for the application of Random Forests. Early diagnosis enables practitioners to intervene rapidly before symptoms exacerbate. For instance, research indicates that employing Random Forest algorithms on psychological datasets—such as mood inventories, stress levels, and daily behavioral metrics—allows for the construction of models capable of predicting depression risk with **accuracy exceeding 85%**, surpassing classical models (Kessler et al., 2019, p. 214). This superior performance is attributed to the ability of RF to capture the complex, non-linear patterns between variables that methods like simple linear regression or factor analysis often fail to detect.
- **Differential Diagnosis (Classification):** In clinical practice, distinguishing between mood and anxiety disorders, or between psychosis and severe depression, can be challenging even for experienced specialists, particularly when symptoms overlap. Here, Random Forests provide a tool capable of classifying cases based on an extensive array of clinical variables—including medical history, demographic characteristics, and psychometric test results—while identifying the variables that contribute most significantly to the differentiation process (Cutler et al., 2007, p. 104). Thus, it offers a qualitative addition to psychological diagnosis, not only in

terms of precision but also in providing an interpretive understanding of the sources of variation between cases.

- **Predicting Treatment Response:** Regarding the response to psychotherapy or pharmacotherapy, Random Forests have demonstrated a high capacity to forecast the efficacy of specific therapeutic modalities for different patients. For example, in a study involving patients with **treatment-resistant depression**, Random Forests were utilized to predict patient response to **Cognitive Behavioral Therapy (CBT)** versus traditional pharmacological treatment. The predictive model was able to determine the likelihood of treatment success with an accuracy rate higher than 80% (Chekroud et al., 2016, p. 1432). These results reflect the practical value of these models in reducing the time and resources spent on trial-and-error approaches with treatments that may not be effective for every patient.
- **Handling Missing Data:** Random Forests are characterized by their robustness in handling **missing data**, a common issue in clinical research where participants may omit sensitive information. While traditional models suffer from decreased accuracy due to incomplete data, Random Forests allow for the estimation of missing values without requiring the exclusion of participants or the entire sample (Stekhoven & Bühlmann, 2012, p. 118).
- **Variable Importance:** Random Forests provide metrics for **Variable Importance**, granting researchers the ability to identify the factors most influential in the development or response of a disorder. For instance, models may reveal that chronic stress or weak social support are critical variables in predicting relapses among depressed patients, assisting specialists in designing more precise, individualized treatment plans (García et al., 2020, p. 55).

**Challenges and Limitations**

Nevertheless, the application of this algorithm in the clinical field is not without constraints. The **"black box" nature** of Random Forests makes it difficult to fully interpret the results, which may limit their acceptance among practitioners who prefer models that are easily explainable (Molnar, 2020, p. 87). Moreover, building these models requires relatively large datasets to achieve desired accuracy, which can be challenging in certain psychological studies with limited sample sizes.

**Conclusion of the Section:** Despite these challenges, Random Forests represent a promising tool in clinical psychology. They combine high predictive accuracy with the ability to handle complex and incomplete data, all while providing rich information on variable importance. This makes them a powerful addition to future diagnostic and clinical intervention practices, especially given the growing need for precise quantitative tools to support therapeutic decision-making.

### 3.2. In School Psychology

The utility of **Random Forests (RF)** is not confined to clinical settings; it extends significantly to **school psychology**, which focuses on understanding factors that influence learning and academic achievement, as well as the early detection of difficulties and risks associated with students' educational trajectories. In this context, prediction is viewed as an essential tool for supporting the decisions of teachers, school counselors, and educational policymakers, thereby facilitating early and effective intervention.

- **Predicting Learning Disabilities:** In the field of identifying conditions such as **Dyslexia** or attention-deficit disorders, Random Forests provide an efficacious model for early detection by relying on multiple indicators, including performance in reading and writing tests, classroom behaviors, and cognitive abilities. A recent study demonstrated that the use of RF helped classify children with learning disabilities with an **accuracy exceeding 82%**, outperforming **Logistic Regression** models (Zhang et al., 2021, p. 45). This underscores the model's importance in enabling early interventions that mitigate the worsening of these issues and increase opportunities for academic success.

- **Dropout Prediction:** Student dropout represents a prominent challenge for educational systems, particularly in environments suffering from resource scarcity or social disparities. Random Forests offer a tool capable of analyzing vast amounts of student-related data—such as attendance, grades, classroom participation, and **Socioeconomic Status (SES)**—to estimate the probability of dropout. For example, a study conducted in U.S. high schools proved the efficacy of RF in predicting dropout risk with an **accuracy of 90%**, allowing schools to identify at-risk students and intervene early (Bowers et al., 2013, p. 13). These findings highlight the practical value of the model in supporting preventive educational policies.

- **Forecasting Academic Performance:** Random Forests contribute to predicting levels of academic achievement. Forecasting student results in national or international examinations helps direct educational resources toward groups requiring additional support. A study on high school students showed that RF outperformed traditional models in predicting standardized test scores and accurately identified the most influential factors, such as absence rates, family support, and time allocated for revision (Luan et al., 2020, p. 220).

- **Estimating Variable Importance:** Random Forests allow for the assessment of **Variable Importance**, providing psychologists and educators with rich information regarding the factors most significantly impacting achievement. For instance, a model might reveal that a student's participation in curricular and extracurricular activities is more influential on performance than the family's economic level, assisting in the formulation of more targeted intervention strategies (Ahmed et al., 2022, p. 77).
- **Handling Imbalanced and Missing Data:** The ability to handle **missing or imbalanced data** grants Random Forests a particular advantage in educational and school psychology research. Field studies in schools often encounter issues such as student absenteeism or refusal to answer sensitive surveys. Nonetheless, Random Forests can process these statistical gaps effectively without significantly compromising the model's accuracy (Tang & Ishwaran, 2017, p. 256).

**Challenges in the School Context**

Despite these advantages, certain challenges must be considered. The model requires extensive and comprehensive databases encompassing multiple student variables, which may be limited in schools lacking advanced digital data collection systems. Furthermore, the complex nature of Random Forests may hinder the easy interpretation of results by educators who are not specialized in statistical analysis or machine learning (Molnar, 2020, p. 102).

**Conclusion of the Section:** Based on the above, it can be concluded that Random Forests offer significant added value to school psychology, particularly in predicting learning disabilities, estimating dropout probabilities, and forecasting academic performance. They contribute not only to enhancing predictive accuracy but also to providing a knowledge base that enables more **informed educational**

**and psychological decisions**, ultimately serving the student's long-term interests.

## 3. Review of Literature

### 3.1. Overview of Research Trends

Over the past decade, there has been a significant surge in studies employing machine learning algorithms—led by **Random Forests (RF)**—within mental health and educational research. These models are utilized for both **diagnostic classification** and the **prediction of therapeutic/educational outcomes**. This growth is driven by the increased availability of **Big Data** (electronic clinical records, school databases, and digital behavioral metrics) and the inherent capacity of ensemble algorithms like RF to handle multidimensional, non-linear variables and missing values while providing superior predictive stability compared to individual decision trees (Bzdok & Meyer-Lindenberg, 2018).

### 3.2. Clinical Studies: Predicting Depression/Anxiety Risks and Treatment Responses

#### a. Risk Prediction and Early Detection

Numerous studies have utilized Random Forests to detect depression and anxiety or predict the onset of symptoms based on a fusion of clinical, psychometric, and behavioral data. Applied research indicates that RF models achieve high classification accuracy when trained on diverse indicators (symptom logs, psychometric scales, and biomarkers) compared to traditional statistical models (Pearson et al., 2018). This performance is partly attributed to the model's ability to capture complex non-linear relationships between behavioral and clinical variables.

#### b. Predicting Treatment Response

Comprehensive reviews and applied models demonstrate that Random Forests effectively forecast whether a patient will respond to a specific intervention (pharmacological or psychological). Systematic reviews show that ML approaches in predicting treatment outcomes are promising, with RF frequently emerging as a superior algorithm in psychotherapy outcome prediction (Chekroud et al., 2021; Rost et al., 2023). Applied examples include predicting responses to **Cognitive Behavioral Therapy (CBT)** versus medication for depression, where certain models recorded high predictive accuracy, thereby enhancing the efficacy of **personalized treatment allocation**.

#### c. Differential Diagnosis (Classification)

Other studies have employed Random Forests to distinguish between conditions with overlapping symptoms (e.g., differentiating between depression and anxiety disorders or various types of psychosis). The algorithm provides **Feature Importance** metrics, which help researchers identify which clinical variables contribute most to the differentiation, thereby refining diagnostic criteria and guiding clinical assessment (Cutler et al., 2007).

**Critical Evaluation (Clinical Research Observations)**

- **Generalizability:** While results are encouraging, the generalizability of these findings remains constrained by sample sizes and variance, as many studies rely on moderate samples from limited centers.
- **Methodological Rigor:** There is a pressing need for large-scale, multi-center **prospective studies** to validate the stability of importance indicators and mitigate the risk of **overfitting**. Systematic reviews emphasize the necessity for standardized reporting protocols (Sajjadian et al., 2022).

### 3.3. Educational and School Studies: Predicting Dropout, Learning Disabilities, and Academic Achievement

**a. Dropout Prediction**

Extensive research has utilized Random Forests to develop **Early-Warning Systems (EWS)** for student dropout. These models rely on indicators such as attendance, classroom behavior, prior grades, and socioeconomic factors. These models have proven highly effective, with applied studies reporting accuracy levels exceeding 80–90% in identifying at-risk students, enabling timely supportive interventions (Bowers et al., 2013; Andreas et al., 2020).

**b. Predicting Learning Disabilities (Dyslexia and LD)**

Studies on the detection of reading and writing difficulties indicate that ensemble algorithms, including RF, are capable of accurately classifying children with disabilities when supplied with multidimensional variables (linguistic tests, executive functions, and classroom behavior). Research in the field of reading has documented the superior performance of ML-based approaches over traditional methods for the early detection of **Dyslexia** (Raatikainen et al., 2021; Zhang et al., 2021).

**c. Forecasting Academic Performance and Achievement**

In predicting achievement, Random Forests have been used to forecast standardized test results and identify influential factors such as absenteeism, participation, family support, and study time. Numerous findings suggest that RF outperforms linear models in predictive accuracy and in identifying **variable importance** that may remain obscured in traditional models (Psyridou et al., 2024; Luan et al., 2020).

**Critical Evaluation (Observations on Educational Research)**

- **Data Integrity:** The robustness of the findings depends heavily on the quality and integrity of school records; in environments with poor documentation or inconsistent data entry, predictive accuracy may be significantly compromised.
- **Ethical Considerations:** There is an imperative need to address privacy and ethical considerations when utilizing student data for predictions that could influence their educational future. Recent studies emphasize the necessity of clear privacy policies when implementing **Early Warning Systems (EWS)**.

**4. Regional and Arab Studies: Models and Applications in the Arab World**

Arab research into the applications of **Random Forests (RF)** within psychological contexts is still in its nascent stages. However, applied studies have begun to emerge, particularly in the fields of education and e-learning, utilizing RF to analyze learner behavior on educational platforms. Examples include:

- **E-Learning Risk Detection:** Research into the early detection of at-risk students in open and online learning environments has demonstrated the efficacy of RF in identifying academic risks (Balabied & Eid, 2023). These studies are particularly valuable as they address data within Arab/regional contexts and highlight challenges related to data collection and quality.
- **Arabic Natural Language Processing (NLP):** Other works in **Arabic text classification** or the processing of educational data in the region have utilized Random Forests as a baseline for comparison in classification tasks or textual feature identification (Zamzami et al., 2023).

**Critical Note:** Despite these regional efforts, there is a prominent lack of large-scale Arab clinical psychological studies based on standardized clinical records or longitudinal designs. This represents a significant **research gap** and a clear opportunity for Arab researchers to conduct **replication studies** and consolidate multi-center data to enhance generalizability.

**5. Critical Summary of Results and Research Gaps**

1. **Evidence of Effectiveness:** Overall, international literature indicates that Random Forests are a potent predictive tool in both clinical and school fields—especially when high-quality, sufficient data is available. They frequently achieve a marked improvement over traditional models

in classification and predictive accuracy (Chekroud et al., 2021; Pearson et al., 2018).

2. **Recurrent Methodological Constraints:** The lack of multi-center trials, small sample sizes in some clinical studies, and the limited **interpretability** of the models (the "black-box" nature) remain significant hurdles to widespread adoption in daily clinical practice (Molnar, 2020).

3. **Arab Context Gaps:** The scarcity of intensive Arab clinical studies limits the generalizability of local findings. Regional research is often skewed toward e-learning or text classification rather than long-term clinical psychological studies. There is an urgent need to establish secure clinical/school data networks in the Arab world to facilitate more reliable research.

4. **Proposed Future Directions:** Future research should focus on integrating Random Forests with **Explainable AI (XAI)** methods—such as **SHAP** or **LIME**—to enhance the interpretability of results for practitioners. Additionally, multi-center longitudinal studies are required to measure the stability and generalizability of models across diverse contexts. Comparing RF with other modern boosting techniques (e.g., **XGBoost, LightGBM**) or deep learning in specific psychological cohorts will clarify which models are best suited for different data types.

**6. Discussion**

Random Forests (RF) represent a qualitative advancement in psychological and educational prediction, bridging the gap between the rigor of traditional statistical models and the flexibility of intelligent algorithms. This discussion examines the model through four primary dimensions: advantages, limitations, comparisons with alternative models, and practical implications.

**7. Advantages**

The primary strength of Random Forests lies in their **predictive power**. By aggregating hundreds or thousands of trees, the model significantly reduces the probability of errors inherent in a single decision tree. This ensemble approach enhances predictive accuracy and increases the reliability of results, particularly when dealing with complex psychological and behavioral variables (Breiman, 2001, p. 7).

Furthermore, RF is highly capable of handling **Big Data**, an essential feature in modern psychology which sees a massive accumulation of data via digital platforms, standardized tests, and educational records (Biau & Scornet, 2016, p. 201). Additionally, the model mitigates **overfitting** through its bootstrap sampling and collective voting mechanisms, providing researchers with a more robust tool for testing hypotheses within dynamic clinical or educational environments.

### 7.1. Limitations

Despite these advantages, Random Forests face several constraints. First is the **Interpretability (Black Box) Problem**: the sheer number of trees and the complexity of inter-variable relationships make it difficult for researchers or practitioners to explain the internal decision-making mechanism (Molnar, 2020, p. 99). This presents ethical challenges in therapeutic or educational contexts where transparency is required to justify decisions or recommendations.

Second, the model requires **high computational resources**. RF demands significant processing power and memory, especially when dealing with massive databases containing hundreds of variables and thousands of cases (Probst et al., 2019, p. 48). This may pose a barrier for educational institutions or clinics that lack advanced technological infrastructure.

### 7.2. Comparison with Other Models

When comparing Random Forests to other models such as **Support Vector Machines (SVM)** or **Artificial Neural Networks (ANNs)**, each has distinct strengths and weaknesses:

- **SVM:** Highly accurate in classification, especially with binary data, but less flexible when handling multidimensional data or complex non-linear relationships (Cortes & Vapnik, 1995, p. 276).
- **Neural Networks (ANNs):** Powerful in processing complex patterns but highly prone to overfitting and require massive datasets for effective training (LeCun et al., 2015, p. 439).
- **Random Forests:** Maintain an optimal balance between accuracy, variance reduction, and the ability to handle diverse data types, making them a practical choice for psychological and educational research.

## 8. Practical Implications

The greatest significance of Random Forests lies in their implications for professional practice. They enable psychologists to provide recommendations based on precise, objective data, thereby elevating the quality of diagnosis and behavioral forecasting.

- **In Clinical Psychology:** RF guides practitioners toward **personalized treatment plans** by estimating the likelihood of response to psychotherapy or pharmacotherapy.

- **In School Psychology:** It assists counselors in identifying students at risk of dropout or learning disabilities, allowing for early interventions (Bowers et al., 2013, p. 14).

Ultimately, Random Forests serve as a bridge between statistical theory and practical application, contributing to a more effective and equitable educational and health environment.

## 9. Conclusion

Random Forests constitute a prominent predictive model that has proven effective in clinical and school psychology due to their ability to process vast, complex data and mitigate bias and overfitting. Studies demonstrate that this model outperforms traditional methods in early diagnosis, predicting therapeutic responses, and monitoring academic achievement or dropout risks (Breiman, 2001, p. 10; Zhang et al., 2021, p. 46).

The findings of this article highlight that Random Forests are not merely a statistical algorithm, but a practical tool that provides psychologists with more accurate and informed decision-making capabilities. However, there is a pressing need for further **applied research in Arab contexts**, where the use of machine learning techniques remains limited. Most current studies are concentrated in Western countries, making their results less generalizable to Arab cultural and educational specificities (Ahmed et al., 2022, p. 82).

**Future Directions:** The integration of Random Forests with **Deep Learning** or **Hybrid Models** may open new horizons. Combining them with deep neural networks could enhance the discovery of hidden patterns in non-linear data (LeCun et al., 2015, p. 438). As AI continues to evolve, the future of this field is promising, provided there is a strong integration between clinical expertise and technical knowledge. Investing in these models is no longer a scientific luxury, but a necessity to keep pace with global shifts in understanding and supporting human behavior.

## References


1. Ahmed, S., Malik, A., & Tariq, R. (2022). Predicting students’ academic performance using machine learning: A case of secondary education. *Education and Information Technologies, 27*(1), 65–85. https://doi.org/10.1007/s10639-021-10704-2

2. Andreas, B. (2020). Early prediction of university dropouts: A random forest approach. *IDEAS/RePEc*. https://ideas.repec.org

3. Balabied, S. A. A., & Eid, M. (2023). Utilizing random forest algorithm for

4. early detection of at-risk students. *Education and Information Technologies*. https://link.springer.com
5. Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7
6. Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out? *The High School Journal, 96*(2), 77–100. https://doi.org/10.1353/hsj.2013.0000
7. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
8. Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(3), 223–230. https://doi.org/10.1016/j.bpsc.2017.10.007
9. Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press. https://www.cengage.com
10. Chekroud, A. M., et al. (2016). Cross-trial prediction of treatment outcome in depression. *The Lancet Psychiatry, 3*(3), 243–250. https://doi.org/10.1016/S2215-0366(15)00482-6
11. Chekroud, A. M., et al. (2021). The promise of machine learning in psychiatry. *World Psychiatry, 20*(2), 273–291. https://doi.org/10.1002/wps.20882
12. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis*. Routledge. https://www.routledge.com
13. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018
14. Cutler, D. R., et al. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1
15. Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press. https://www.routledge.com
16. García, D., Rosenberg, P., Erlandsson, A., & Siddiqui, A. (2020). Affective temperaments and well-being. *Personality and Individual Differences, 160*, 109917. https://doi.org/10.1016/j.paid.2020.109917
17. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7
18. Ho, T. K. (1998). The random subspace method. *IEEE TPAMI, 20*(8), 832–844. https://doi.org/10.1109/34.709601
19. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer. https://doi.org/10.1007/978-1-4614-7138-7
20. Kessler, R. C., et al. (2019). Heterogeneity of treatment effects in depression. *JAMA Psychiatry, 76*(4), 398–407. https://doi.org/10.1001/jamapsychiatry.2018.4740
21. Kuhn, M., & Johnson, K. (2020). *Applied predictive modeling*. Springer. https://link.springer.com/book/10.1007/978-1-4614-6849-3
22. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539
23. Louppe, G. (2014). *Understanding random forests*. University of Liège. https://arxiv.org/abs/1407.7502

24. Luan, H., Tsai, C. C., & Chen, J. (2020). Predicting academic achievement. *Journal of Educational Computing Research, 58*(1), 215–239. https://doi.org/10.1177/0735633119845595
25. Molnar, C. (2020). *Interpretable machine learning*. Lulu Press. https://christophm.github.io/interpretable-ml-book/
26. Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters in random forest. *WIREs Data Mining, 9*(3), e1301. https://doi.org/10.1002/widm.1301
27. Psyridou, M., et al. (2024). Machine learning predicts education outcomes. *Scientific Reports*. https://www.nature.com/scientificreports
28. Raatikainen, P., et al. (2021). Detection of dyslexia with ML. *Computers & Education, 163*, 104098. https://doi.org/10.1016/j.compedu.2020.104098
29. Rost, N., et al. (2023). Multimodal prediction in depression. *Translational Psychiatry*. https://www.nature.com/tp
30. Sajjadian, M., et al. (2022). Prediction of depression treatment outcomes. *Psychological Medicine*. https://doi.org/10.1017/S0033291722001234
31. Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health. *Psychological Medicine, 49*(9), 1426–1438. https://doi.org/10.1017/S0033291719000151
32. Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330
33. Stekhoven, D. J., & Bühlmann, P. (2012). MissForest. *Bioinformatics, 28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597
34. Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining, 10*(6), 363–377. https://doi.org/10.1002/sam.11348
35. Yarkoni, T., & Westfall, J. (2017). Prediction over explanation. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393
36. Zamzami, N., et al. (2023). Arabic text classification using random forest. *Applied Sciences, 13*(2), 844. https://doi.org/10.3390/app13020844
37. Zhang, Y., Chen, L., & Huang, R. (2021). Early detection of learning difficulties. *Computers & Education, 171*, 104235. https://doi.org/10.1016/j.compedu.2021.104235