AI-Based Analysis of Infant Motor Development and Age Prediction

Po Feng-Chen (Leader), Sun Hao-Wen, Wang Chun-Yi

1 Abstract

This study investigates the relationship between infants' motor development and age using data analysis and machine learning techniques. Analyzing 33,474 questionnaire responses from the SRDA Child Development Database, we found a strong correlation between motor proficiency and age. Various predictive models were developed, including regression and classification approaches, to estimate infants' age based on their motor skill levels. The models, particularly those using SVM and neural networks, achieved high accuracy. Motivated by the need for early developmental assessment amid declining birthrates, the study aims to provide parents and educators with a scientific tool for monitoring growth and identifying potential. The research highlights the potential of data-driven approaches in supporting early childhood education and personalized developmental guidance.

2 Scientific Methods and Tools

- **Data Processing:** Used Python tools (Pandas, NumPy) for filtering, classification, and aggregation.
- **Model Training:** 80:20 train-test split.
- **Selected Features:** 44 questionnaire items related to motor proficiency.
- **Machine Learning Models:** Random Forest (100 estimators), XGBoost (300 estimators), Support Vector Regression (SVR), Neural Networks.
- **Visualization:** Used Matplotlib to generate charts reflecting trends and results.

3 Result Analysis

3.1 Proficiency Trends

We calculated the average total proficiency score by age group. The results clearly show that total proficiency increases with age.

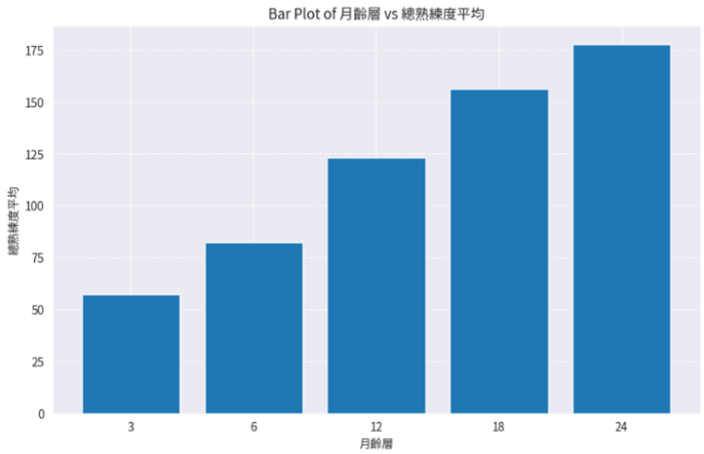

Figure 1: Average Total Proficiency by Age Group

3.2 Machine Learning Models

All models used the 80:20 train-test split. Modeling was structured in two tasks: classification and regression.

3.2.1 Classification: Age Groups

Age groups were divided into five classes: 3, 6, 12, 18, and 24 months (with a ±1-month buffer for classification).

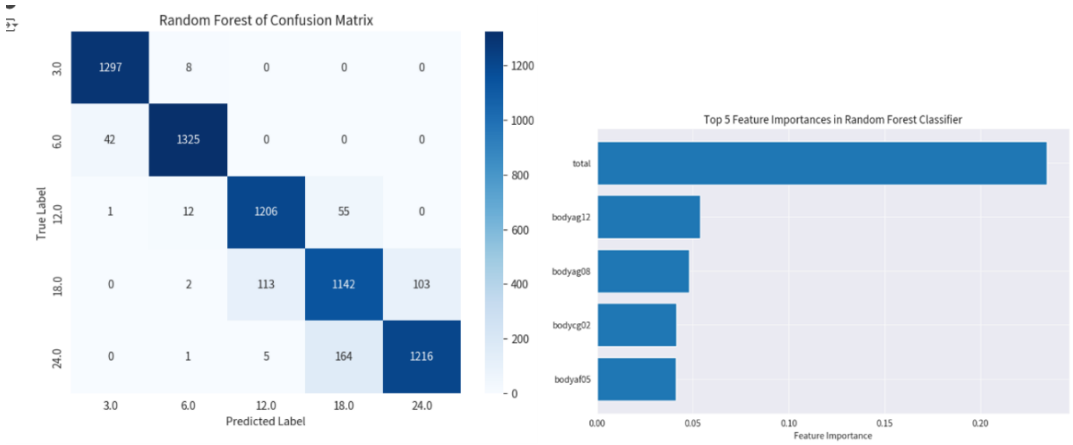

Figure 2: Random Forest Confusion Matrix (Left) and Top 5 Motor Skills Contributing to Age Prediction (Right)

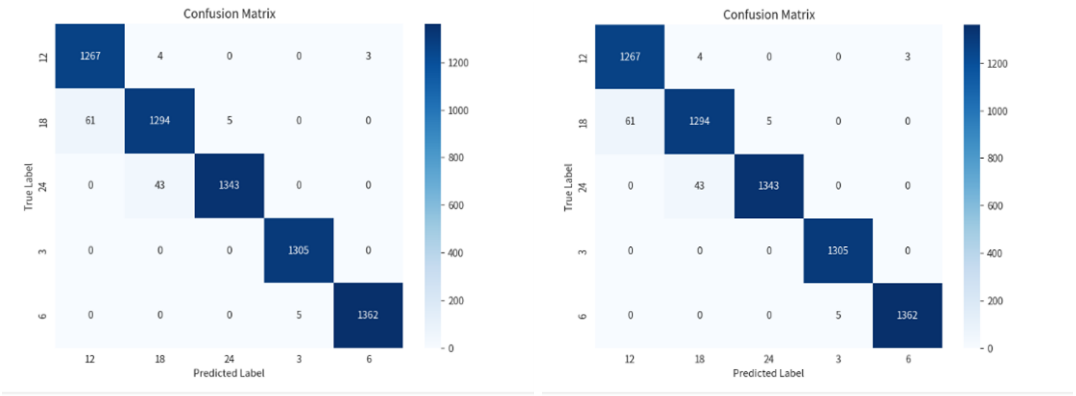

Figure 3: SVM Confusion Matrix (Left) and Multiclass Neural Network Confusion Matrix (Right)

- Random Forest: Accuracy = 92%
- SVM: Accuracy = 97%
- Multiclass Neural Network: Accuracy = 98%
- Random Forest provides feature importance analysis, revealing which movements most influence age prediction.

3.2.2 Regression: Predicting Actual Age

Predicted actual age in months. Accuracy assessed with an error margin of ±1 month (5% of age range).

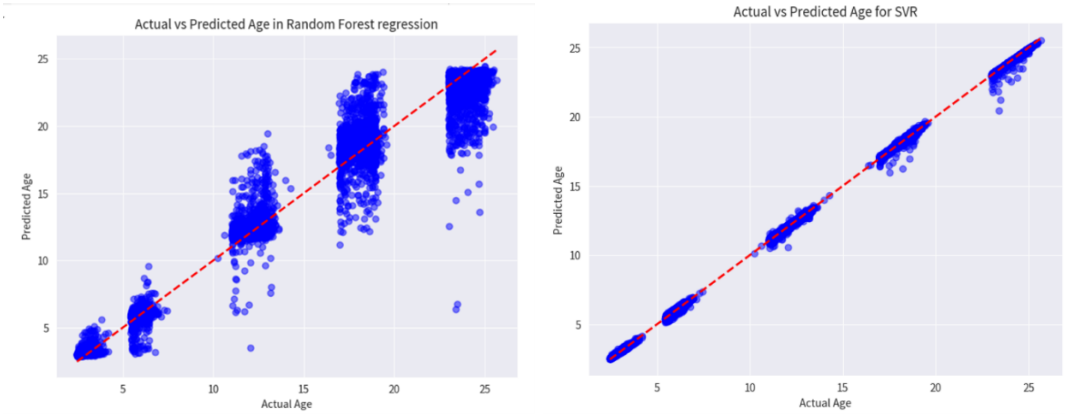

Figure 4: Predicted vs. Actual Ages: Random Forest (Left) and SVM (Right)

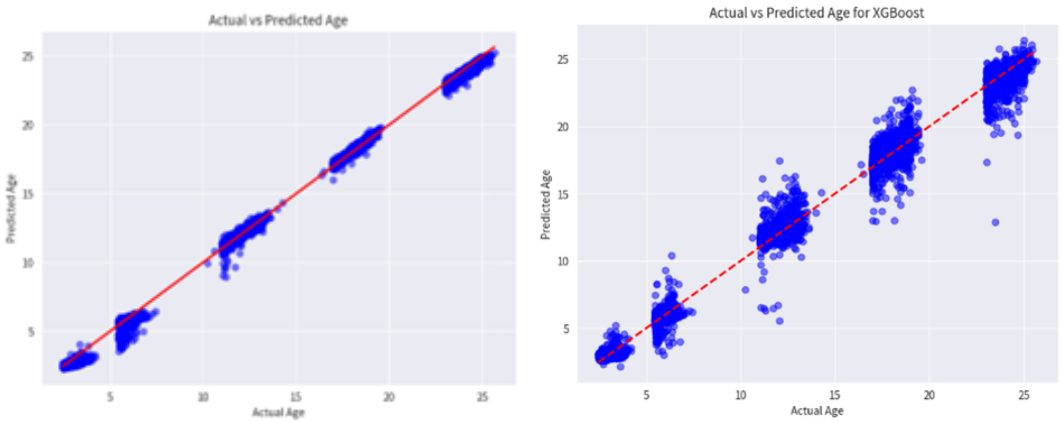

Figure 5: Predicted vs. Actual Ages: Neural Network (Left) and XGBoost (Right)

- Random Forest: Accuracy = 75.12%
- SVM: Accuracy = 99.76%
- Neural Network: Accuracy = 99.09%
- XGBoost: Accuracy = 89.67%
- Random Forest showed more outliers, indicating lower performance.

4 Discussion and Inference

Our findings demonstrate a strong correlation between age and motor performance. Whether through regression or classification, our models achieved impressive results. The SVM model, with a 99.76% accuracy, showed outstanding predictive power.

That said, our current model only considers motor-related variables, which may explain its high accuracy. For future work, we plan to incorporate more diverse variables to increase generalizability and explore potential talents in different developmental areas. We hope the model remains robust even as variable scope expands, proving its practical value.