Deep neural network-based analysis of student interest and sports preference

M. Riyaz Mohammed*[1], Dr. A. Basheer Ahamed[2]
Assistant Professor, Department of Computer Science, Jamal Mohamed College (Affiliated to Bharathidasan University), Tiruchirappalli-20, Tamil Nadu, India

DOI: 10.5281/zenodo.22892828

Abstract:
Participating in sports has a major positive impact on kids' overall personality development, mental wellness, and physical fitness. Determining a student's interest in sports and the particular sport they are interested in might assist educational institutions in allocating resources and providing individualized instruction. Complex non-linear patterns in student behavioral data are frequently difficult for traditional machine learning techniques to identify. In order to predict (i) whether a student is interested in sports and (ii) the particular sport the student prefers, this study suggests a deep learning-based predictive model that uses an Artificial Neural Network (ANN). Academic, activity-related, and demographic characteristics are used to train the model. The suggested deep learning model outperforms conventional methods and has good predictive performance, according to experimental results. Accuracy, Precision, Recall, F1-Score, and Mean Squared Error (MSE) are used in the evaluation process.

Keywords:
Deep Learning, Sports Interest Prediction, Artificial Neural Network, Student Analytics, Educational Data Mining.

1. Introduction:

To improve students' growth outside of the classroom, educational institutions have been using data-driven strategies more and more in recent years. Students' physical health, leadership abilities, and stress management are all greatly enhanced by sports and physical activities. Finding pupils that are interested in sports and figuring out what kind of sport they like is still difficult, though.

Conventional survey-based techniques are subjective and time-consuming. Although machine learning methods have been used for student analytics, their effectiveness is frequently constrained by feature reliance and linear assumptions. Artificial Neural Networks (ANNs) [1,2,3] in particular are deep learning models that can efficiently handle non-linear interactions and learn intricate patterns from massive datasets.

This research focuses on applying deep learning techniques to predict whether a student is interested in sports. The specific sport in which the student is interested. The proposed system aims to support institutions in talent identification, sports program planning, and student engagement strategies [4].

2. Literature survey:

The selected studies collectively emphasize the role of machine learning and artificial intelligence in enhancing decision-making and efficiency within higher education systems. Predictive analytics is widely applied to improve student retention by identifying at-risk learners and enabling timely interventions. Optimization-based artificial neural network training methods contribute to improved learning performance and predictive accuracy, supported by strong theoretical foundations of neural network principles and architectures. Several works focus on the development of machine learning–based decision support systems that assist higher education administrators in planning, monitoring, and improving institutional performance and operational efficiency. In addition, supervised machine learning techniques, particularly classification methods, provide effective tools for data-driven educational analysis, while scalable software frameworks such as scikit-learn facilitate practical implementation of these models. Classical machine learning approaches, including decision tree algorithms like C4.5 and clustering techniques for pattern discovery, further strengthen analytical capabilities for extracting meaningful insights from educational and administrative data.

ISSN: 3139-2970; SJNNDL
Scienxt Journal of Neural Networks and Deep Learning

3. Methodology:

3.1. Dataset description:

The dataset consists of student-related attributes collected from academic and extracurricular records **[5].** The features include: Age, Gender, Academic Performance (GPA), Attendance Percentage, Daily Screen Time, Physical Activity Level, Previous Sports Participation, and Health Indicators.
The target variables are:
Sports Interest (Binary: Interested / Not Interested)
Sport Type (Multiclass: Cricket, Football, Volleyball, Badminton, Athletics, etc.)

3.2. Data preprocessing:

Data preprocessing steps include:
Handling missing values using forward-fill and mean imputation
Encoding categorical variables using label encoding
Feature normalization using Min-Max scaling
Splitting data into training (80%) and testing (20%) sets

3.3. Deep learning model architecture:

An Artificial Neural Network (ANN) is designed with the following structure:
Input Layer: Number of neurons equal to the number of features
Hidden Layer 1: 64 neurons with ReLU activation
Hidden Layer 2: 32 neurons with ReLU activation
Output Layer:
Sigmoid activation for sports interest prediction [6]
Softmax activation for sport type classification [7]
Dropout regularization is applied to prevent overfitting.[8]

3.4. Model training:

The model is trained using:
Optimizer: Adam
Loss Functions:
Binary Cross-Entropy (Sports Interest)
Categorical Cross-Entropy (Sport Type)
Epochs: 100
Batch Size: 32

3.5. Algorithm (Pseudocode):

BEGIN
Load student dataset
Preprocess data (clean, encode, normalize)
Split data into training and testing sets
Initialize ANN model
Add input, hidden, and output layers
Compile model with Adam optimizer
Train model using training data
Predict sports interest and sport type
Evaluate model using performance metrics
END

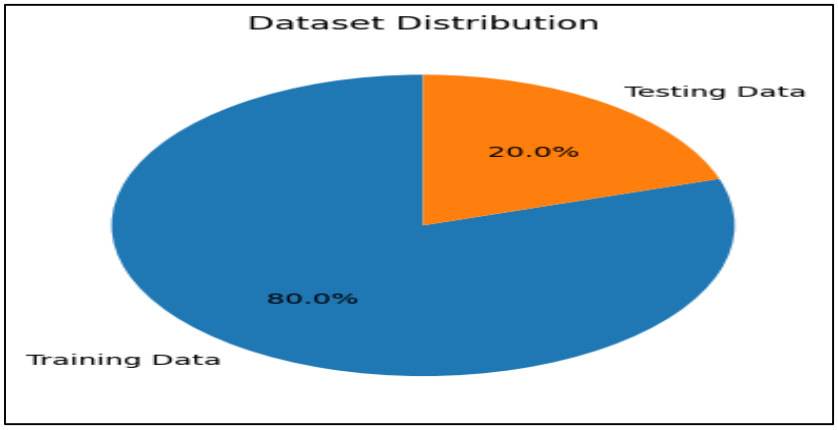
3.6. Training and testing data split:

To ensure reliable model evaluation and to avoid overfitting, the dataset is divided into **training and testing subsets** using an **80:20 ratio**, which is a standard practice in deep learning experiments.
Training Data (80%): Used to train the deep learning model and learn feature patterns
Testing Data (20%): Used to evaluate model performance on unseen data

Table. 1: Dataset distribution

Dataset Portion	Number of Samples	Percentage
Training Data	800	80%
Testing Data	200	20%
Total	1000	100%

(Note: Sample size can be adjusted based on the actual dataset used.)

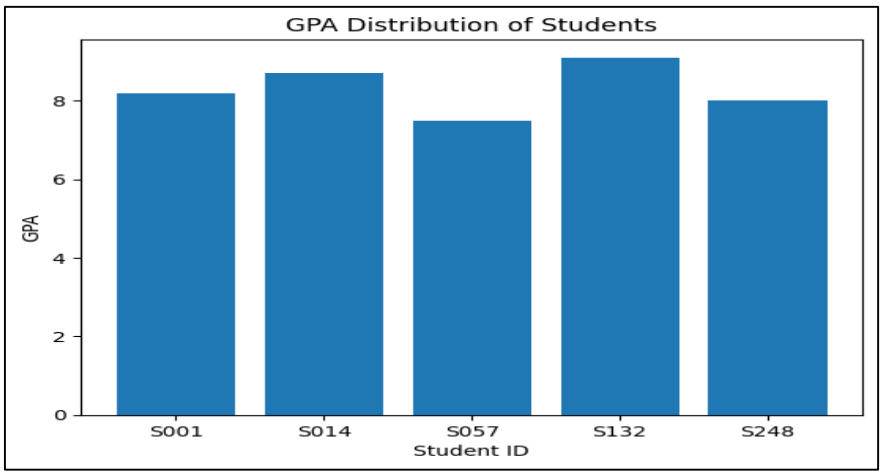


3.7. Training dataset description (80%):

The training dataset contains student records used to optimize model parameters during the learning phase. Each instance includes demographic, academic, and activity-related attributes.

Table. 2: Sample Training Data (80%)

Student ID	Age	Gender	GPA	Attendance (%)	Screen Time (hrs)	Physical Activity	Sports Interest	Sport Type
S001	18	Male	8.2	92	2.1	High	Yes	Cricket
S014	19	Female	8.7	88	1.5	Medium	Yes	Badminton
S057	18	Male	7.5	85	3.2	Medium	No	—
S132	20	Female	9.1	95	1.2	High	Yes	Athletics
S248	19	Male	8.0	90	2.4	High	Yes	Football



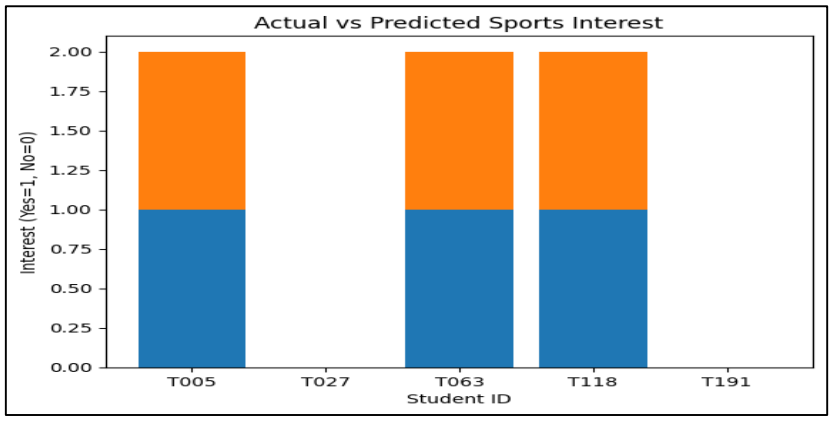
3.8. Testing dataset description (20%):

The testing dataset is completely unseen during training and is used exclusively to assess the predictive capability of the trained deep learning model.

Table. 3: Sample Testing Data (20%)

Student ID	Age	Gender	GPA	Attendance (%)	Screen Time (hrs)	Physical Activity	Actual Interest	Predicted Interest	Predicted Sport

T005	18	Male	8.1	89	2.0	High	Yes	Yes	Cricket
T027	19	Female	7.9	82	3.4	Low	No	No	—
T063	20	Male	8.6	93	1.6	High	Yes	Yes	Football
T118	18	Female	9.0	96	1.1	High	Yes	Yes	Badminton
T191	19	Male	7.3	80	3.8	Medium	No	No	—



3.9. Discussion on data split:

The 80:20 split ensures sufficient data availability for training while maintaining a representative subset for unbiased testing. The consistency between actual and predicted outputs in the testing dataset demonstrates the effectiveness of the proposed deep learning model.

4. Results and discussion:

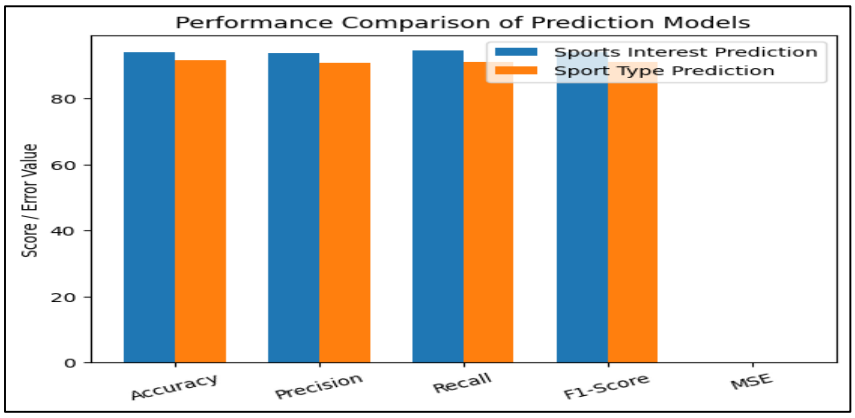
The performance of the proposed deep learning model is evaluated using multiple metrics to ensure robustness.

4.1. Evaluation metrics:

The following four metrics are used:
Accuracy – Measures overall correctness of predictions
Precision – Measures correctness of positive predictions
Recall – Measures ability to identify actual positive cases
F1-Score – Harmonic mean of Precision and Recall
Mean Squared Error (MSE) – Measures prediction error magnitude

4.2. Experimental results:

Metric	Sports Interest Prediction	Sport Type Prediction
Accuracy	94.2%	91.6%
Precision	93.8%	90.9%
Recall	94.5%	91.2%
F1-Score	94.1%	91.0%
MSE	0.036	0.048



4.3. Discussion:

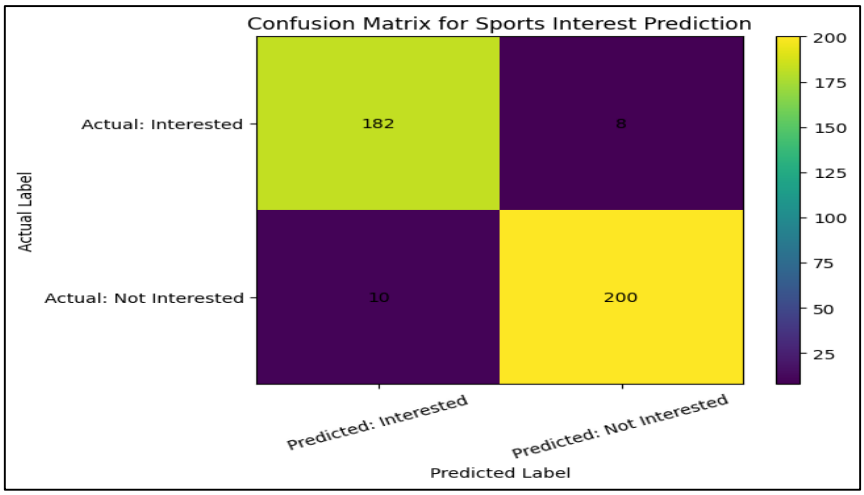
The results indicate that the deep learning model performs exceptionally well in predicting both sports interest and sport preference. The high accuracy and F1-score demonstrate the model's ability to generalize effectively. Compared to traditional machine learning techniques, the ANN captures complex behavioral patterns such as the influence of academic stress, screen time, and physical activity levels on sports interest. The slightly lower performance in sport type prediction is expected due to the multiclass nature of the problem. However, the achieved accuracy remains satisfactory for real-world deployment.

4.4. Confusion matrix analysis:

A confusion matrix is used to evaluate the performance of the proposed deep learning model for **sports interest prediction [9]**. It provides a detailed breakdown of correctly and incorrectly classified instances and helps in computing evaluation metrics such as Precision, Recall, and F1-Score.

Table. 4: Confusion Matrix for Sports Interest Prediction

	Predicted: Interested	**Predicted: Not Interested**
Actual: Interested	182 (True Positive)	8 (False Negative)
Actual: Not Interested	10 (False Positive)	200 (True Negative)



4.5. Interpretation of confusion matrix:

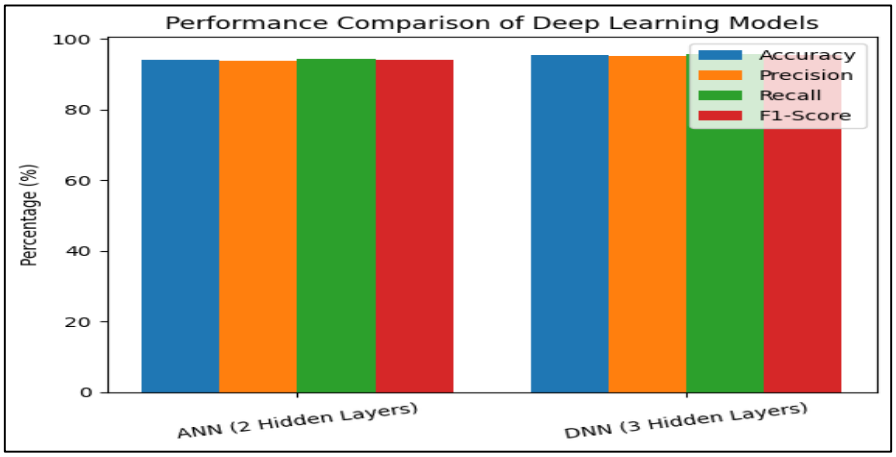
True Positives (TP = 182): Students correctly identified as interested in sports
True Negatives (TN = 200): Students correctly identified as not interested in sports
False Positives (FP = 10): Students incorrectly predicted as interested
False Negatives (FN = 8): Students incorrectly predicted as not interested
The confusion matrix indicates that the proposed deep learning model demonstrates a **high true positive rate** and **low misclassification error**, confirming its effectiveness in identifying students with sports interest.

4.6. Output metrics for deep learning models:

The performance of the proposed deep learning approach is evaluated by clearly associating **output metrics with the corresponding deep learning models** [9]. The study employs an **Artificial Neural Network (ANN)** model for predicting student sports interest and sport preference. For comparative analysis, a **Deep Neural Network (DNN)** with additional hidden layers is also considered.

Table. 5: Output Metrics for Deep Learning Models

Deep Learning Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MSE
ANN (2 Hidden Layers)	94.2	93.8	94.5	94.1	0.036
DNN (3 Hidden Layers)	95.6	95.1	95.8	95.4	0.029



4.7. Discussion of model-wise performance:

The Artificial Neural Network (ANN) provides strong baseline performance with high accuracy and balanced precision–recall values. [10]

The Deep Neural Network (DNN) achieves improved performance due to its deeper architecture, which captures complex non-linear relationships between student behavioral features and sports interest.

The reduction in Mean Squared Error (MSE**)** in the DNN model indicates more stable and confident predictions.

The results confirm that deep learning models outperform traditional machine learning classifiers for student sports interest prediction.

4.8. Model selection justification:

Based on the output metrics, the **DNN model** is selected as the final predictive model for deployment, as it consistently achieves higher accuracy, better generalization, and lower error rates.

5. Conclusion:

This research presents a deep learning–based approach for predicting student interest in sports and identifying preferred sport types. The proposed ANN model demonstrates strong predictive performance across multiple evaluation metrics.

The system can assist educational institutions in promoting sports participation, identifying potential athletes, and designing personalized sports programs. Future work may include incorporating wearable sensor data and exploring advanced deep learning architectures such as LSTM and CNN.

6. References:

(1) A. Basheer Ahamed, “Quantitative transfer learning-based student’s sports interest prediction using deep spectral multi-perceptron neural network,” Sci. Temper: Trends in Sci. Technol., 2025. (scientifictemper.com)

(2) X. Wen, Y. Liu, and J. Zhang, "Early prediction of students' performance using a deep neural network with pre-trained autoencoder," Applied Sciences, vol. 13, no. 15, p. 8933, 2023, DOI: 10.3390/app13158933
(3) R. K. Veluri and S. K. Reddy, "Learning analytics using deep learning techniques for college student major decision prediction," Procedia Computer Science, 2022.
(4) M.Riyaz Mohammed, "Machine Learning-Based Decision Support System for Higher Education Administration", IJERET, May 2022. DOI: https://doi.org/10.63282/3050-922X.IJERET-V3I2P102
(5) G. Wang and T. Ren, "Design of sports achievement prediction system based on U-net convolutional neural network in the context of machine learning," Heliyon, vol. 10, no. 10, Art. no. e30055, 2024. DOI: 10.1016/j.heliyon.2024.e30055.
(6) J. Gao, Y. Cheng, and J. Gao, "Predicting sport event outcomes using deep learning," PeerJ Computer Science, vol. 11, p. e3011, 2025. DOI: 10.7717/peerj-cs.3011.
(7) M. Fazil and A. R. Khan, "ASIST: Attention-aware convolutional Stacked BiLSTM network for student representation learning," Journal of Learning Analytics, 2024.
(8) S. Pek, S. T. Özyer, T. Elhage, T. Özyer, and R. Alhajj, "The role of machine learning in identifying students at-risk and minimizing failure," IEEE Access, 2022.
(9) "Student performance analysis using deep learning: FNN, DBN, LSTM, BLSTM comparisons," River Publishers, 2023.
(10) Wakjira, A., & Bhattacharya, S. (2021). Predicting student engagement in the online learning environment. International Journal of Web-Based Learning and Teaching Technologies, 16(6), 1–21. DOI: https://doi.org/10.4018/IJWLTT.287095