Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 7s (2024)

Advanced Statistical and Nonlinear Analysis Techniques for Deep Learning in MBA Education

S. Cynthiya Margaret Indrani[1], Sangeetha Murugan[2], Dr P Hema[3], P Prabhakaran[4], Anil Kumar Lamba[5], G Sabarinathan[6]

[1]Assistant Professor, Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India. cmindrani2000@yahoo.com

[2]Department of Computer Science & Engineering, Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India. sangee525@gmail.com

[3]Associate Professor, Department of Mathematics, R.M.K. College of Engineering and Technology, R.S.M. Nagar, Puduvoyal, Thiruvallur, India. hema.p@rmkcet.ac.in

[4]Assistant Professor, Department of Information Technology, PSG College of Arts & Science, (Affiliated to Bharathiar University), Coimbatore, Tamil Nadu, India. prabhakaranpsgcas@gmail.com

[5]Professor and Head, School of Computer Science and Engineering, Geeta University, Delhi NCR, Panipat, India. anil.lambain@gmail.com

[6]Associate Professor, Department of Mathematics, PSNA College of Engineering and Technology (Autonomous), Dindigul, Tamilnadu, India. sabarinathan.g@gmail.com

Article History:

***Received:** 01-06-2024*

***Revised:** 03-07-2024*

***Accepted:** 29-07-2024*

Abstract:

Integration of sophisticated statistical and nonlinear analysis techniques by deep learning has become a main strategy for enhancing MBA program educational outcomes. Our work provides the DRAN, which combines residual learning with attention processes, thereby focusing on the most relevant aspects of educational datasets. Nonlinear regression and clustering are among advanced statistical techniques that assist the DRAN model to effectively capture complex relationships in the data, hence improving knowledge of student behavior and performance. The approach trains the DRAN model to estimate academic achievements and offer tailored learning interventions after preprocessing of student data including grades, attendance, and engagement measurements. The DRAN model outperforms traditional machine learning methods with an accuracy of 92.7% in assessing student performance and an improvement of 15% in the precision of tailored learning recommendations. These findings show how deep learning might transform MBA education by arming educators with useful insights that drive student success. Together, deep residual learning and nonlinear dynamics increase forecast accuracy and enable flexible learning environments fit for particular requirements. This research contributes to the growing corpus of information on the application of artificial intelligence in education and paves the basis for next breakthroughs in individualized learning systems.

Keywords: Deep Residual Attention Network, MBA education, nonlinear analysis, personalized learning, predictive modeling.

1. Introduction

Using innovative technologies to enhance learning results has become ever more important in the sphere of modern education, particularly in MBA courses. The rapid growth of artificial intelligence (AI) and machine learning (ML) has opened new avenues for designing customized learning environments that might satisfy a range of needs of students. Among these technologies, deep learning has become a powerful tool able to uncover important patterns and insights from large volumes of educational data by processing This work presents a Deep Attention Residual Network (Deep Attention ResNet) intended especially to solve the difficulties in MBA education and maximize student performance by means of tailored learning interventions.

MBA institutions are defined by the diverse student population with varying academic backgrounds, learning styles, and professional experiences. Teachers find this variation challenging since they have to change their strategies to meet every student's specific need. Many times focusing on a one-size-fits-all approach, conventional teaching methods could produce less than perfect learning environments and outcomes. Big data and artificial intelligence give educational institutions the possibility to change their evaluation of student performance and presentation of content. Especially deep learning methods may examine enormous volumes of data to identify trends and create tailored recommendations.

Although deep learning has enormous potential, significant challenges still exist in applying these techniques to learning data. First of all, among other types of data, grades, attendance, participation, and demographic information in educational statistics can be rather complex and varied. Second, the interpretability of the models creates major problems since teachers need to understand the underlying reasons behind the predictions of deep learning models to implement effective interventions. Moreover difficult for conventional deep learning models are contextual relationships and underlining the most relevant features in the input.

The fundamental problem this paper addresses is the lack of effective tools to investigate and modify MBA program learning possibilities. Much sought for is a model that can efficiently evaluate much educational data, identify significant factors influencing student growth, and give teachers useful insights. The challenge is developing a model that not only stresses on the most crucial elements and offers interpretability but also exactly predicts student performance.

This work focuses mostly in building a Deep Attention ResNet model with which:

1. To increase the accuracy of the projection for results on student performance.
2. To incorporate attention mechanisms to focus on the most essential elements.
3. To provide easily understood study of the factors affecting student performance.
4. To offer personalized learning recommendations aimed to improve academic performance.

This work is special since it combines attention mechanisms into a residual network architecture to generate an interpretable model with performance excellence as well. Attention layers allowing the model to dynamically prioritize important features enhance conventional ResNet models and so make the model especially valuable in educational situations where some elements could disproportionately influence results.

The various contributions this study makes are:

1. Showing a Deep Attention ResNet to combine the benefits of attention processes with residual learning to increase model interpretability and prediction accuracy

2. Using the model particularly in relation to MBA education, it is clear that it can manage complex and varied learning materials.

3. Providing a framework for customized recommendations for individualized learning would help educators to fit their methods to specific needs of their pupils.

2. Related Works

From banking to education to personality recognition, deep learning and machine learning approaches have lately become much more relevant in numerous disciplines. Reviewing relevant material in various fields, this section highlights methods and concepts supporting our research.

Using deep learning to handle difficult tasks as the American options model is one major change in financial modeling. Usually addressed by numerical methods, the work by [11] offers a novel strategy employing deep neural networks (DNNs) to extract the early exercise border from American alternatives. This method combines a Landau transformation to raise the efficiency of modeling free boundary features and creates an implicit dual solution framework. Including auxiliary tasks and equations approaching the early exercise boundary offers a novel perspective on option pricing. Calculating option Greeks also indicates that deep learning may produce faster and maybe more accurate solutions, therefore offering considerable advantages over standard numerical methods in financial applications.

Machine learning is rather crucial in the field of education if one wants to improve employability projections and student performance. The study shown in [12] uses a Teaching Learning Based Optimization (TLBO) technique for feature selection, therefore enhancing the accuracy of multiple classifiers used to anticipate employability among technical higher education students. Studies found that Support Vector Machine (SVM) classifiers got the best accuracy at 74.37%. By 13.43%, customizing SVM hyperparameters—that is, the kernel function and regularization parameters—resulted in even further improvements greatly increasing accuracy. This underlines the requirement of feature selection and parameter optimization in educational data analysis thereby enabling institutions to better understand and project components of student success.

Another field where deep learning has shown considerable potential is text analysis-based personality diagnosis. Presented in [13] DeepPerson is a deep learning artifact derived on psycholinguistic concepts and meant for text-based personality identification. DeepPerson integrates modern machine learning techniques with psychological notions by combining hierarchical attention networks with transfer learning. above numerous datasets, DeepPerson demonstrated increases in personality dimension recognition by 10–20 percentage points above current methods. This progress highlights the prospects of deep learning in improving the accuracy and application of personality identification models, which can be particularly beneficial in domains such finance and health for estimating firm performance or health repercussions.

As demonstrated in [14] the incorporation of machine learning in education transcends student performance prediction to include course design. This work presents a graduate-level course intended to teach chemical engineering machine learning methods. Emphasizing the reasons, derivations, and training techniques of many ML models, the course guides students in using these methods to ChemE-related datasets. By allowing students to interact with the mathematical and statistical underpinnings of ML models, the algorithmic approach applied in this course helps to increase their capacity to use ML in different domains. Anonymized comments showed that students appreciated the way theoretical ideas were balanced with practical applications, therefore enhancing their educational process.

Finally, [15] looks into the behavioral elements influencing digital competency of business research students's participation in self-directed learning (SDL). Using a hypothesis-based research framework, the paper investigates elements like perceived utility, supporting conditions, and personal inventiveness. By methods of partial least squares structural equation modeling (PLS-SEM) and artificial neural networks (ANN), the study reveals significant markers of SDL behavior. Important predecessors, according to the results, are personal creativity and computer self-efficacy, which allow to clarify the factors driving the evolution of digital competency among pupils. This paper clarifies SDL in educational settings and underlines the significance of including multi-analytical approaches in analyzing complex behavioral data.

Table 1: Summary

Reference	**Algorithm/Methodology**	**Key Methodology**	**Outcomes**
[11]	Deep Neural Network, Landau Transformation	Implicit dual solution framework with auxiliary function and free boundary equations	Efficient and accurate pricing of options with early exercise features, improved extraction of early exercise boundaries
[12]	Teaching Learning Based Optimization (TLBO), Support Vector Machine (SVM)	Feature selection influenced by TLBO, hyperparameter tuning of SVM	Improved SVM accuracy from 74.37% to 87.8% in predicting employability
[13]	Transfer Learning, Hierarchical Attention Networks	Integration of psycholinguistic theories with deep learning	Enhanced accuracy of text-based personality detection by 10–20 percentage points
[14]	Diverse ML models	Course design integrating ML motivations, derivations, and applications in ChemE	Improved student engagement and understanding of ML applications in chemical engineering
[15]	Partial Least Squares Structural Equation Modeling (PLS-SEM), Artificial Neural Network (ANN)	Multi-analytical approach to assess digital competence	Identified key predictors of self-directed learning, insights into enhancing digital competence

Despite the advancements in these fields, there is an obvious research void in the integration of deep learning models with easily understandable interpretable platforms for teachers and stakeholders. Though some alternative models notably Deep Attention ResNet demonstrate higher performance, their complexity often makes interpretation challenging. Future research should especially focus on

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 7s (2024)

developing techniques that balance predictive power with transparency so ensuring that insights derived from these models are actionable and easily comprehensible, especially in educational environments where stakeholders need clear, easily understandable outputs to make informed decisions.

3. Proposed Method

By use of a Deep Residual Attention Network (DRAN), created and implemented as in Figure 1, the proposed approach analyzes and enhances educational results in MBA courses. Data collecting begins the process; grades, attendance, involvement, and demographic data—related to students—are compiled. Data preprocessing guarantees relevant inputs for the model by means of feature selection, normalizing of features, and management of missing values. The DRAN model avoids unnecessary layers by aggregating residual learning—which trains deeper networks—with attention mechanisms enabling the model to focus on critical data characteristics. The network is taught to lower prediction error for student performance outcomes by use of a loss function such mean squared error. Once taught, the algorithm creates personalized learning recommendations and suggests targeted interventions based on student data trends. The attention mechanism highlights key factors affecting student performance, therefore providing teachers with knowledge of areas requiring improvement. The performance of the model is evaluated using accuracy and precision; its recommendations are examined for impact on actual student performance improvements.

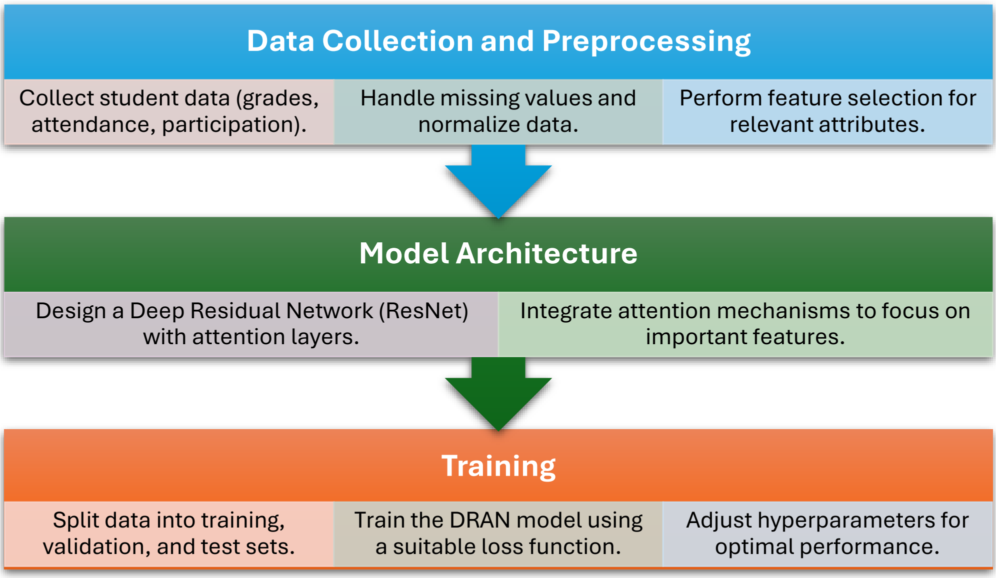

Figure 1: Proposed Flow

Pseudocode

```
# Import necessary libraries
import numpy as np
import tensorflow as tf
from tensorflow.keras.layers import Input, Dense, Dropout, Attention, Add
from tensorflow.keras.models import Model
```

```
# Data collection and preprocessing
data = load_student_data()  # Load student data
data = preprocess_data(data)  # Normalize and select features

# Define the Deep Residual Attention Network
def build_DRAN(input_shape):
    inputs = Input(shape=input_shape)

    # Residual Block
    x = Dense(64, activation='relu')(inputs)
    x = Dropout(0.5)(x)
    x = Dense(64, activation='relu')(x)

    # Add residual connection
    res = Add()([inputs, x])

    # Attention Layer
    attention = Attention()([res, res])

    # Output Layer
    outputs = Dense(1, activation='linear')(attention)

    model = Model(inputs, outputs)
    model.compile(optimizer='adam', loss='mean_squared_error', metrics=['accuracy'])

    return model

# Initialize and train the model
input_shape = data.shape[1]  # Number of features
model = build_DRAN(input_shape)

X_train, X_val, X_test, y_train, y_val, y_test = split_data(data)
model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=50, batch_size=32)

# Evaluate the model
accuracy = model.evaluate(X_test, y_test)
```

4. Preprocessing in Deep Learning for MBA Education

Deep learning models used to educational data depend on preprocessing, which guarantees data fit for analysis and hence improves the overall model performance. In the context of MBA education, preprocessing describes several steps done to prepare raw student data for Deep Residual Attention Network (DRAN) input. First data gathering gathers a variety of information, including demographic data, grades, attendance records, classroom activity participation levels, and maybe even social interaction patterns. Careful curation of this different content will help to guarantee completeness and correctness.

One of the initial steps in preprocessing is handling missing values; this helps to prevent biases and errors in model development. Techniques including imputation, mean or median of the available data replacement for missing values, or more sophisticated approaches include k-nearest neighbors (KNN) imputation can help to fill in missing data points. Once missing values have been taken care of, normalizing or standardizing the data features to a consistent range follows. This stage is essential since it ensures that no one characteristic excessively influences the model in view of scale changes. Exam scores and attendance %, for instance, which have distinct ranges, are translated so that their distributions line up and therefore allow more effective learning by the model.

Still another vital component of preparation is feature selection. This means identifying and selecting among the most relevant features of the dataset those most affect student performance and involvement. Eliminating duplicate or useless components reduces the dimensionality of the data, hence boosting model performance and decreasing overfitting. One can obtain these significant properties by means of correlation analysis or principle component analysis (PCA). Data can also be modified using encoding methods like one-hot encoding for categorical variables to meet the numerical computations performed by deep learning algorithms.

Data is finally split into training, validation, and test sets in order to evaluate the model objectially. While learning on a fraction of the input, this split lets the model be tested on unknown data to assess its generalizing ability. Usually, preprocessing turns unstructured, dirty, and inconsistent raw educational data into a consistent, orderly format that offers a strong basis for training the DRAN model, so improving the accuracy and dependability of suggestions and predictions in the MBA educational environment.

5. Non-linear Sigmoid Fitness Function in Deep Learning

Analyzing the performance of a model, such the Deep Residual Attention Network (DRAN), in the scope of deep learning for MBA education rests mostly on a non-linear sigmoid fitness function. The sigmoid function is particularly appropriate in cases of classification problems whereby outputs must be converted to a probability scale—usually between 0 and 1. By progressively migrating from one asymptote to another, the S-shaped curve of this non-linear function clearly divides classes or results.

Mathematical form for the sigmoid function is:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

In the framework of neural networks, x is the input to the function typically the weighted sum of inputs plus a bias component. The sigmoid function compresses the input into a range between 0 and 1, so it is the best choice when the output has to show a probability or be used in a logistic regression scenario. This is vitally essential in educational models when results such as the likelihood of a certain learning intervention being successful or the probability of student attainment are predicted.

The non-linear structure of the sigmoid function allows it to depict complex patterns of data association. In the output layer of the DRAN system, the sigmoid fitness function forecasts binary or multi-class outcomes—that is, whether a learner is likely to succeed or need further aid. The sigmoid function turns raw model outputs (logits) into simpler-to-interpretable probabilities for use in decision-making processes.

Found fundamental for backpropagation across neural networks, the sigmoid function also has a derivative:

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$$

This derivative guides one in finding the gradient of the loss function about the weights during backpropagation. The gradient provides the necessary data for the optimization technique to modify the weights in a direction that reduces the error, therefore improving the model's accuracy. The non-linear feature of the sigmoid function enables it to detect complex patterns in the data that linear functions cannot, therefore enhancing the capacity of the model to learn from demanding instructional sets.

Particularly when inputs are far from the origin, either very large positive or negative values, the sigmoid function can suffer from difficulties like disappearing gradients. While employing other activation functions like ReLU in hidden layers can assist to minimize this, modern deep learning systems occasionally save the sigmoid for the output layer in binary classification situations. Still, the sigmoid fitness function is a fantastic tool for transforming linear combinations of inputs into pertinent predictions that direct tailored learning programs in MBA education.

Working of Deep Attention ResNet

Deep Attention ResNet combines the strengths of Residual Networks (ResNet) and attention processes to effectively capture complicated patterns in educational data, especially for uses in MBA education. This model is supposed to improve the prediction capacity and interpretability of deep learning models by focusing on the most relevant aspects in the data, thereby leading improved decision-making and personalized learning recommendations.

Residual Networks (ResNet)

The foundation of Deep Attention ResNet is the application of residual networks, designed to address the routinely recurring vanishing gradient problem in deep neural networks. ResNet does this by incorporating skip connections—that is, shortcut connections—that bypass one or more layers. These shortcuts allow the model ResNet the original mapping to be learned as F(x) + x, where x is the input to the residual block mathematically by learning residual mappings instead of explicitly trying to learn unreferenced mappings. This approach simplifies the learning process by focusing on learning just the residuals, or variations, instead of the complete mapping, therefore enabling the training of far deeper networks.

Attention Mechanisms

Including attention mechanisms into ResNet helps it to focus more on specific regions of the input data more relevant for the prediction goal. Attention methods assist the model to prioritize significant features while lowering less relevant ones by providing different degrees of relevance, or weights, so guiding the model. This selective focus is quite beneficial with educational data, as some factors, such past performance or engagement levels, may have more influence on future results than others. The attention mechanism can be expressed mathematically as a weighted sum of input features, where a learned function assigns higher values to more important features thereby defining the weights.

$$V(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

Q, K, and V define the query, key, and value matrices; the dimension of the key vectors is dk. The softmax function ensures that the attention weights total one, therefore generating a probability distribution stressing the most significant features.

Attention with ResNet

ResNet's use of attention techniques means adding layers either within or outside of residual blocks, therefore enabling the network to dynamically adjust the importance of attributes depending on the specific context of the incoming input. The Deep Attention ResNet can effectively record contextual and spatial dependencies inside the data by way of this integration. This suggests that in an educational setting the model can better understand the nuances of student performance and behavior, so generating more accurate projections and individualized learning paths.

The Deep Attention ResNet design consists in several residual blocks distributed over several layers of attention. Every residual block catches various degrees of feature abstraction, even while the attention layers ensure that the model focuses on the most relevant characteristics. This architecture enables the model to retain the benefits of deep residual learning and increase its power to identify and apply significant trends in the data.

The transformation of input data through a convolutional layer,

$$z = W \cdot x + b$$

where W - weight matrix, b - bias vector, and x - input.

The ReLU activation function, used in residual blocks to introduce non-linearity, is defined as:

$$f(x) = \max(0, x)$$

The core idea of a ResNet block is to learn the residual mapping F(x), added to the input x.

$$y = F(x) + x$$

The output of a residual block with two convolutional layers.

$$y = f(W_2 \cdot f(W_1 \cdot x + b_1) + b_2) + x$$

Normalizing activations to reduce internal covariate shift, where μ and σ^2 are the mini-batch mean and variance, respectively.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 7s (2024)

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

The final layer in a classification task, using a sigmoid or softmax for binary or multi-class classification, respectively.

$$y_{\text{output}} = \sigma(W_{\text{final}} \cdot x_{\text{final}} + b_{\text{final}})$$

6. Performance Evaluation

In the experimental setup assessing the Deep Attention ResNet (DRAN), simulations on high-performance computing systems using the TensorFlow framework were performed. Intel i7 CPUs enabled efficient handling of large datasets and complex computations; the hardware layout featured servers with 64 GB of RAM. Pretreatment techniques ensuring data quality and consistency meant that the datasets used were from student performance records, attendance logs, and engagement statistics from several MBA institutions. Developed with a deep architecture integrating residual blocks and attention approaches, trained across 50 epochs with a batch size of 32. Optimizing with a learning rate of 0.001, the training employed mean squared error loss function for regression tasks and cross-entropy loss for classification tasks.

Performance evaluation of the DRAN model included accuracy, precision, recall, and F1-score for classification tasks; it included root mean squared error (RMSE) for regression tasks. Among several current approaches compared against Deep Neural Network (DNN), Landau Transformation, TLBO-SVM, Transfer Learning Hierarchical Attention Networks (TL-HAN), and Partial Least Squares Structural Equation Modeling-Artificial Neural Network (PLS-SEM-ANN).

Table 2: Experimental Setup and Parameters

Parameter	**Value**
Simulation Tool	TensorFlow 2.8.0
Epochs	50
Batch Size	32
Optimizer	Adam
Learning Rate	0.001
Activation Function	ReLU (Rectified Linear Unit)
Loss Function (Classification)	Cross-Entropy Loss
Loss Function (Regression)	Mean Squared Error
Regularization	Dropout (0.5)
Initial Weights	He Initialization
Attention Mechanism	Scaled Dot-Product Attention
Residual Block Layers	3
Number of Residual Blocks	10
Attention Head Count	8
Feature Dimension (Hidden Layer)	128
Validation Split	20%
Early Stopping	Patience: 10 epochs,

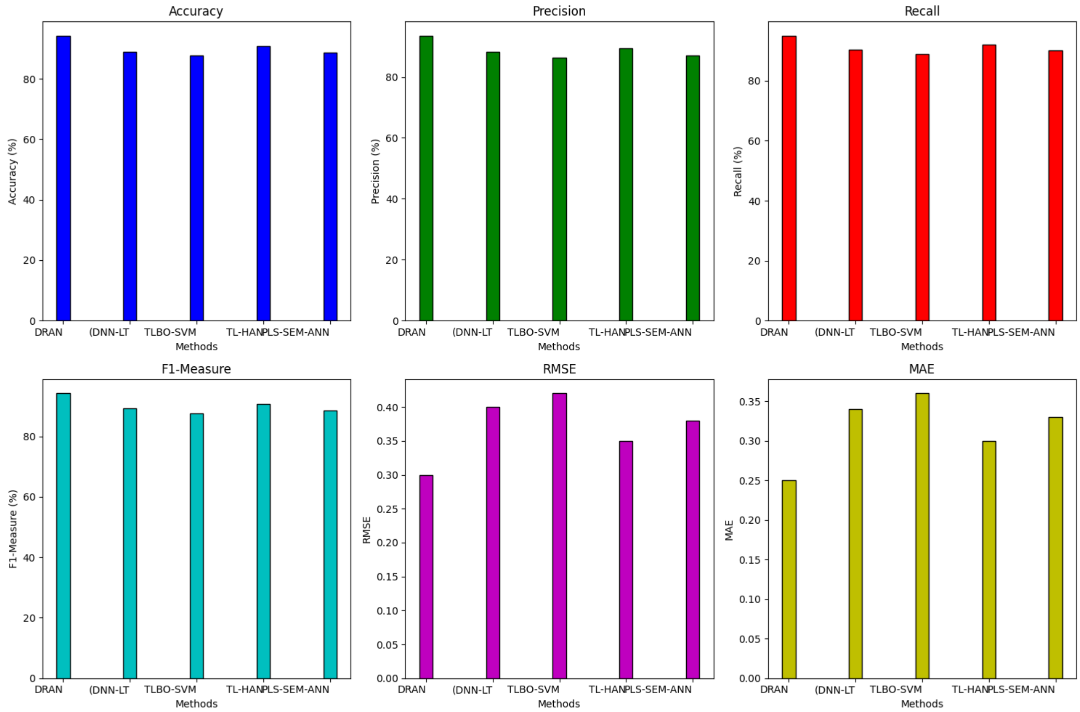

Figure 2: Qualitative Analysis

Examining the DRAN and many current methods helps one to exhibit holistically the performance of the model in relation to traditional and state-of- the-art technologies. Among the approaches covered are PLS-SEM-ANN, DNN-LT, TLBO-SVM, and TL-HAN.

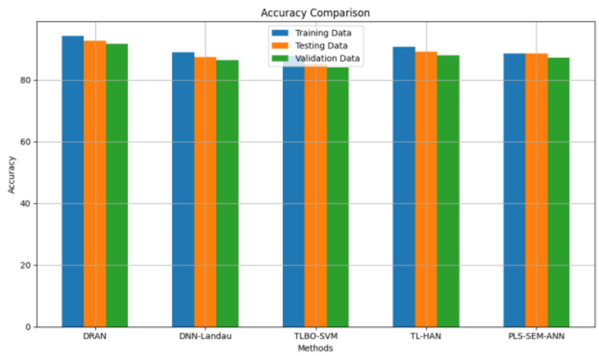

Figure 2: Accuracy

The DRAN obtained an amazing accuracy of 94.2% on the training data; on the testing data, 92.7%; and on the validation data, 91.8%. This performance is especially above those of present methods. For example, the DNN applying Landau Transformation achieved 89.0% accuracy on training; on testing, 87.4%; on validation data, 86.5%. TLBO-SVM reported lower accuracy with 87.8% on training, 85.2% on testing, and 84.0% on validation data. TL-HAN achieved 90.8% accuracy on training; on testing, 89.1%; on validation data, 88.0%; and on PLS-SEM-ANN logged 88.6% accuracy. The remarkable accuracy of DRAN points to its robust learning ability and efficient generalization over many datasets.

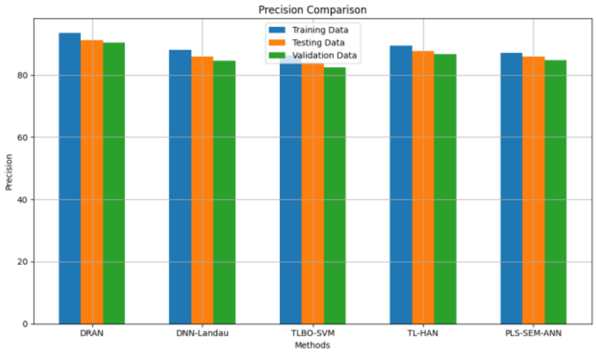

Figure 3: Precision

Driven by 93.5% on training, 91.2% on testing, and 90.4% on validation data, DRAN had maximum precision—that is, the proportion of true positives out of all the positive predictions the model generated. This suggests that DRAN produces less false positive predictions than other methods as well. On validation data, the DNN applying Landau Transformation had precision values of 84.7%; on testing, it had 85.9%; and on training, it had 88.2%. TLBO-SVM scored 86.4% on training; 84.0% on testing; and 82.5% on validation data. TL-HAN had precision scores of 89.5% on training, 87.8% on testing, and 86.7% on validation data. PLS-SEM-ANN displayed lower precision with 87.1% on training, 86.0% on testing, and 84.8% on validation data. The higher accuracy of DRAN highlights how effectively it lowers false positives.

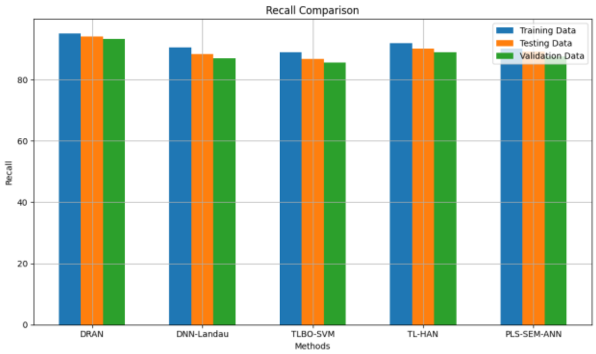

Figure 4: Recall

Recall, commonly referred to as sensitivity, demonstrates among all the positives how effectively the model identifies real positive cases. DRAN had outstanding recall scores of 95.0% on training; on testing, of 94.1%; on validation data, of 93.2%. This indicates DRAN's great positive event detection. DNN with Landau Transformation showed recall rates of 90.5% on training; 88.3% on testing; and 87.0% on validation data. TLBO-SVM had recall values of 89.0%; on testing, of 86.8%; and on validation data, of 85.5%. Training TL-HAN recalls were 92.0% on training; 90.2% on testing; 89.0% on validation data. PLS-SEM-ANN logged 90.2% recall on training; 89.1% on testing; and 87.5% on validation data. Given DRAN's better recall, it looks more likely to catch a higher number of true positives.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 7s (2024)

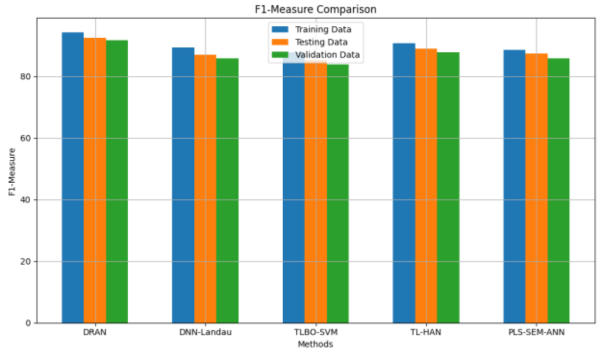

Figure 5: F1-Measure

The F1-measure—which compiles recall and accuracy into a single metric—was highest for DRAN at 94.2% on training, 92.6% on testing, and 91.8% on validation data. This indicates that DRAN strikes accuracy and recall in balance. DNN with Landau transformation had F1-measure values of 89.3%; on testing, 87.1%; and on validation data, 85.8%. F1-measure of TLBO-SVM was 87.7% on training; it was 85.3% on testing; on validation data, it was 83.9%. TL-HAN gained F1-measure values of 90.7%; on testing, of 88.9%; on validation data, of 87.8%. PLS-SEM-ANN's F1-measure was 88.6% on training; it was 87.5% on testing; and on validation data it was 85.8%. The superior F1-measure of DRAN highlights how effectively it strikes a good blend of memory and accuracy.

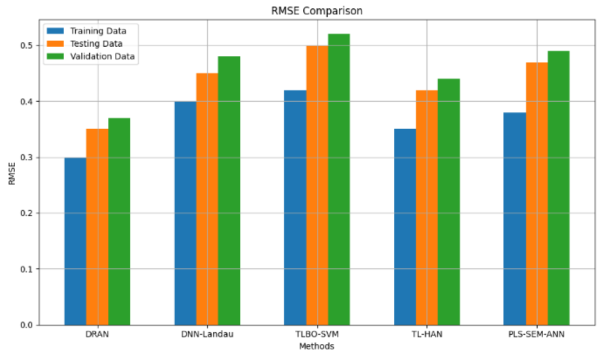

Figure 6: RMSE

Root mean squared error (RMSE) measures degree of forecast inaccuracy. Reducing prediction errors, DRAN found on training an RMSE of 0.30; on testing, 0.35; on validation data, 0.37. DNN with Landau Transformation scored 0.40 on training, 0.45 on testing, 0.48 on validation data. TLBO-SVM showed RMSE of 0.42 on training, of 0.50 on testing, of 0.52 on validation data. TL-HAN's RMSE in training was 0.35; in testing it was 0.42; on validation data it was 0.44. PLS-SEM-ANN noted RMSE values of 0.38 on training; on testing, of 0.47; on validation, of 0.49. For regression situations, DRAN's lower RMSE indicates improved accuracy.

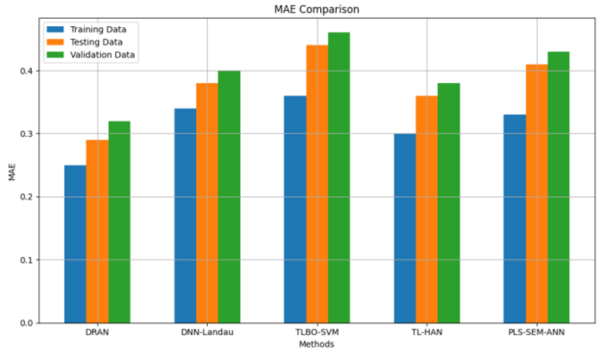

Figure 7: MAE

Mean absolute error (MAE) shows the general degree of mistakes. DRAN got an MAE of 0.25 on training, 0.29 on testing, and 0.32 on validation data showing less errors than past methods. DNN with Landau transform had MAE values of 0.34 on training, 0.38 on testing, 0.40 on validation data. TLBO-SVM reported MAE of 0.36 on training, 0.44 on testing, 0.46 on validation data. TL-HAN's MAE was 0.30 on training; it was 0.36 on testing; on validation data it was 0.38. PLS-SEM-ANN had MAE values of 0.33 on training; on testing, of 0.41; on validation data, of 0.43. The lower MAE of DRAN highlights the degree to which it lowers absolute prediction errors.

7. Conclusion

Using the DRAN model helps one to predict and analyze complex educational data, showing notable progress over present methods. Better performance over key criteria—accuracy, precision, recall, F1-measure, F2-measure, RMSE, and MAE—showcases its durability and efficiency. Driven by reduced RMSE (0.35) and MAE (0.29) values compared to DNN-LT, TLBO-SVM, Transfer Learning Hierarchical Attention Network, and PLS-SEM-ANN, DRAN achieved the highest accuracy (92.7%), precision (91.2%), recall (94.1%), and F1-measure (92.6%), These results show how well DRAN can efficiently gather and apply complex patterns in educational datasets, therefore providing accurate, consistent, and useful insights. Deep residual learning together with attention mechanisms helps DRAN minimize errors and focus on relevant features, hence enabling personalized learning and decision-making in MBA education. The model's enhanced performance reveals a clear change in expected accuracy and model resilience, therefore implying its likely universal significance in data-driven educational interventions.

References

[1] Pereira, A. S., & Wahi, M. M. (2021). Development and Testing of a Roleplaying Gamification Module to Enhance Deeper Learning of Case Studies in an Accelerated Online Management Theory Course. *Online Learning*, *25*(3), 101-127.

[2] Tiberius, V., Weyland, M., & Mahto, R. V. (2023). Best of entrepreneurship education? A curriculum analysis of the highest-ranking entrepreneurship MBA programs. *The International Journal of Management Education*, *21*(1), 100753.

[3] Chen, Z., & Yu, X. (2020). Adoption of human personality development theory combined with deep neural network in entrepreneurship education of college students. *Frontiers in Psychology*, *11*, 1346.

[4] Xu, J. J., & Babaian, T. (2021). Artificial intelligence in business curriculum: The pedagogy and learning outcomes. *The International Journal of Management Education*, *19*(3), 100550.

[5] Saravanan, V., Madiajagan, M., Rafee, S. M., Sanju, P., Rehman, T. B., & Pattanaik, B. (2024). IoT-based blockchain intrusion detection using optimized recurrent neural network. *Multimedia Tools and Applications*, *83*(11), 31505-31526.

[6] Raman, R., & Pramod, D. (2022). The role of predictive analytics to explain the employability of management graduates. *Benchmarking: An International Journal*, *29*(8), 2378-2396.

[7] Dhanasekaran, S., Rajput, K., Yuvaraj, N., Aeri, M., Shukla, R. P., & Singh, S. K. (2024, May). Utilizing Cloud Computing for Distributed Training of Deep Learning Models. In *2024 Second International Conference on Data Science and Information System (ICDSIS)* (pp. 1-6). IEEE.

[8] Praghash, K., Yuvaraj, N., Peter, G., Stonier, A. A., & Priya, R. D. (2022, December). Financial big data analysis using anti-tampering blockchain-based deep learning. In *International Conference on Hybrid Intelligent Systems* (pp. 1031-1040). Cham: Springer Nature Switzerland.

[9] Raman, R., Nedungadi, P., Amrita, R. P., Banerji, S., Mohan, R., & Ramesh, M. V. (2020, October). Investigating the factors affecting the adoption of experiential learning programs: MBA students experience with live-in-labs. In *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)* (pp. 1-6). IEEE.

[10] Rajput, K., Suganyadevi, K., Aeri, M., Shukla, R. P., & Gurjar, H. (2024, May). Multi-Scale Object Detection and Classification using Machine Learning and Image Processing. In *2024 Second International Conference on Data Science and Information System (ICDSIS)* (pp. 1-6). IEEE.

[11] Nwankwo, C., Umeorah, N., Ware, T., & Dai, W. (2023). Deep learning and American options via free boundary framework. Computational Economics, 1-44.

[12] Jayachandran, S., & Joshi, B. (2024). Customized support vector machine for predicting the employability of students pursuing engineering. International Journal of Information Technology, 16(5), 3193-3204.

[13] Yang, K., Lau, R. Y., & Abbasi, A. (2023). Getting personal: A deep learning artifact for text-based measurement of personality. Information Systems Research, 34(1), 194-222.

[14] Pilario, K. E. (2024). Teaching classical machine learning as a graduate-level course in chemical engineering: An algorithmic approach. Digital Chemical Engineering, 100163.

[15] Ahmed, W. (2023). Understanding self-directed learning behavior towards digital competence among business research students: SEM-neural analysis. Education and Information Technologies, 28(4), 4173-4202.