Analysis of Factors Influencing the Outcomes of AR Based Education Approach in STEM Learning Using Machine Learning

MSc Research Project
Data Analytics

Priyadharsini Packirisamy
Student ID: x22199098

School of Computing
National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Priyadharsini Packirisamy
Student ID:	x22199098
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Furqan Rustam
Submission Due Date:	16/09/2024
Project Title:	Analysis of Factors Influencing the Outcomes of AR Based Education Approach in STEM Learning Using Machine Learning
Word Count:	7656
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Priyadharsini Packirisamy
Date:	16th September 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Analysis of Factors Influencing the Outcomes of AR Based Education Approach in STEM Learning Using Machine Learning

Priyadharsini Packirisamy
x22199098

Abstract

Augmented reality (AR) based education approach is an evolving domain. Several studies have been conducted to analyse the impacts of AR-based education. However, these studies have certain limitations, like usage of rudimentary techniques such as basic mathematical and statistical methods and analysing the generic binary outcomes like positive or negative impacts. In practice, to implement a successful AR-based education, numerous factors will weigh in. Few examples are knowledge of technology, accessibility to the application, subject matter interest, and infrastructure challenges. Another limitation observed in these studies are the lack of adequate data both in terms of number of participants and data dimensionality. The aim of this research is to identify the potential factors that influence the outcomes of AR-based education approach which will facilitate addressing the areas of improvement. Machine learning algorithms have been used to determine these factors, as they are well suited to identify the patterns and to solve real-world problems effectively. This overcomes the inadequacy of the basic analysis techniques that were previously used. For fair analysis, an extensive dataset collected as part of the ARETE's Pilot 2 project with 1988 participants across 11 countries and varied age groups for two subjects - math and science, divided as intervention and control groups have been chosen. This fills the gap in previous studies of inadequate data. The data has been preprocessed, and existing trends in the data has been analysed using exploratory data analysis. This data has been used to train the neural networks RNN, CNN, and LSTM and the machine learning model SVM to identify the influencing factors. The overall accuracy yielded by the models are 99%, 98%, 96%, and 94% for RNN, CNN, LSTM, and SVM respectively. This research revealed that age, gender, test book of the subject, media usage and availability of technology are some of the most influential factors in the AR-based learning approach. Also, the overall impact shows positive attitude towards the AR-based education approach.

1 Introduction

The objective of this project is to analyse the factors influencing the outcome of the AR-based education approach to identify the potential areas of improvement during the implementation of AR-based learning techniques and to determine the overall impact while revealing the linear and non-linear relationships of the features and the target variables. Four machine and deep learning models consisting of RNN, CNN, LSTM, and SVM have been built, implemented and evaluated for this project.

1.1 Research Problem Background

Numerous studies have been conducted on the impact analysis of AR-based education approach. However, they appear to be rudimentary in terms of analysis methods as in the studies by Belda-Medina and Marrahi-Gomez (2023) and O'Connor and Mahony (2023), which involved analysis using basic statistical and mathematical methods. Some researches appeared to have inadequate input data collection and sample selection techniques in terms of data diversity and dimensionality such as the studies by Belda-Medina and Marrahi-Gomez (2023) and O'Connor and Mahony (2023). Few other researches focused only on binary outcome analysis like positive or negative impact such as the studies by Chang et al. (2022), Garzón et al. (2020) and Cao and Yu (2023) rather than a broad spectrum of outcomes.

1.2 Motivation

Incorporation of AR techniques in education is a rapidly growing technique that has great potential to study and analyse the factors involved in a successful implementation of AR-based learning. The analysis techniques of the previous studies urges the need for developing robust methods to efficiently learn and analyse this subject area. This will facilitate not only to focus on the specific areas of improvements which can be addressed but also the technical, infrastructural and attitude challenges in the domain and reveal hidden patterns during the process. As the current methods of analysis are mostly basic, this research will incorporate novelty in the analysis technique by laying the foundation for robust technologies like machine learning in the field for future analyses.

1.3 Research Question

To what extent does the underlying individual factors such as age, gender, demographic, language and so on, impact the student attitude, and self-efficacy towards the AR-based learning approach and influence the outcomes?

1.4 Research Objectives

Firstly, this project will not only perform impact analysis of the AR-based learning approach, but also the input factors that influence these outcomes will be determined. Secondly, it uses machine learning models for impact analysis rather than using naive techniques which is well suited for real-world problems as in this case. Lastly, the usage of an extensive dataset comprising of diverse sample for efficient analysis will eliminate the data inadequacy issues of the previous studies. Usage of machine learning techniques, will also be helpful to determine the multivariate and non-linear relationship between the input and target variables.

1.5 Benefits of the Proposed Solution

The solution will contribute greatly in handling the existing and future problems in the field of AR-based education approach and impact analysis. It does so by creating robust prototypes of highly efficient models which are reusable and interoperable. The models have yielded excellent reproducibilty indicating the high efficiency of the models. The project also addresses the set backs in the previous studies. Apart from the contributions to the domain, the research project has also revealed insightful data and patterns as to what and how certain factors play a role in influencing the attitude and self-efficacy as a result of the AR-based learning.

1.6 Structure of the Document

The structure of this document begins with abstract stating a crisp summary of the project, followed by introduction, review of related works, methodology, design, implementation, evaluation of the results, discussion of project specific objectives and limitations, and wrapping with conclusion and future works.

2 Related Work

A detailed analysis and review of the existing studies is crucial for every research to determine its objectives and existing trends in the field. The detailed review of studies related to this project has been analysed and the same has been discussed in the upcoming sections.

2.1 Current AR Techniques and Comparative Reviews

Projects by Dirgantara Deha et al. (2023) and Nazeer et al. (2023) show the importance of AR based education as they develop exciting prototypes applicable to real-life learning experiences. It is crucial to study how the AR based learning systems impacts the users. Comparative studies by Escobedo et al. (2014) and Radu (2012) show the various factors to consider before implementing the AR based education. Radu (2012) reveals several useful factors to consider for success of AR based education approach. Systematic reviews by Buchner et al. (2022) suggests the consideration of other factors like cognition load and external factors that might impact the outcomes.

2.2 A Different Vantage Point on the Implementation of AR based Education

While most of the studies analyse, the impacts of AR in education, this study by Kusuma Zamahsari et al. (2024), reviews the challenges of setting up an AR based learning environment. It is quite interesting as this study addresses the environmental, technical, usage and adaptability hurdles while incorporating an AR environment. It is important to take the all the challenging factors into account before implementing an AR based learning environment and this study illustrates few of the challenges that should be taken into consideration.

2.3 Limitations and Gap of the Previous Researches

Upon review of these 3 meta-analysis based studies by Chang et al. (2022), Garzón et al. (2020) and Cao and Yu (2023), it is evident that all of these methods ignored the underlying factors. While these papers did an elaborated study on a great number of papers, they either conclude the study by stating that AR based education has positive impacts or moderately positive impacts. These meta-regression models have their limitations. They fail to analyse other potential factors like environmental set up challenges, personal influences such as skills, capability and interests and any other external factors that might affect the efficiency of the performance and outcomes.

Using sequential mixed methods, this study Belda-Medina and Marrahi-Gomez (2023) have examined 3 hypotheses - students' attitude, vocabulary acquisition and students' motivation. Tested with 130 9th-grader students aged between 14 and 15, the analysis yielded positive results. Although being statistical method, it lacks in consideration of the underlying factors like gender and age. Additionally, the sample size is not widely categorized focusing only a specific age group which makes the study insufficient. Another study by O'Connor and Mahony (2023) using Partial Least Square (PLS) Structural Equation Modelling (SEM) methods with 65 students using an online questionnaire, faces a similar issue. The sample size of the participants and the questionnaire were inadequate. Despite getting desired positive results, the accuracy of these studies are not precise as many of the influencing factors are not taken into consideration. This paper by Kusuma Zamahsari et al. (2024) in particular, states intriguing perspective infrastructure an physical factors influencing the outcomes.

2.3.1 Contribution of Machine Learning in the Domain

The limitations are usage of rudimentary techniques like basic mathematical and statistical methods, and only analysing the generic binary outcomes like positive or negative impacts. In practice, to implement a successful AR based education, numerous factors will weigh in. Few examples are, knowledge of technology, accessibility to appliances, subject matter interest, and infrastructure challenges. Another limitation observed in these studies are the lack of adequate data both in terms of number of participants and data dimensionality. The contribution of machine learning to this domain is that this is an evolving technology that requires extensive analysis and machine learning is vital to handle real-world problems like this that involves large and complex datasets. Additionally, the input data shows non-linear patters and multivariate relationships that makes machine learning a suitable choice to identify the potential factors influencing the outcome of AR based education, as it is quite difficult to get the same insights through means of basic methods. Usage of machine learning also enables scalability, reusability and interoperability for future studies in the domain. In this research, all the limitations stated in the above section will be addressed as the aim of this research is to identify the various underlying factors that affect the AR based education approach and student response. Deep learning methods will be used to analyse any and all factors affecting these results and identify the potential areas of improvement. It will also open doors to the usage of robust techniques such as machine and deep learning approaches in the domain of AR based education.

3 Methodology

To begin with, the project is built based on KDD methodology. Out of multiple algorithms, RNN, CNN, LSTM, and SVM have been finalized to implement for this project. Here is the rationale behind choosing these algorithms. The input dataset consists entirely of numeric data. Upon analysis, the above methods tend to yield best results when handling with numeric data. The prediction in hand is a multi-class classification problem. The chosen methods have been suggested as the most effective methods to handle multi-class classification problems on numerical data. In this project, not only the success of the AR based learning is determined whether it is positive or negative but also the individual factors which has the most impact on the target outcome are also identified. In this case, the chosen deep and machine learning models have been quite efficient to assign class weights in order to determine the importance of the individual dependent variables. Amongst all the models used for research and development, the 4 models stated above yielded the highest accuracy and were efficient in handling multi-class classification problems. The Figure 1 shows the high level architecture of the research project. This will be explained in the following sections.

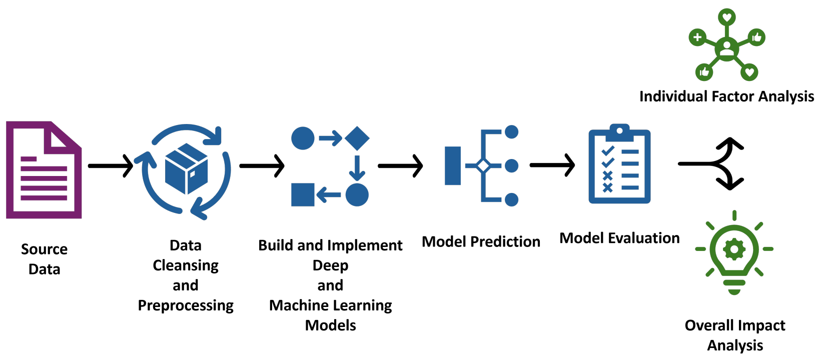

Figure 1: Methodology (Original Illustration)

3.1 Research Process

In this section, the overall process and procedures involved to implement the research will be discussed. The process consists of several phases like data collection, preprocessing, model implementation and training, model prediction, evaluation and interpreting and disclosing the findings.

3.1.1 Dataset Source and Description

The dataset used for this project is sourced from the URL - [1]. This is a comprehensive dataset collected for the Pilot 2 project by ARETE across 11 countries and 1988 participants to analyse the impacts of AR in STEM learning. The input data is an excel file with raw categorical data and the filename is 20230118_P2 Students Test_share_ids_info.xlsx. It has 412 columns and 1988 rows. Few of the columns are Student ID, Gender, Year of Birth, Country, and few other columns to collect the response of the students on their attitude and self efficacy towards Math and Science on

[1]Dataset Source: https://zenodo.org/records/7877072

before, after and memory retention criteria. There are 16 qualified features which can be used as input variables.

3.1.2 Data Cleansing and Preprocessing

Next step is to clean and preprocess the input data. This involves a identifying and handling null and missing values, handling invalid data, and any necessary preprocessing steps before feeding the data into the deep and machine learning models. For the dataset in hand, the following data cleansing and preprocessing step have been carried out.

3.1.3 Build and Implement the Models

The chosen models will be implemented for this project. The detailed implementation will be discussed in Section 5. The models will be built using Python. They will be trained with 80%-20% validation split.

3.1.4 Model Prediction

Once the models are implemented and trained, the will make predictions for the multi-class classification problem addressed in this project. Twelve target columns have been chosen for prediction.

3.1.5 Model Evaluation

The models will be evaluated individually based on several evaluation metrics like accuracy, precision, F1-score, and AUC-ROC score. They will also be assessed by cross-comparison to validate and compare the accuracy and efficacy of each in predicting the target variables. Individual independent variables will also to analysed to determine the impact of underlying factors on the outcome of this research.

3.1.6 Result Interpretation

Based on the results from each model, the most efficient model will be identified. The overall impact of AR based education on students to determine whether the impacts are positive or negative will be disclosed. The underlying factors that affect the outcome most will be enclosed. All the results and findings will be furnished transparently.

3.2 Evaluation Methodology

To evaluate the efficiency of the models, the metrics chosen are accuracy, precision, recall, F1-score, AUC-ROC score and confusion matrix as explained by **?** Although accuracy is one of the important metrics to determine the efficacy of the model, other metrics play vital part in evaluating the model. For instance, confusion matrix not only visualises the actual and predicted label depicting whether they match or not, but also shows the True Positive, True Negative, False Positive and False Negative values.

4 Design Specification

This section elicits the architecture of the models implemented in the project. For all the neural network models, ReLU has been chosen to activate the hidden layers as it handles the vanishing gradient problem well. To activate the output layer, 'softmax' has been used as it is well suited for multi-class classification problems. Additionally, stratified K-fold cross validation has been used for better performance. This will be explained in detail in the Section 5. Also several requirements including software, functional have been discussed.

4.1 Architecture of RNN

The RNN model built for this model is a sequential Simple RNN model. To build the sequential model, Sequential() function has been used. The input layer is defined with the shape of the input features which in this case is 16. A SimpleRNN layer is added using model.add(SimpleRNN) with 100 units. The model is built to retrieve the last output in the output sequence. A fully connected dense layer is added with 50 units. The activation function for the SimpleRNN and Dense layers is ReLU. Another dense layer which is the output layer with 4 target classes has been added for the multi-class classification. The high level architecture of the model is shown in the Figure 2.

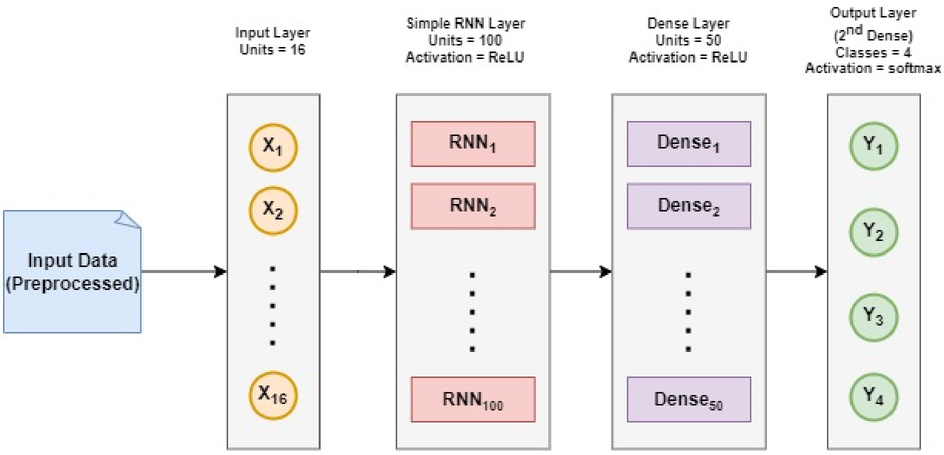

Figure 2: Architecture of the RNN Model (Original Illustration)

4.2 Architecture of CNN

The CNN model in this project has been initialized with Sequential() function to build a sequential model. The input layer has been defined with the shape of the input features. 1D convolutional layer has been added with model.add(Conv1D) with 64 units. To down sample the output from the 1D convolutional layer, a max pooling layer ha been added using model.add(MaxPooling1D). To flatten the output from the previous layer into 1D vector, a Flatten layer has been added using model.add(Flatten()). model.add(Dense) a fully connected dense layer with 50 units has been added. ReLU has been used as the activation function for 1D convolutional and fully connected dense layers. The output layer has been defined for 4 classes with softmax activation function. The high level depiction of the CNN model used in this project is shown in Figure 3.

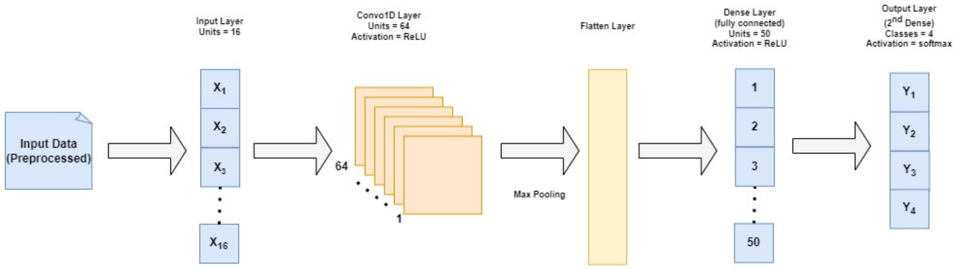

Figure 3: Architecture of the CNN Model (Original Illustration)

4.3 Architecture of LSTM

LSTM is a variant of the RNN model. In this project, bidirectional LSTM model has been built. After initializing the sequential model with Sequential() function, input layer with 16 units has been added. Bidirectional LSTM layer has been added using model.add(Bidirectional(LSTM) with 100 units. This returns all sequences of output. Another LSTM layer with 100 units to learn complex patterns has been added using model.add(LSTM). Using model.add(Dense) a fully connected dense layer with 50 units has been defined. All these layers use ReLU for the activation function. The output layer with 4 classes has been added using another dense layer. This layer uses softmax for activation. The architecture of the LSTM model for this project is shown in Figure 4.

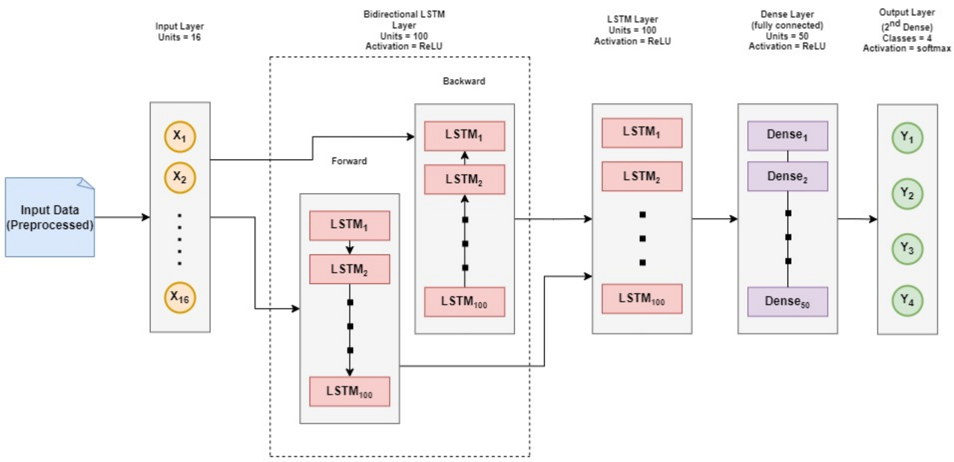

Figure 4: Architecture of the LSTM Model (Original Illustration)

4.4 Architecture of SVM

The SVM model for this project has been built with hyperparameter tuning to increase the accuracy. SVM has three hidden hyperparameters - C, gamma, and kernel. They cannot be manually tuned. Hence GridSearchCV has been used for hyperparameter tuning. GridSearchCV will train the models with all possible combinations of the hyperparameter tuning and retrieve the model with the best tuned hyperparameters. This model is then used for prediction and evaluation. The architecture of the SVM model is shown is the Figure 5.

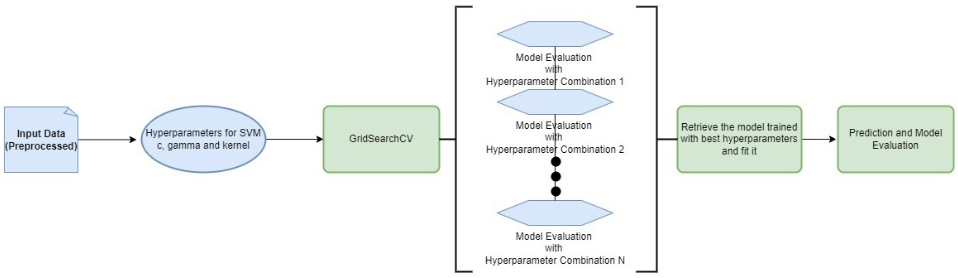

Figure 5: Architecture of SVM with Hyperparameter Tuning (Original Illustration)

4.5 Requirements

The three main functional requirements of this project aredata availability, code reusability and result reproducibility. It is vital that the chosen dataset has a comprehensive set of data that is as diverse as possible for fair analysis. The subsets of the data should also have sufficient data samples. Also, in this project, there are 6 target variables, all with exact same class labels that fall under multi-class classification. The models built must be reusable and the results should be consistent across multiple data samples indicating code reproducibility.

As for the non-functional requirements, data security must be ensured by using data masking techniques, sensitive data must be protected to prevent any breach of confidentiality and ethical implications. Appropriate measures and considerations should be taken into account while publishing the results. The results should be furnished with transparency and all factors must be evaluated without bias. The code should be scalable and interoperable which means it should able to perform well for huge volume of data and it must also be able to easily integrated across platforms.

5 Implementation

5.1 Implementation Details

For this research on impacts analysis to identify the underlying factors affecting the outcomes of AR based education, RNN, CNN, LSTM and SVM models have been implemented to predict six target variables. These variables are shown in the Table 1.

Table 1: Target Variables

Variable Name	**Description**
Pre_A_MAT_Enjoy_Learning	Pre-test: Attitude towards Mathematics - Enjoy Learning
Post_A_MAT_Enjoy_Learning	Post-test: Attitude towards Mathematics - Enjoy Learning
Ret_A_MAT_Enjoy_Learning	Retention: Attitude towards Mathematics - Enjoy Learning
Pre_SE_SCI_Usually_Do_Well	Pre-test: Self-Efficacy in Science - Usually Do Well
Post_SE_SCI_Usually_Do_Well	Post-test: Self-Efficacy in Science - Usually Do Well
Ret_SE_SCI_Usually_Do_Well	Retention: Self-Efficacy in Science - Usually Do Well

As the independent variables are same across the categories and all the six dependent variables have the exact same unique class labels that fall under the category of multi-class classification, most of the implementation steps are common for the models. The project has been implemented in Jupyter Notebook using Anaconda Navigator.

5.1.1 Data Import, Preprocessing and Analysis

In step 1, the input file is read into a dataframe using pandas. Basic data cleansing steps to check and handle duplicate data, null, missing and invalid values have been applied. The input data has many columns which are not required for the project. These columns have been dropped and rest of the columns have been renamed for easy understanding. The main dataframe has been copied into 6 dataframe with retaining only 16 qualified features as input variables and the respective target variable and rest of the variables have been dropped. The dataframe for the variable 'Ret_SE_SCI_Usually_Do_Well' has been shown as reference in Figure 6. This measure has been taken as all of these variables have different proportions of invalid data, and removing invalid data from the main dataset, will result it unnecessary data loss. Steps have been taken to ensure that the Pre, Post and Ret variables will have the same set of Student IDs for Math and Science respectively to ensure fair analysis. Once the preprocessing has been completed, these individual data-frames have been loaded into separate pickle files to reuse in other Jupyter Notebook files.

```
<class 'pandas.core.frame.DataFrame'>
Index: 354 entries, 27 to 1121
Data columns (total 17 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   TB                                  354 non-null    int64
 1   CNTY                                354 non-null    int64
 2   LANG                                354 non-null    int64
 3   Gender                              354 non-null    int64
 4   Year                                354 non-null    Int64
 5   Frequency_of_Language_Usage_at_Home 354 non-null    int64
 6   Own_Computer                        354 non-null    int64
 7   Shared_Computer                     354 non-null    int64
 8   Table                               354 non-null    int64
 9   Own_Room                            354 non-null    int64
 10  Internet_Connection                 354 non-null    int64
 11  Mobile_Phone                        354 non-null    int64
 12  Gaming_System                       354 non-null    int64
 13  Media_Use_Home                      354 non-null    int64
 14  Media_Use_School                    354 non-null    int64
 15  Media_Use_Other                     354 non-null    int64
 16  Ret_SE_SCI_Usually_Do_Well          354 non-null    int64
dtypes: Int64(1), int64(16)
memory usage: 50.1 KB
```

Figure 6: Dataframe for Ret_SE_SCI_Usually_Do_Well)

To analyze the data further, check the trends and determine class imbalance and distribution, the variables have been plotted using matplotlib.

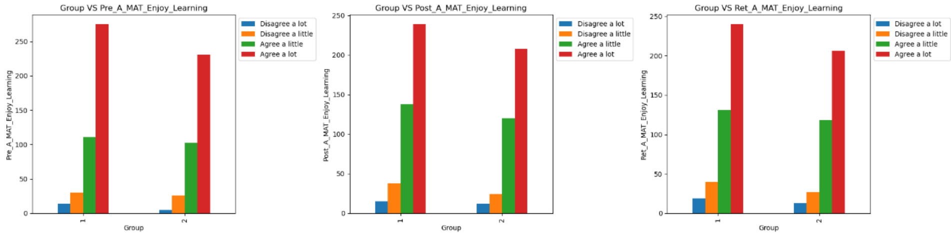

Figure 7: Group vs Attitude)

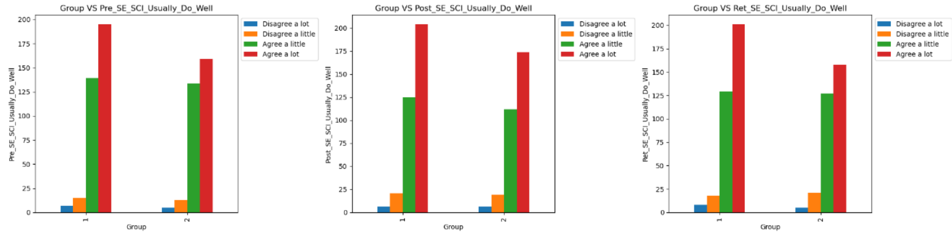

Figure 8: Group vs Self Efficacy)

The Figure 7 and Figure 8 show that students in the intervention group agree a lot more about their attitude towards enjoying the learning and self efficacy about doing well in the subject compared to the control group. This shows that the overall impacts of AR based education is positive.

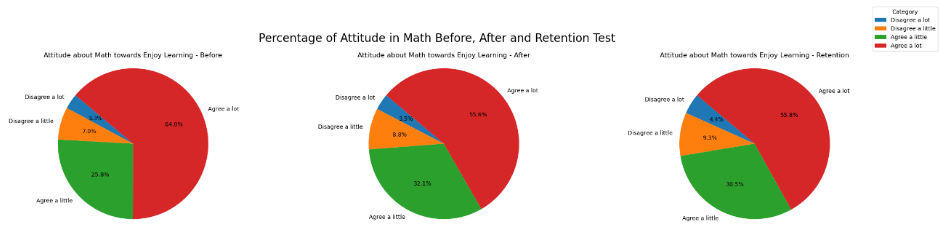

Figure 9: Class Distribution of Attitude

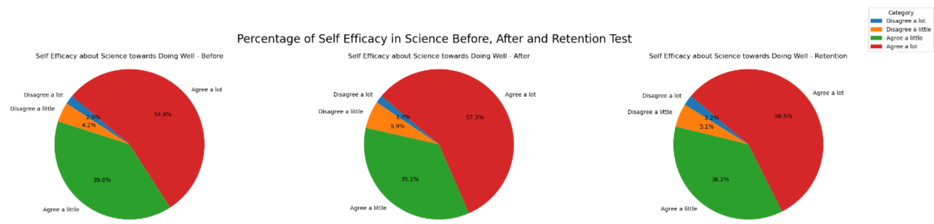

Figure 10: Class Distribution of Self Efficacy

The Figure 9 and Figure 10 depict that there the class label 4 for "Agree a lot" category is more indicating class imbalance.

5.1.2 Implementation of RNN, CNN and LSTM with Stratified K-Fold Cross Validation

To ensure code resuabilty, the steps are defined as functions. A Class for predictor has been initialized for each model. Random seed has been set to maintain uniformity accross the code for reproducibility. Tensorflow environment has been enabled for deterministic operations to maintain stability. To handle the class imbalance issue, data has been upscaled by using the highest count of the class which is 4 in this case. The unique class labels are 1, 2, 3 and 4 for the target variables. The in data in X has been scaled using StandardScaler() to adjust the range of the data. The data is reshaped into 3D format for all the models. The target variable is encoded using OneHotEncoder and reshaped into 2D array, as it has categorical values.

The RNN, CNN, and LSTM models are defined. The detailed design is explained in design section. To compile the model, the Adam optimizer has been chosen due to its adaptive learning rate and efficiency. As for the loss function, 'categorical_crossentropy' has been opted as it is best suited for classification problems and its ability to calculate the difference between the true label and the predicted probabilities thereby minimizing loss and increasing the efficiency of predictions. As 'accuracy' is the measure of the proportion between accurate predictions and total predictions and it is the basic metric to evaluate any model, it has been chosen the evaluation metric for the models.

Stratified K-Fold with 5 folds has been defined. The model is trained in 5 validation splits with EarlyStopping and ReduceLROnPlateau to avoid overfitting and stop the learning rate when the accuracy is stable. The models runs for 50 epochs and batch_size=32. Other hyperparameters like class_weights are generated in the code. The model with the best accuracy is identified, and the predictions are done using the model. The classification report, confusion matrix, AUC-ROC, model loss and accuracy and variable importance plot are generated to evaluate the metrics. To run the models for all variables, the predictor class is passed with appropriate input parameters like dataset name and variable name for code reusability.

5.1.3 Support Vector Machines (SVM)

Similar to LSTM model, model reusability has been ensured for building SVM model as well by defining the model as function. This SVM model is incorporated with hyperparameter tuning using GridSearchCV. The validation data has been separated into 3 folds for cross validation. The best hyperparameters identified for this model are C = 1, gamma = 1 and kernel = 'rbf'. The model has been trained using the 'best_estimator' by assigning the optimal parameters identified in the previous step. To predict, the functions are called for all 6 variables by passing the dataframe and target variable name. The model has been evaluated based on accuracy, precision, recall, F1-score and AUC-ROC, confusion matrix. Finally, variable importance plot has been implemented to show which of the input features have the most influence on the outcomes.

5.1.4 Hyperparameters and Optimization

The hyperparameter settings are shown in the Table 2.

Table 2: Hyperparameter Values and Tuning Range of the Models

Model	Hyperparameter Values	Tuning Range
RNN	random_state=9098 learning_rate=0.001 class_weight based on training data EarlyStopping : monitor='val_loss', patience=5, restore_best_weights=True ReduceLROnPlateau: monitor='val_loss', factor=0.5, patience=3	Default is 42 0.01, 0.001, 0.0001, 0.00001
CNN	random_state=9098 learning_rate=0.001 class_weight based on training data EarlyStopping : monitor='val_loss', patience=5, restore_best_weights=True ReduceLROnPlateau: monitor='val_loss', factor=0.5, patience=3	Default is 42 0.01, 0.001, 0.0001, 0.00001
LSTM	random_state=9098 learning_rate=0.001 class_weight based on training data EarlyStopping : monitor='val_loss', patience=5, restore_best_weights=True ReduceLROnPlateau: monitor='val_loss', factor=0.5, patience=3	Default is 42 0.01, 0.001, 0.0001, 0.00001
SVM	C=1 gamma=1 kernel=rbf	[0.1, 1, 10, 100] [1, 0.1, 0.01, 0.001] Linear, Polynomial, RBF, Sigmoid, Precomputed

5.2 Tools and Languages

The tools and languages used in this project are given in the Table 3

Table 3: Tools and Languages

Software, Tools, Libraries, and Packages	Usage
Anaconda Navigator 3	To access the IDE for Jupyter Notebook.
Jupyter Notebook	A web-based interactive computing platform to build, execute, and evaluate data analytics code in Python.
Python 3.12.4	Programming language versatile to develop Machine learning models.
pandas, numpy	Data Preprocessing.
matplotlib, seaborn, datetime, and itertools	Visualization.
sklearn	Data preprocess, build and evaluate machine learning models.
tensorflow	Build deep learning models.

6 Evaluation

The 4 models have been evaluated using standing evaluation metrics that are accuracy, precision, recall, F1-Score, AUC-ROC, training and validation accuracy and loss. Confusion matrix has also been plotted to validate the performance of the classification model. Each model has been used to predict all 6 target variables. As all the target variables have 4 unique class labels, the metrics of the models are displayed as tables for all the variables. Table 4 and Table 5 show the performance metrics of the RNN and CNN respectively. For LSTM and SVM models, the Table 6 shows the accuracy and average metrics for Precision, Recall and F1-Score. They are displayed in the order of their performance showing the best performing model first.

Table 4: Performance Metrics of RNN Model
tab:Metrics for LSTM and SVM

Variable	Accuracy	Class	Precision	Recall	F1-Score
Pre_A_MAT_Enjoy_Learning	1.00	1	1.00	1.00	1.00
		2	1.00	1.00	1.00
		3	1.00	1.00	1.00
		4	1.00	1.00	1.00
Post_A_MAT_Enjoy_Learning	0.98	1	1.00	1.00	1.00
		2	0.98	1.00	0.99
		3	0.96	1.00	0.98
		4	1.00	0.94	0.97
Ret_A_MAT_Enjoy_Learning	0.99	1	1.00	1.00	1.00
		2	0.98	1.00	0.99
		3	1.00	1.00	1.00
		4	1.00	0.98	0.99
Pre_SE_SCI_Usually_Do_Well	0.99	1	1.00	1.00	1.00
		2	1.00	1.00	1.00
		3	0.97	1.00	0.99
		4	1.00	0.97	0.99
Post_SE_SCI_Usually_Do_Well	0.97	1	1.00	1.00	1.00
		2	0.98	1.00	0.99
		3	0.95	0.95	0.95
		4	0.95	0.93	0.94
Ret_SE_SCI_Usually_Do_Well	0.98	1	1.00	1.00	1.00
		2	1.00	0.97	0.99
		3	0.95	0.97	0.96
		4	0.97	0.97	0.97

Table 5: Performance Metrics of CNN Model

Variable	Accuracy	Class	Precision	Recall	F1-Score
Pre_A_MAT_Enjoy_Learning	0.98	1	0.95	1.00	0.97
		2	1.00	1.00	1.00
		3	0.98	0.95	0.96
		4	0.98	0.96	0.97
Post_A_MAT_Enjoy_Learning	0.96	1	1.00	1.00	1.00
		2	0.96	0.98	0.97
		3	0.90	0.96	0.93
		4	0.98	0.89	0.93
Ret_A_MAT_Enjoy_Learning	0.97	1	1.00	1.00	1.00
		2	0.96	1.00	0.98
		3	0.94	1.00	0.97
		4	1.00	0.90	0.95
Pre_SE_SCI_Usually_Do_Well	0.98	1	1.00	1.00	1.00
		2	1.00	1.00	1.00
		3	0.97	0.95	0.96
		4	0.95	0.97	0.96
Post_SE_SCI_Usually_Do_Well	0.98	1	1.00	1.00	1.00
		2	0.98	1.00	0.99
		3	0.95	0.98	0.96
		4	0.97	0.93	0.95
Ret_SE_SCI_Usually_Do_Well	0.98	1	1.00	1.00	1.00
		2	1.00	0.97	0.99
		3	0.95	0.97	0.96
		4	0.97	0.97	0.97

Table 6: Performance Metrics for LSTM and SVM Models

Model	Variable	Accuracy	Precision	Recall	F1-Score
LSTM	Pre_A_MAT_Enjoy_Learning	0.99	0.99	0.99	0.99
	Post_A_MAT_Enjoy_Learning	0.98	0.98	0.98	0.98
	Ret_A_MAT_Enjoy_Learning	0.97	0.87	0.88	0.87
	Pre_SE_SCI_Usually_Do_Well	0.96	0.97	0.96	0.96
	Post_SE_SCI_Usually_Do_Well	0.90	0.88	0.88	0.88
	Ret_SE_SCI_Usually_Do_Well	0.97	0.90	0.90	0.90
SVM	Pre_A_MAT_Enjoy_Learning	0.97	0.97	0.97	0.97
	Post_A_MAT_Enjoy_Learning	0.94	0.95	0.94	0.94
	Ret_A_MAT_Enjoy_Learning	0.93	0.94	0.93	0.93
	Pre_SE_SCI_Usually_Do_Well	0.91	0.92	0.91	0.91
	Post_SE_SCI_Usually_Do_Well	0.95	0.95	0.95	0.94
	Ret_SE_SCI_Usually_Do_Well	0.91	0.91	0.91	0.91

Amongst the algorithms implemented, RNN outperformed the other models with an overall 99% accuracy followed by CNN with 98% accuracy. These models were also consistent with the 6 different sub-datasets and target variables used in this project proving result reproducibility. Although LSTM yielded 96% accuracy which is good, it was inconsistent. The reason why RNN and CNN performed well over LSTM which is considered more efficient in general is that simple RNN has fewer parameters making it less susceptible to overfitting and one-dimensional CNNs are effective at capturing spatial hierarchies and local patterns directly within the data. The lack of long-term dependency in the data may also cause bidirectional LSTM to perform inadequately. Additionally, the computational cost in terms of memory and utilization is higher for LSTM compared to simple RNN and one-dimensional CNN models. This might affect the accuracy of the LSTM model while in contrast, the low computational cost of RNN and CNN models enabled them to yield higher and consistent performance. On the other hand, SVM performed with 94% accuracy, however, it has considerable difficulty in handling high dimensional data, therefore producing less accurate and inconsistent predictions. The Figure 11 shows the K-fold cross validation results with 5-fold in RNN model for one of the target variables – 'Pre_A_MAT_Enjoy_Learning'. It also shows the implementation of early stopping in the model.

Classification Report for Fold 1:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	55
2	1.00	1.00	1.00	55
3	0.84	0.89	0.87	55
4	0.88	0.84	0.86	55
accuracy			0.93	220
macro avg	0.93	0.93	0.93	220
weighted avg	0.93	0.93	0.93	220

Classification Report for Fold 2:

	precision	recall	f1-score	support
1	0.65	0.58	0.62	55
2	0.60	0.55	0.57	55
3	0.62	0.33	0.43	55
4	0.39	0.65	0.49	55
accuracy			0.53	220
macro avg	0.57	0.53	0.53	220
weighted avg	0.57	0.53	0.53	220

Classification Report for Fold 3:

	precision	recall	f1-score	support
1	0.76	0.76	0.76	55
2	0.61	0.76	0.68	55
3	0.50	0.36	0.42	55
4	0.57	0.58	0.58	55
accuracy			0.62	220
macro avg	0.61	0.62	0.61	220
weighted avg	0.61	0.62	0.61	220

Classification Report for Fold 4:

	precision	recall	f1-score	support
1	0.64	0.85	0.73	55
2	0.60	0.69	0.64	55
3	0.64	0.25	0.36	55
4	0.54	0.60	0.57	55
accuracy			0.60	220
macro avg	0.60	0.60	0.58	220
weighted avg	0.60	0.60	0.58	220

Classification Report for Fold 5:

	precision	recall	f1-score	support
1	0.57	0.84	0.68	55
2	0.69	0.49	0.57	55
3	0.37	0.25	0.30	55
4	0.51	0.58	0.54	55
accuracy			0.54	220
macro avg	0.54	0.54	0.52	220
weighted avg	0.54	0.54	0.52	220

Figure 11: Classification Report Matrix for Attitude in Math

6.1 Case Study 1 - Attitude in Math towards Enjoy Learning

RNN model has been chosen for detailed discussion due to its high efficiency and performance amongst the models for all variables. This has been explained in Section 6. The classification reports for Pre_A_MAT_Enjoy_Learning, Post_A_MAT_Enjoy_Learning and Ret_A_MAT_Enjoy_Learning have been shown in Figure 12. As depicted in the figure, for the variable Pre_A_MAT_Enjoy_Learning, the RNN model has overall accuracy of 100% and the precision, recall and F1-score across all classes is 100%. The precision, recall and F1-score for each of the classes is also 100%. This indicates that the model has classified all the classes accurately. For the variable Post_A_MAT_Enjoy_Learning, overall accuracy is 98%. Also the precision, recall and F1-score is 98%. As for the overall accuracy is 99% and the precision, recall and F1-score is also high as 99% indicating that the model is well-balanced.

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	55
2	1.00	1.00	1.00	55
3	1.00	1.00	1.00	55
4	1.00	1.00	1.00	55
accuracy			1.00	220
macro avg	1.00	1.00	1.00	220
weighted avg	1.00	1.00	1.00	220

Pre_A_MAT_Enjoy_Learning

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	48
2	0.98	1.00	0.99	48
3	0.96	1.00	0.98	48
4	1.00	0.94	0.97	47
accuracy			0.98	191
macro avg	0.98	0.98	0.98	191
weighted avg	0.98	0.98	0.98	191

Post_A_MAT_Enjoy_Learning

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	48
2	0.98	1.00	0.99	48
3	1.00	1.00	1.00	48
4	1.00	0.98	0.99	48
accuracy			0.99	192
macro avg	0.99	0.99	0.99	192
weighted avg	0.99	0.99	0.99	192

Ret_A_MAT_Enjoy_Learning

Figure 12: Classification Report Matrix for Attitude in Math

The confusion matrix in Figure 13 depicts that the RNN model is performing exceptionally well across all classes, with only a small number of incorrect classifications occurring in class 4 for the variables Post_A_MAT_Enjoy_Learning and Ret_A_MAT_Enjoy_Learning.

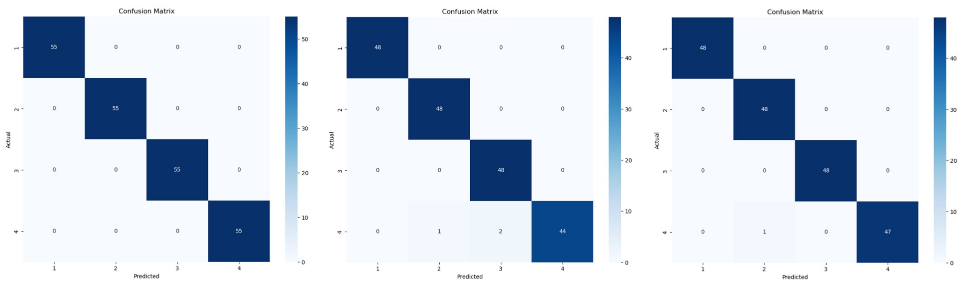

Figure 13: Confusion Matrix for Attitude in Math

The training and validation accuracy for the variables in Figure 14 indicate that the model is learning and improving consistently while Figure 15 shows that the loss has been minimized steadily.

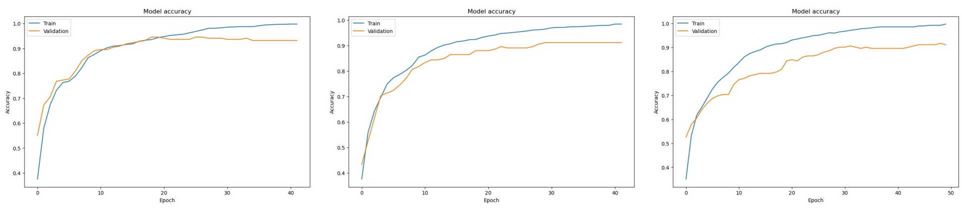

Figure 14: Model Accuracy for Attitude in Math

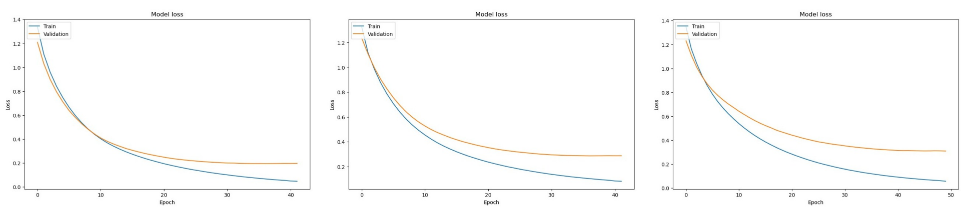

Figure 15: Model Loss for Attitude in Math

The model has a perfect AUC-ROC score of 1 for all the classes. This indicates that the model is able to distinguish the classes accurately. ROC for all the classes in each variable in shown in the Figure 16.

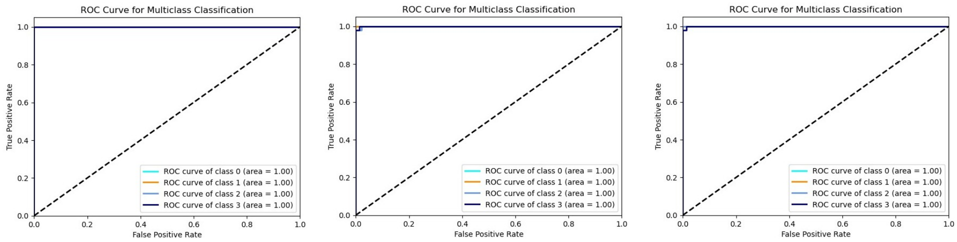

Figure 16: ROC Curve for Attitude in Math

Finally, to identify the features influencing the outcome of the AR based education, variable importance plot has been plotted. This plot in the Figure 17 shows that the year of birth that is age has the highest influence in the outcomes.

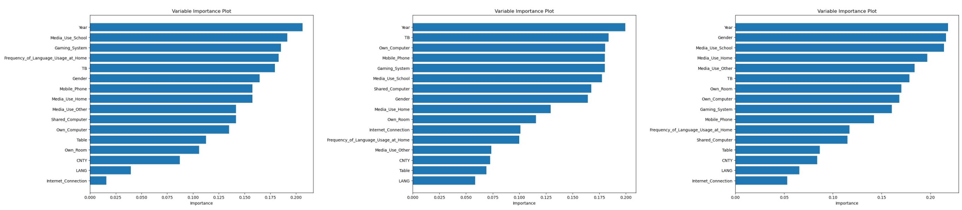

Figure 17: Variable Importance for Attitude in Math

The next step is to determine how age influences the outcome of the attitude. To analyse this, year has been converted to age and plotted against the target variables. The Figure 18 shows that the age group of 13 has better attitude towards Math followed by age group of 12.

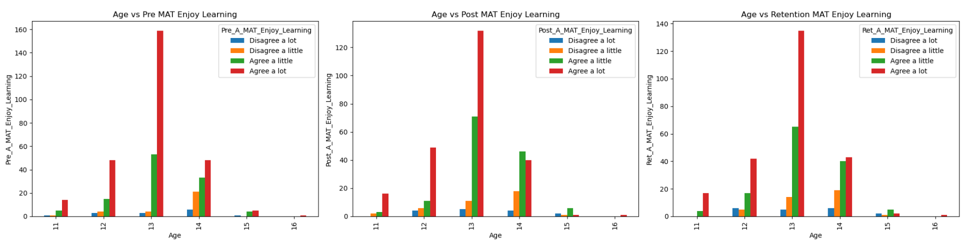

Figure 18: Age vs Self Attitude

6.2 Case Study 2 - Self Efficacy in Science towards Doing Well in Exams

The model which yielded the highest performance to determine the self efficacy in Science is RNN. The variables chosen to analyse the self efficacy are Pre_SE_SCI_Usually_Do_Well, Post_SE_SCI_Usually_Do_Well and Ret_SE_SCI_Usually_Do_Well. Beginning with the classification report in Figure 19, the overall accuracy is 99%. The precision, recall and F1-score across the classes and for each of the classes is 99% for the variable Pre_SE_SCI_Usually_Do_Well. The value of all the metrics is 97% and 98% for the variables Post_SE_SCI_Usually_Do_Well and Ret_SE_SCI_Usually_Do_Well respectively. This shows that the model has excellent performance.

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	39
2	1.00	1.00	1.00	39
3	0.97	1.00	0.99	39
4	1.00	0.97	0.99	39
accuracy			0.99	156
macro avg	0.99	0.99	0.99	156
weighted avg	0.99	0.99	0.99	156

Pre_SE_SCI_Usually_Do_Well

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	41
2	0.98	1.00	0.99	41
3	0.95	0.95	0.95	41
4	0.95	0.93	0.94	40
accuracy			0.97	163
macro avg	0.97	0.97	0.97	163
weighted avg	0.97	0.97	0.97	163

Post_SE_SCI_Usually_Do_Well

Final Model Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	40
2	1.00	0.97	0.99	40
3	0.95	0.97	0.96	40
4	0.97	0.97	0.97	40
accuracy			0.98	160
macro avg	0.98	0.98	0.98	160
weighted avg	0.98	0.98	0.98	160

Ret_SE_SCI_Usually_Do_Well

Figure 19: Classification Report Matrix for Self Efficacy in Science

The confusion matrix in Figure 20 shows that class 1 has been predicted correctly. However, there are some inaccurate classifications in the classes 2, 3 and 4.

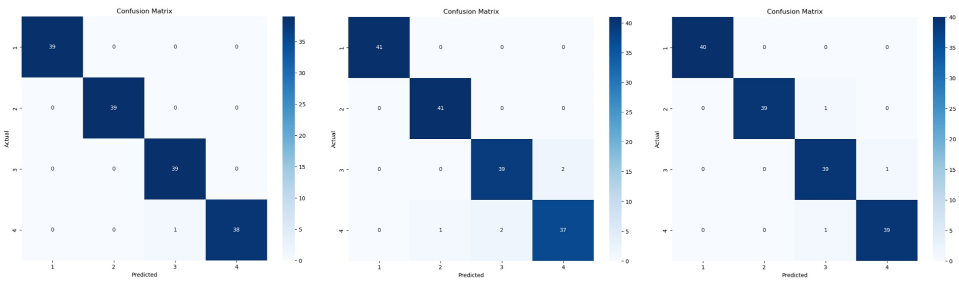

Figure 20: Confusion Matrix for Self Efficacy in Science

As for the training and validation accuracy and loss shown in Figure 21 and Figure 22 indicate that the model is persistently learning and improving while constantly minimizing the loss.

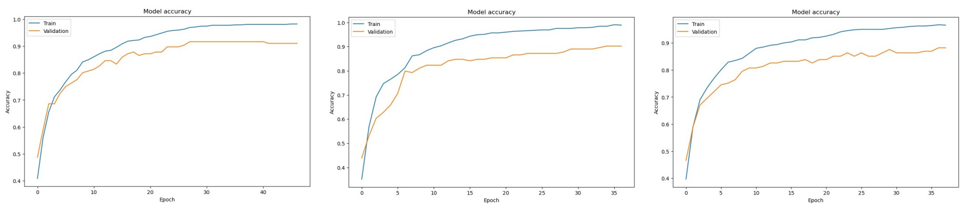

Figure 21: Model Accuracy for Self Efficacy in Science

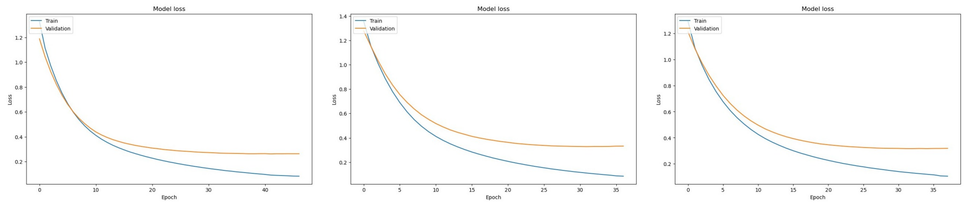

Figure 22: Model Loss for Self Efficacy in Science

AUC-ROC score for each class shown in Figure 23 is 1 indicating that the model performs well with respect to classifying the true and false positives for each variable and class.

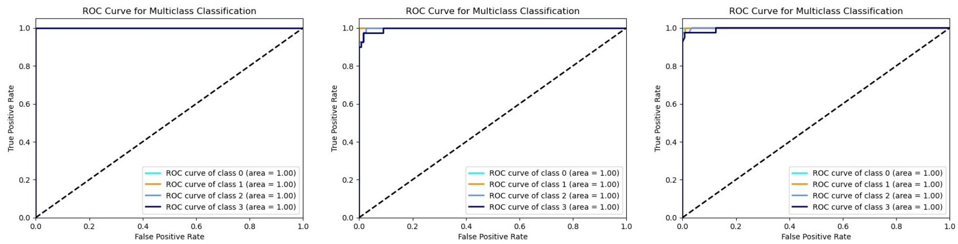

Figure 23: ROC Curve for Self Efficacy in Science

Lastly, Figure 23 illustrates that the most influencing factor for the variables Pre_SE_SCI_Usually_Do_ Post_SE_SCI_Usually_Do_Well and Ret_SE_SCI_Usually_Do_Well are Test Book(TB) which has different test books for Math and Science, Media Use in School and Gender.

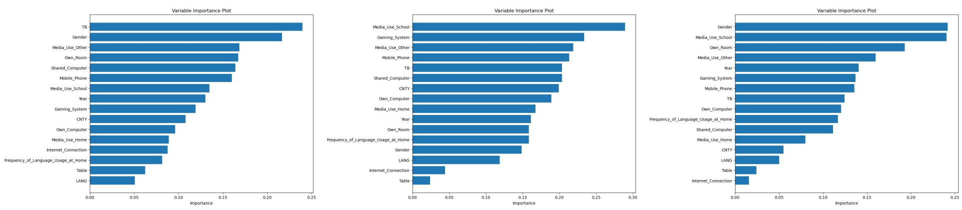

Figure 24: Variable Importance for Self Efficacy in Science

To check exactly how these factors influence the outcome, the input variables have been plotted as grouped bar chart against the target variables for further analysis.

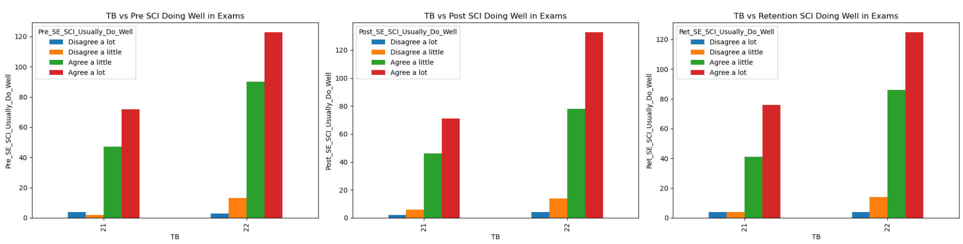

Figure 25: Test Book vs Self Efficacy

In the Figure 25 test book (TB) has been plotted against the target variables. The test book for science is S1 and S2. Although there is no clear indication of what these mean, the 'Agree a lot' class on self efficacy is higher for S2. It could be deduced as for test book S2 the self efficacy has increased.

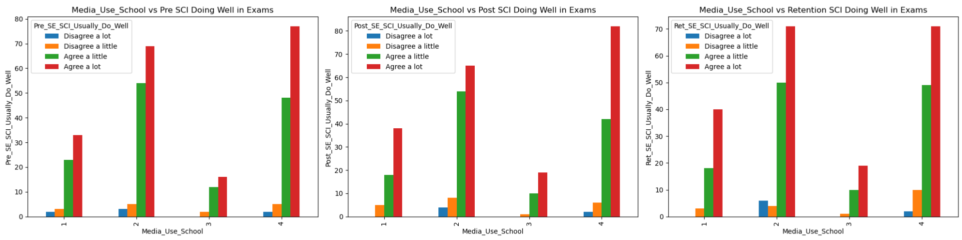

Figure 26: Media Use School vs Self Efficacy

The Figure 26 shows Media use in school plotted against the target variables. Values 1-4 in the input variable indicates the frequency of use where 1 being the highest. The bar charts illustrates that when the usage of media is almost never, the self efficacy increases followed by moderate usage have the second highest impact for increase in self efficacy.

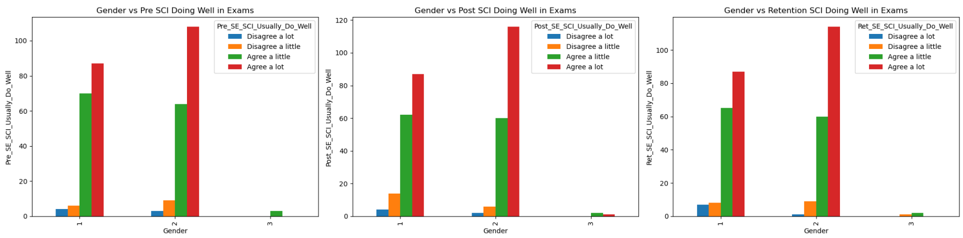

Figure 27: Gender vs Self Efficacy

The comparison of Gender versus Efficacy in Figure 27 shows that almost both gender show high efficacy rates, males show a slightly increased efficacy.

6.3 Discussion

To begin with, almost all the models yielded high performance, RNN performed well in comparison, followed CNN. Although, RNN gave 100% on all the metrics for the first variable, indicating that there might be an issue of over-fitting, the results were consistent when validated with different datasets, variables and sets of class distribution as shown in Table 4. From this, it could be concluded that the model performs exceptionally well on different dataset and it could be safely stated that there is no issue of over-fitting. The average accuracy of the model is 99%. Next, CNN model performed almost as good as RNN with a mere difference of 1% consistently accurate for all the variables yielding average accuracy of 98% as illustrated in Table 5. Though LSTM and SVM model gave a good average accuracy of 96% and 94% respectively, they performance was inconsistent for different variables as shown in the Table 6.

Next, while determining the factors influencing the outcomes of AR based education, each model showed a different input variable. It was inconsistent across the models. The assumption is to take the output from the best performing model and determine the impact. Additionally, the overall impact has also been analysed. Figure 7 and Figure 8 depict that intervention group, the one that received the AR based education, shows higher positive outcome on attitude and self efficacy in contrast to the control group.

Although the models were highly efficient in handling the multi-class classification problem, the input data appears to have too many redundant variables to determine the attitude and self efficacy attributes for example, the attitude variables on "I Like" and "Enjoy Learning" are almost similar. Due to large number of target variables, and imbalance between the test samples for the pre, post and retention attributes there was a significant data loss while handling these data discrepancies in preprocessing. As a result, there was inconsistency in determining the individual factor influence. Although, on comparison, an idea of the most influencing factors can be determined. The significant demographic factors are age and gender. Similarly, infrastructural dependencies such as internet connection, and availability of own or shared computer have also been observed. Also the overall attitude and self efficacy was higher for Science compared to Math. This can be viewed in Figure 9 and Figure 10. In conclusion, despite some minor limitations, the research project addressed its objectives of selecting adequate dataset which was observed as inadequate in the studies by Chang et al. (2022), Garzón et al. (2020), and Cao and Yu (2023), sample selection issue which focused on the same age group, performing impact analysis using machine learning models in contrast to the basic statistical methods conducted by Belda-Medina and Marrahi-Gomez (2023) and O'Connor and Mahony (2023), identifying the overall impact, and individual factors influencing the outcomes of AR based learning which most of the studies lacked. This research project has addressed the inadequacies of the previous studies while laying the foundation for future studies in the field of AR based education using machine learning.

7 Conclusion and Future Work

In conclusion, the research question to identify to what extent does the underlying individual factors such as age, gender, demographic, language and so on, impact the student attitude, and efficacy towards the AR based learning approach and influence the outcomes have been identified at average level. The main objectives of this project are to address the setbacks of the previous researches that were lack of adequate sample data, incorporate machine learning models to analyse the impacts and identify the underlying factors influencing the outcomes of AR based learning in order to address those factors while implementing AR based education approach. In this project using efficient machine learning models which include neural networks like RNN, LSTM and CNN and also common classification models like SVM, 6 variables with respect to attitude and self efficacy were classified and the factors influencing the outcomes has been identified.

Although the input dataset is extensive, it is redundant. In the future, the dataset can be normalized to avoid the redundancy. This will eliminate the data loss issue and inconsistency in identifying the influencing factors. This study was focused on 6 target variables categorized into two categories. In future, other variables on attitude and self efficacy can be analysed and a more comprehensive conclusion can be derived with the results.

The main objective of this study is to identify the potential factors that influence the outcome of the AR based learning. In that regard, factors like age, gender, test book, and media use in school has high influence on the outcomes. A certain category of students belonging to the age group of 13 showed high agreement towards enjoying learning through AR. This paves the way for further analysis as to why the other age groups were less interested and focus on improving the same. Similarly, female students did not show much improvement in self-efficacy after the AR based learning although, male students did. This is yet another area of investigation. Likewise, the students' attitude and self-efficacy towards science was better when compared to math. This exhibits a non-linear pattern where the effect of AR on attitude and self-efficacy is much stronger in subjects of high interest in this case, science, but minimal in subjects of low interest like math. For instance, if a student loves science, AR tools that enhance learning in this area might greatly increase their positive attitude and self-efficacy, but the same tools might not have much impact if used in a subject they find uninteresting. There were other factors that directly impacted the outcome like availability of internet connection which comes under infrastructure and technical accessibility. Students with better access to technology had higher positive attitude. In this research some potential factors that influence the outcome of AR based learning have been identified and how certain patterns of these factors affect the outcome have been revealed. Additionally, the overall impact of the AR based learning shows higher positive attitude and self-efficacy in the intervention group that received the AR based education in comparison to the control group that received conventional learning methods. The machine learning algorithms in this project also lay the foundation for future research in the field of AR based learning by providing code reusability, result reproducibility, interoperability and scalability that serves as a great value addition.

References

Belda-Medina, J. and Marrahi-Gomez, V. (2023). The impact of augmented reality (ar) on vocabulary acquisition and student motivation, *Electronics* **12**(3).
URL: *https://www.mdpi.com/2079-9292/12/3/749*

Buchner, J., Buntins, K. and Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review, *Journal of Computer Assisted Learning* **38**(1): 285–303.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12617*

Cao, W. and Yu, Z. (2023). The impact of augmented reality on student attitudes, motivation, and learning achievements—a meta-analysis (2016–2023), *Humanities and Social Sciences Communications* **10**(352): 1–12.

Chang, H.-Y., Binali, T., Liang, J.-C., Chiou, G.-L., Cheng, K.-H., Lee, S. W.-Y. and Tsai, C.-C. (2022). Ten years of augmented reality in education: A meta-analysis of (quasi-) experimental studies to investigate the impact, *Computers Education* **191**: 104641.
URL: *https://www.sciencedirect.com/science/article/pii/S0360131522002123*

Dirgantara Deha, M. T., Santoso, R. J. and Tho, C. (2023). Implementation of augmented reality for solar system subject in primary school, *2023 International Conference on Information Management and Technology (ICIMTech)*, pp. 714–719.

Escobedo, L., Tentori, M., Quintana, E., Favela, J. and Garcia-Rosas, D. (2014). Using augmented reality to help children with autism stay focused, *IEEE Pervasive Computing* **13**(1): 38–46.

Garzón, J., Kinshuk, Baldiris, S., Gutiérrez, J. and Pavón, J. (2020). How do pedagogical approaches affect the impact of augmented reality on education? a meta-analysis and research synthesis, *Educational Research Review* **31**: 100334.
URL: *https://www.sciencedirect.com/science/article/pii/S1747938X19303525*

Kusuma Zamahsari, G., Nur Amalia, M., Rifah, L., Permana, F., Romadhon, S. and Prihatini, A. (2024). A systematic review in educational settings: Numerous challenges to the adoption of augmented reality, *2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pp. 1–6.

Nazeer, S., K B, S. and Roobini, M. (2023). Interactive learning using mobile based augmented reality application, *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–7.

O'Connor, Y. and Mahony, C. (2023). Exploring the impact of augmented reality on student academic self-efficacy in higher education, *Computers in Human Behavior* **149**: 107963.
URL: *https://www.sciencedirect.com/science/article/pii/S074756322300314X*

Radu, I. (2012). Why should my students use ar? a comparative review of the educational impacts of augmented-reality, *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 313–314.