International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Proposed Hybrid CNN-RNN Architecture for Student Performance Prediction

Sam Yit Xiong[1], Esraa Faisal Malik Gasim[2], Chew XinYing*[3], Khaw Khai Wah[4], Lee Ming Ha[5]

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: Education is one of the important factors for the development of a country, thus, an early prediction system to predict the student performance is needed. With such system, education institutes would have a strong capacity to detect slow learners and investigate the major variables impacting their academic performance and preventing students from dropping out of school owing to poor marks or failing the exam. There are several studies that uses traditional machine learning to predict student performance, however, hybrid deep learning in predicting student performance has never been reported. Whereas machine learning requires feature selection, deep learning models may conduct automatic feature selection in the training model. However, it may suffer from the curse of dimensionality as it collects more and more features from student data. This paper proposes a hybrid deep learning model with combination of Convolutional Neural Network (CNN) and Recurrent neural network (RNN), i.e. CNN-RNN, where CNN captures the local dominant features and reduce the curse of dimensionality and RNN obtains the semantic correlation between features. The results of the experiments indicate that the hybrid CNN-RNN prediction model performs better than deep learning model by 3.16%, where the accuracy increased from 73.07% in ANN to 79.23% in hybrid CNN-RNN.

Keywords: *Convolutional Neural Network, Recurrent neural network, CNN-RNN, Hybrid Deep Learning, Education, Student Performance Prediction*

1. Introduction

Education is critical for a country's long-term economic development as talent is required to move the country forward. Educational organization are one-of-a-kind and play a crucial role in the growth of any country [1]. Thus, a new term called educational data mining (EDM) has raised. EDM is an emerging discipline that uses data gathered from educational environments to uncover relevant and valuable knowledge. By utilizing data mining (DM) methods, EDM offers researchers with a deeper knowledge of student behaviour and the environments in which learning occurs [2]–[4].

With DM, education practitioner can find out what is the key factors affecting student's performance. By doing so, a quick action can be taken to assist students, improve educational quality, and boost school resource management. Furthermore, academic success is influenced by a variety of circumstances and everyone has a varied learning ability as a result of their backgrounds [5].

Accurately evaluating student's performance and finding the relationship between collected data and student performance is a challenging task due to variety of sources of educational data. Additionally, various teaching strategies are needed for students with different backgrounds. By analyzing single variate, the data pattern cannot be completely known, and analysing multivariate is challenging by using traditional method. Therefore, a more robust approach for making predictions based on multivariate input is required. All the data collected holds some trends and pattern that are important for the prediction, with limited human expert, the important details may be overlooked.

Each institute collecting different set of features and researchers used to spend more efforts just to extract the important indicators. According to [6] having a deep learning technique that is less required feature selection will help in the domain. As stated by [7], convolutional net which is composed of a feature detection layer, has demonstrated the ability of CNN in feature extraction for image data. Based on the current research, there are machine learning and deep learning methods being introduced to predict student performance. However, there is room for improvement by implementing more recent technique to capture the pattern of the data to generate a better performing model [8].

Several methods such as statistical, clustering, association relationships and prediction has been used in the field of EDM. However, Statistical approaches may not always be enough for establishing the relationship between numerous factors and student performance [9]. On the other hand, classification method has been widely used EDM approach. historical data such as student grades, and demographic data are used to forecast the future outcomes of the students' results in which classification methods to improve decision making and optimize success for the student such as Decision Trees (DT), Random Forest (RF) , Naïve Bayes (NB), Support Vector Machine (SVM), Artificial

[1] School of Computer Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, MALAYSIA
ORCID ID : -
[2] School of Management, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, MALAYSIA
ORCID ID : -
[3] School of Computer Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, MALAYSIA
ORCID ID : 0000-0001-5539-1959
[4] School of Management, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, MALAYSIA
ORCID ID : 0000-0003-2646-6477
[5] Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak Campus, Sarawak, MALAYSIA
ORCID ID : 0000-0003-2096-5865
** Corresponding Author Email: xinying@usm.my*

Neural Network (ANN) Logistic Regression models (LR), and K nearest neighbours (KNN) techniques have been employed [10]–[25]. Cleary shown from the literature, NN was the most common approach used for predictions of student's result and has shown effective results. With more focus on few of the literature, [26] used fuzzy c-mean to assess student performance by clustering to visualize and cluster ancestors. The clustering approach may be used to find valuable information and patterns in educational data, instead of relying simply on examination grades.

While [5] utilized feed forward neural network to predict the GPA after the first year of study. This research used one input layer, two hidden layer and one output layer. The input data were; type of study program, gender of the student, high-school graduation GPA, age of the student, difference in years from the moment the student high school until the student enrolls in university. The training algorithm used was resilient backpropagation, modifies the rule of weight updating and reverts only weight updates that have caused sign changes of the partial derivate and an error increase. This rule combines information about the sign of the error function derivative which is error surface information with the magnitude of the network error when the decision of reverting an update step is taken.

[27] implemented ANN with Encog framework which was able to achieve 86% accuracy. This author stated that data collection is a challenging task for mid-size universities as they have fewer student's record for analysis. The primary goal of this study is to demonstrate the feasibility of training and modelling a limited dataset size, as well as the feasibility of developing a prediction model with acceptable accuracy. [6] mentioned that identifying the key indication is especially essential when the data is restricted as extraneous data might lead the model to overfit and become unusable. On the other hand, the heat map visualization and hierarchical clustering methods are used to identify the main indicator that could help in predicting dissertation and courses grades [27].

In [6] research, the machine learning techniques being used is MLP-ANN, NB, SVM with radial kernel function (RBF), KNN and Linear Discriminant Analysis (LDA), MLP-ANN and LDA were chosen because some research discovered their efficiency dealing with small dataset.

To the best of our knowledge, hybrid deep learning in predicting student performance has never been reported. Herein, the primary objective of this paper is to propose a hybrid CNN-RNN to predict the student performance and to identify the main factor has the highest correlation with the student performance, Table 1 shows the advantage features of recurrent neural network (RNN), Convolutional Neural Network (CNN).

According to the literature studies, CNN and RNN, as well as hybrid CNN-RNN, are often used in image and video classification; hence, this research will aid in understanding how CNN-RNN performs in structured data. The incorporation of data pre-processing into the architecture and hyper parameters optimization for student data will be covered in the next section. The remainder of this paper is organised as follows; Section 2 presents the materials and methodology along with the proposed method used in this study. Section 3 discusses the results. Finally, Section 4 concludes with conclusions and further work.

2. Materials and methods

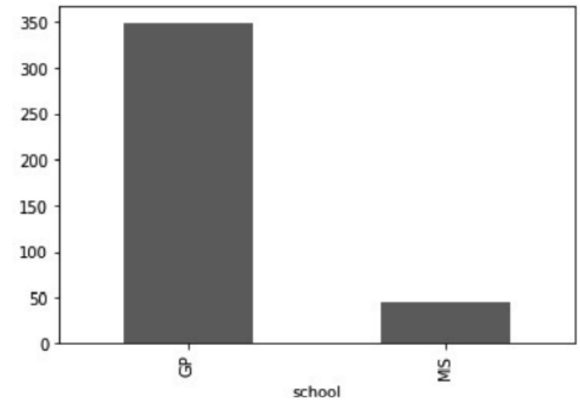
This section outlines the data science project lifecycle along with the proposed model to achieve the research objectives effectively. In addition, it will describe the techniques and tools used to facilitate this research.

2.1. Data Collection

The original datasets were collected by [8] through questionnaires and school database. However, the datasets used in this research were obtained from a secondary source called Kaggle repository.

2.2. Data Preprocessing

Several features are being plotted as graph to analyze the data distribution to find out if there is any imbalance issue and to understand the data distribution. The class imbalance issue has been noticed in two features as shown in Fig.1 and Fig.2 This had been taken care when splitting data into training and validation data, both test and validation data will have data with "MS" school and age range from 19-22. Furthermore, the data being used in this research is structured data, each feature consider a channel for input for CNN. In this case, a reshape was done to match the input for CNN API, which is converting all the feature in spatial dimension and become single dimension data. According to [28] Machine learning algorithms perform effectively when the features in the data are of the same scale. Moreover, it's a preferred pre-processing approach for ANN as it yields better results [29]. This is done to reduce the impact of one element on another. Aside from that, normalization substantially accelerates the training process [30], [31]. As a result, all features in the datasets were normalized.



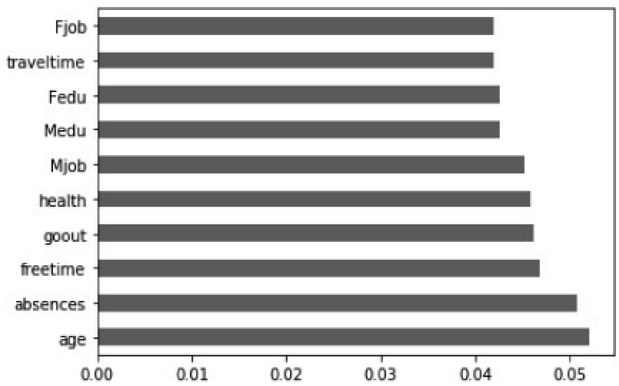


Fig. 2. The age variable distribution

2.3. Feature Selection

It is critical to determine the features that are associated for each task. When the data is too vast to analysis it must be reduced so that worthless features may be eliminated, resulting in a more robust model [32].

Initially, SelectKBest is used to select the top 10 features to be used for learning. It is a function in sklearn library which scores each feature according to k highest score. Features with a higher score are more strongly correlated to the output. By removing unnecessary characteristics using the score function chi2, the model's accuracy improved, and overfitting was reduced. Table 2 illustrates the scores for each of the feature in the data and the features.

Table 2. Score for each feature by SelectKBest

Feature	Score	Feature	Score
absences	**294.3628**	higher	11.3438
failures	**255.7793**	activities	11.1314
school	**60.0358**	address	11.1066
studytime	**40.6074**	romantic	10.7417
traveltime	**35.6419**	goout	10.1056
Dalc	**35.4910**	freetime	9.0309
Mjob	**34.8889**	famsize	8.1309
reason	**33.6175**	famsup	7.5338
Medu	**33.5680**	Fjob	7.2778
Fedu	**32.3688**	guardian	6.3069
Walc	25.7720	internet	5.6943
schoolsup	21.7564	age	5.2234
health	14.5187	nursery	3.1871
paid	13.2521	famrel	3.0397
sex	12.9301	Pstatus	0.4874

***(Bold features) Features selected to train the classification model.**

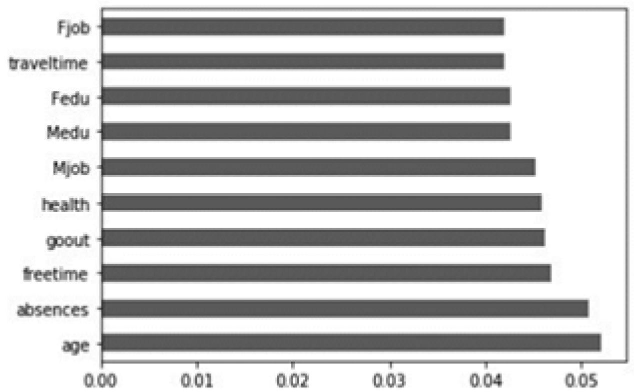


Fig.3. Top 10 features selected using ExtraTreeClassifier

The following technique is ExtraTreeClassifier, which is an ensemble method in the sklearn package that use a meta estimator to fit many randomized decision trees. The higher the score, the more important the characteristic is to the output variable. Fig.3 shows the top 10 features chosen by ExtraTreeClassifier, as well as the features score.

Finally, using Pearson correlation coefficient, which is a statistic approach that evaluates linear connection between two variables, an examination of how the characteristics relate to each other or the prediction output can be conducted. The correlation might be positive or negative. A heatmap is a type of visualization tool that allows us to identify the relationship between elements. In Fig.4 the correlation coefficient is display using colour and annotations, annotations with negative value mean the features is negative correlated and vice versa. With the help of heatmap, the correlation between features can be visualized. The identified factors that affecting student performance will be discussed in Section 3.

2.4. Data reshape

The data being used in this research is structured data, each feature consider a channel for input for CNN. In this case, the data were reshaped to match the input for CNN API, which is converting all the feature in spatial dimension and become single dimension data without changing it data.

2.5. Proposed Hybrid CNN-RNN architecture

The main objective of this paper is to propose a Hybrid CNN-RNN architecture to predict Student performance. The data will be modelled using Supervised classification approaches. The output (G3) will be encoded as Pass or Fail; the predicted output will be pass if the G3 (Final grade) is more than 10, otherwise fail. There are difficulties in determining how many hidden layers, node size, RNN type, and CNN parameter to use. The use of several layers is justified by the assumption that this issue is not linearly separable.

The planned architecture is divided into seven levels. The first layer consists of convolution layers, which are followed by three fully connected networks with ReLu activation functions, followed by an LSTM layer, and lastly by the classification output.

Fig.5 depicts the overall process for this study, from data preparation through model validation. Then, Fig.6 shows the architecture used in this study. Fig.7 shows the Psedo code for Hybrid CNN-RNN used in student performance prediction and Table 3 shows the specifics for each layer for CNN, RNN, and ANN. This architecture is an enhanced version of the [33] with modification to accept tabular data instead of image data, extra data reshape layer is added and hyper parameters is tuned to predict student performance. Due to the fact that the architecture used in [33] focuses on video and image data, therefore it cannot be used to predict student performance since the structure of student data is not spatial and sequential, as opposed to video and image.

Lastly, sparse categorical cross entropy and Adam optimizer were used as loss function and optimizer in this research, respectively. The optimizer being used due to it appropriate for non-stationary objectives and problems with very noisy and/or sparse gradients [34].

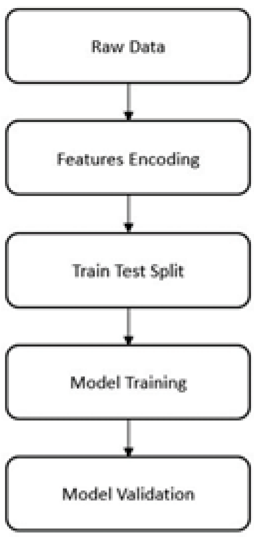


Fig.5. Overall process of the proposed hybrid model

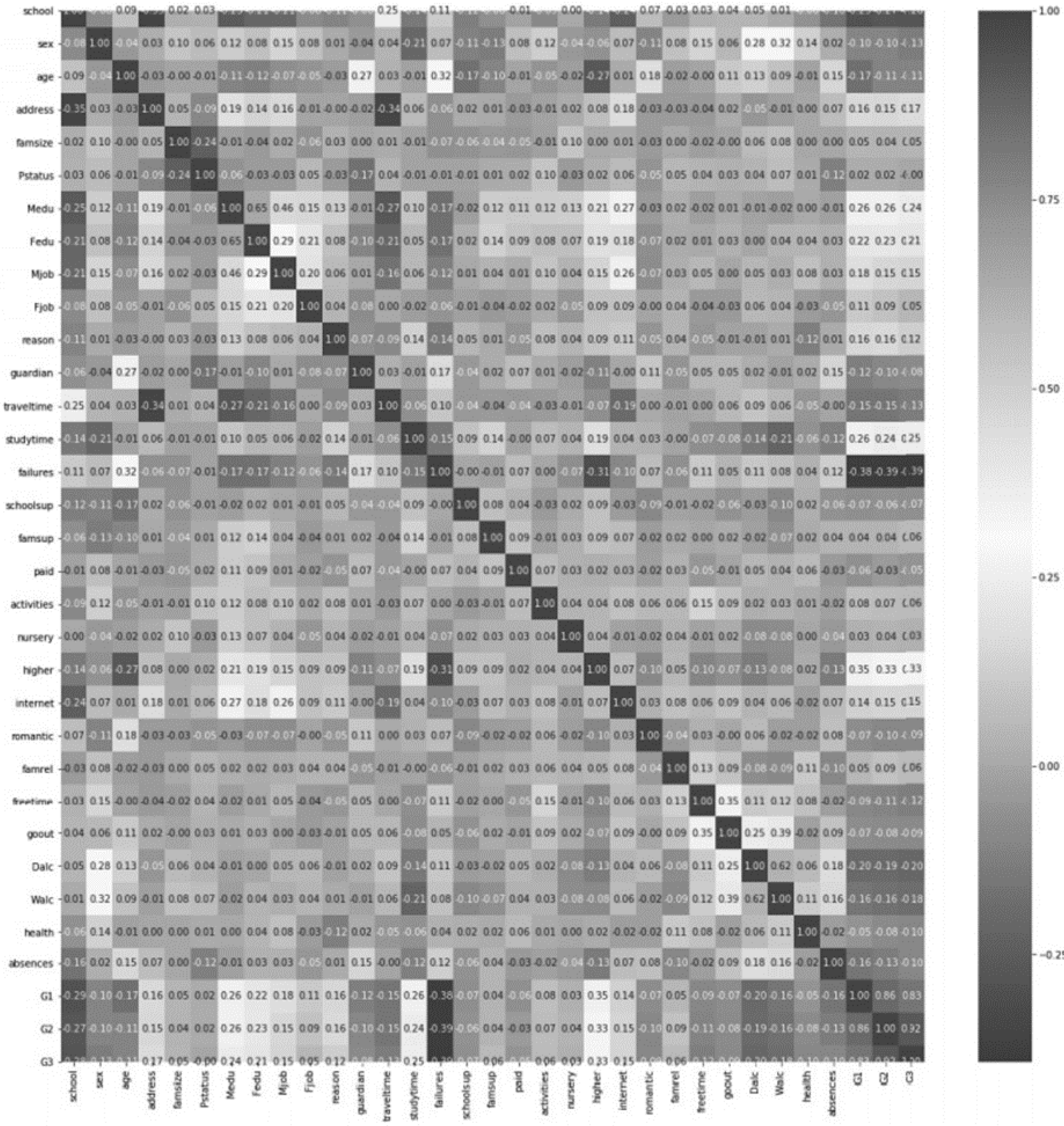


Fig.4. Heatmap for student correlation coefficient

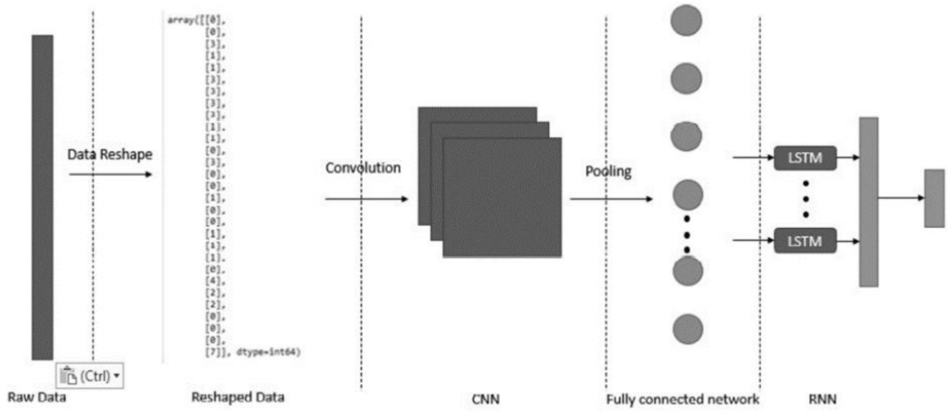


Fig.6. Hybrid CNN-RNN Architecture

```
Algorithm: Pseudo code for Hybrid CNN-RNN used in student performance
prediction
Input   A set of training example
Output  Trained model
        Initialize weight vector for each node
        Initalize 10 random kernel to convolution layer
        Reshape input
        for i = 1 to (predefined iteration) do
           for j = 1 to (number of training dataset) do
              x1 = Convolution(training_data[j])
              x2 = Pooling(x1)
              x3 = ANN(x2)
              x4 = LSTM(x4)
              y_predict = Softmax(x4)
              y_loss = cost_function(y_desired, y_predict)
              backprogate(y_loss)
           end for
        end for
```

Fig.7. Psedo code for Hybrid CNN-RNN used in student performance prediction

Table. 3. Hybrid CNN-RNN Architecture

Layers	Name	Configurations	Modules
1	Conv1	10 kernels, kernel size 3	CNN
2	Pooling1	64 inputs	
3	FC1	30 outputs, ReLu	
4	FC2	30 outputs, ReLu	ANN
5	FC3	30 outputs, ReLu	
6	LSTM1	LSTM, 10 hidden unit output	RNN
7	Softmax		Classification

Epoch: need to find the optimum epoch to prevent underfitting or overfitting of the model. Monitor and compare the loss of the training data and validation data to find the point that both datasets having the lowest lost. At certain point, the loss of the validation data increasing instead of decreasing, this show that the model is already overfit. By analyzing at the chart in fig.8, the loss of the validation increasing while the loss of the training data decreasing. This is the sign of overfitting, which need to be mitigated by using several techniques.

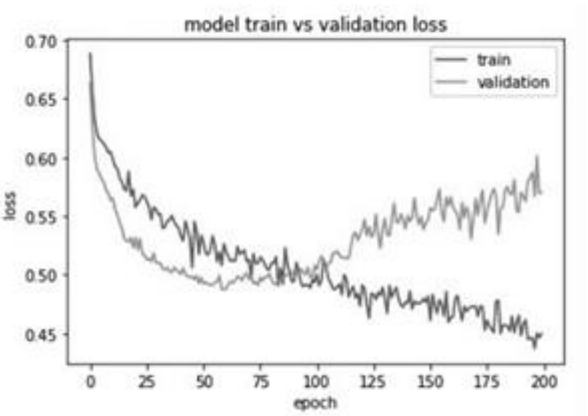


Fig.8. Loss vs Epoch Chart

It's clear that, when less 100 epochs, the validation loss starts increase while training data decreasing. To mitigate this overfitting issue, early stopping is implemented, notice that at the 100 epochs, accuracy already at 80%.

2.6. Prediction Performance Evaluation

For classification, accuracy, recall, and precision, confusion matrix is the commonly used measurements to evaluate how the trained model behave.

Table. 4. Confusion Matrix

True positive (TP) • Reality: Pass • Prediction: Pass	False Positive (FP) • Reality: Fail • Prediction: Pass
False Negative (FN) • Reality: Pass • Prediction: Failed	True Negative (TN) • Reality: Failed • Prediction: Failed

The formula for accuracy: $((TP+TN))/((TP+TN+FP+FN))$

The formula for recall: $TP/((TP+FN))$

The formula for precision $TP/((TP+FP))$

Accuracy is the most obvious performance measure, it just simply a ratio of correctly predicted to the total number of predicted, it is a great measure when you are having a balance dataset. The recall is the ratio of TP (correctly predicted as pass). This mean in the validation data, the students that are pass, how many were labeled correctly. Lastly, precision is the ratio of correctly positive prediction to the total predicted positive, this answer the question of all students labelled as Pass, how many Pass?

2.7. Experimental environment

Jupyter notebook, a platform for supporting interactive data science and scientific computing, was utilized in the research. Furthermore, Python was used as the programming language, and the primary library for training models is TensorFlow, pandas is used for data pre-processing, and matplotlib is used for graph displaying. To enhance the training time, CUDA core was applied in which it speed up the training by 3 times, therefore provide more opportunity to tune the hyper parameter of the Neural Network. The computer system specification is shown in table 5.

Table. 5. System Specification to Run Experiment

CPU	Intel i5-9400F
RAM	32GB DDR-2666
Hard Disk	1TB SSD (Read: 3230 MB/s, Write: 1625 MB/s)
GPU	GeForce GTX 1060 6GB

3. Results and Discussion

The outcomes of the following machine learning and deep learning algorithms will be discussed in this section (KNN - SVM - ANN - CNN - Hybrid CNN-RNN). Furthermore, the sample data is divided into 80 percent training data and 20 percent validation data using the numpy module. The experiment is then repeated in two sets, with all features selected and features selected using SelectKBest (10 features) and Feature Importance (10 features). The selected 10 features are in the following Table 6.

Tab. 6. Feature selected

SelectKBest		ExtraTreeClassifier		Correlation	
Feature	Score	Feature	Score	Feature	Correlation Coefficient
absences	294.36	age	0.05	failures	0.39
failures	255.78	absences	0.05	higher	0.33
school	60.04	famrel	0.05	school	0.28
studytime	40.61	goout	0.05	studytime	0.25
traveltime	35.64	health	0.05	Medu	0.24
Dalc	35.49	Walc	0.05	Fedu	0.21
Mjob	34.89	freetime	0.05	Dalc	-0.20
reason	33.62	Medu	0.04	Walc	-0.18
Medu	33.57	Fedu	0.04	address	0.17
Fedu	32.37	Mjob	0.04	internet	0.15

Result using all features are shown in Table 7 and the comparison of the prediction performance for all features is shown in Fig.9.

Table 7. Accuracy, Precision, Recall and Epoch with all features

	Accuracy	Precision	Recall	Epoch
RNN	73.07	79.16	90.48	50
CNN	76.15	83.19	90.91	10
CNN-RNN	79.23	83.07	94.49	55
ANN	76.15	77.95	97.05	10
KNN	75.69	79.52	97.11	
SVM	74.15	82.50	90.80	

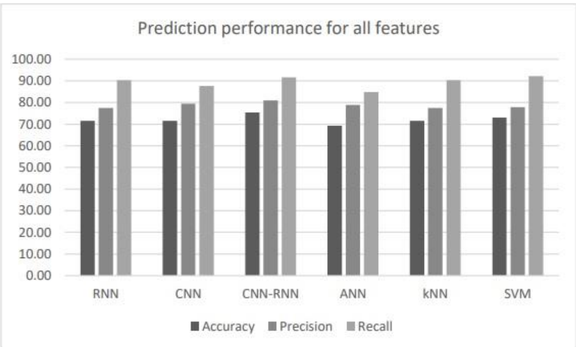


Fig. 9. Prediction Performance for All Features

Result using 10 features as in Table 8, Table 9 and Table 10, which the features is selected by SelectKBest, ExtraTreeClassifier and Feature Correlation respectively. Furthermore, the compression of the prediction performance for top 10 selected features is shown in Fig. 10.

Table. 8. Model Performance and Epoch with Top 10 Features Selected (SelectKBest)

	Accuracy	Precision	Recall	Epoch
RNN	78.69	84.68	94.59	55
CNN	77.69	84.87	90.18	5
CNN-RNN	81.54	85.48	94.64	50
ANN	74.15	85.34	87.61	10
kNN	72.30	78.99	89.52	
SVM	74.46	82.25	94.44	

Table. 9. Model Performance and Epoch with Top 10 Features Selected (ExtraTreeClassifier)

	Accuracy	Precision	Recall	Epoch
RNN	71.54	77.50	90.29	150.00
CNN	71.53	79.48	87.73	50.00
CNN-RNN	75.38	80.99	91.58	100.00
ANN	69.23	78.94	84.91	50.00
kNN	71.53	77.50	90.29	
SVM	73.07	77.86	92.23	

Table. 10. Model performance and Epoch with top 10 features selected (Feature Correlation)

	Accuracy	Precision	Recall	Epoch
RNN	78.46	83.61	92.72	100.00
CNN	80.76	86.77	92.11	50.00
CNN-RNN	82.31	86.29	94.69	50.00
ANN	81.53	86.17	93.81	20.00
kNN	80.00	83.87	94.54	
SVM	78.46	83.61	92.72	

For ANN, CNN, RNN and Hybrid CNN-RNN the validation accuracy fluctuates in every epoch, the stopping epoch is decided by analyzing the chart epoch vs loss and epoch vs accuracy. Following method is being used for all neural network and the stopping epoch is display in Table 9.

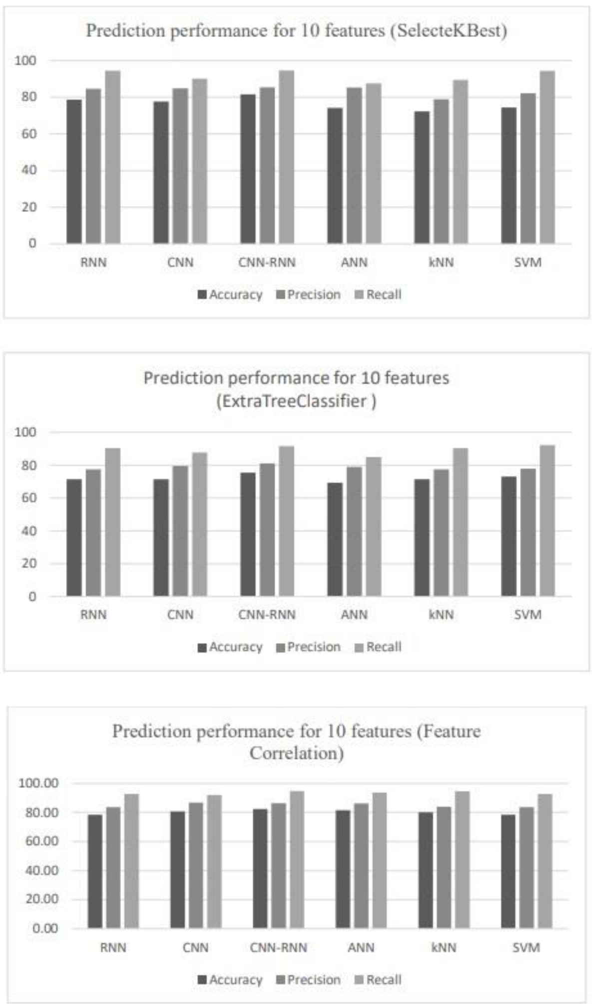


Fig.10. Prediction Performance for Top 10 Selected Features

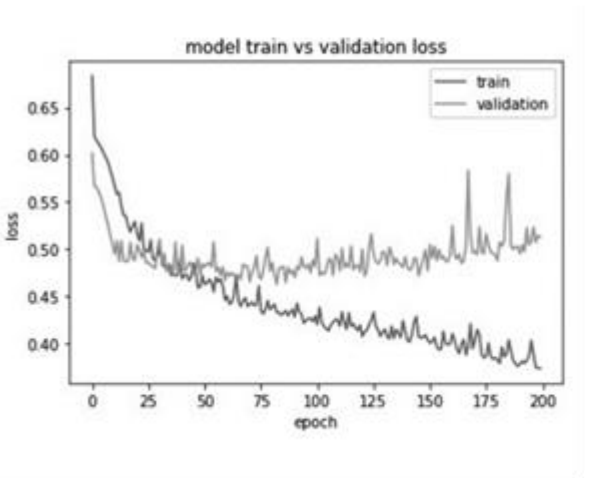


Fig.11. Hybrid CNN-RNN / Epoch vs loss for 1000 epoch

By analyzing Fig.11, the loss starts to increase after epoch 50, this indicates the model starts overfitting. Moreover, in Fig.12, the accuracy of validation dataset starts to decrease at epoch 50 while accuracy for training dataset still increasing, this also another indicator of model overfit. Thus, we can decide to stop training at 50 epochs. Table 11 is the identified factors that affecting student performance, this feature is selected using feature correlation method. The reason for selecting the features chosen by feature correlation is that by utilizing the features to train the model, it achieved the best accuracy, and the coefficient sign can provide insight into whether variables are negatively or positively affect the student performance.

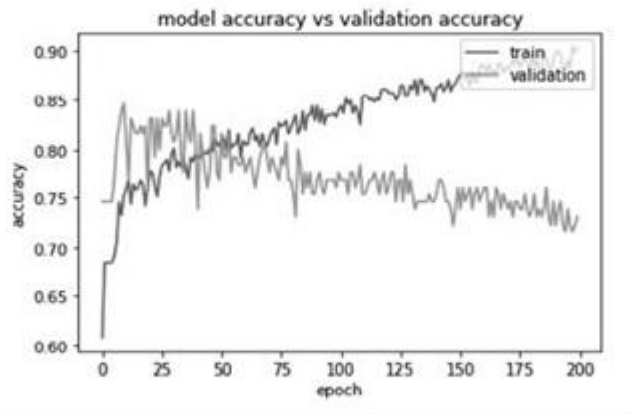


Fig.12. Hybrid CNN-RNN / Epoch vs Accuracy for 1000 epoch.

Tab. 11. Factors that affecting student performance

Feature	Correlation Coefficient	Description
failures	0.39	number of past class failures
higher	0.33	wants to take higher education
school	0.28	Student's school (binary: Gabriel Pereira or Mousinho da Silveira)
studytime	0.25	Weekly study time
Medu	0.24	Mother's education
Fedu	0.21	Father's job
Dalc	-0.20	Workday alcohol consumption
Walc	-0.18	Weekend alcohol consumption
address	0.17	Student's home address type
internet	0.15	Internet access at home

Prediction over features selected by SelectKBest and feature correlation yield a better accuracy compared to using all the features. Aside from improved accuracy, the training needed fewer iterations to get the greatest accuracy. To further analyze the model's performance, the accuracy and recall are also calculated; having either a high false positive or a high false negative may result in a failure to provide suitable education to the learner.

The same dataset is being use for several classification method, RNN performs the worst when using all the features for prediction, ANN and CNN having the same performance in the prediction accuracy and epoch to achieve the best accuracy is the same (Table 7). The proposed algorithm Hybrid CNN-RNN outperform all other in accuracy (Table 7). The trade-off of Hybrid CNN-RNN require more epoch to achieve optimum accuracy compared to the other three deep learning methods. To prove that hybrid CNN-RNN is better than others classification model, the result is being compared to another research which are using the same datasets "Using Data Mining to predict Secondary School Student Performance" [8]. To compare it in the same way, another set experiment is set up and able to get 96.15% while the previous research achieves 85% using RF.

On the other hand, the second objective of this research was to identify the main factors that have the highest correlation with the student performance using feature selection. Three methods are being use, SelectKBest, ExtraTreeClassifier and correlation matrix with heatmap. In the heatmap, G1 and grade G2 is the top 2 features that have positive correlation to G3. G1 and G2 is the result for 1st and 2nd semester respectively, Thus, a conclusion has been reached that a student's performance is impacted by prior results, but these two variables are not being used for the prediction model because they are not available before the student join the school. The research on student performance prediction after course commencement appears to be extensive. However, early prediction prior to course commencement remains a difficulty [35].

Alternatively, there are few available elements before the students started their education at school, an overlook at the father's and mother's education levels might been conducted which have a positive association with output. Additionally, travel time, alcohol consumption either on weekend or weekday having negative correlation to the performance, reducing time spend on these 3 activities can help to improve the student performance. Comparing the feature selected by SelectKBest and ExtraTreeClassifier, the top 10 features selected by Feature Correlation yield a better result, SelectKBest (81.54%), ExtraTreeClassifier (75.38%) and Feature Correlation (82.31%). By using the features selected by the other two feature selection technique, SelectKBest and ExtraTreeClassifier having lower accuracy and ANN training requires more epoch to converge. In conclusion, features selected by Feature Correlation is a best method in this research, it gives better accuracy and faster converge in neural network model training, the prediction accuracy also improved by 4.5% average compared to using prediction on all features. By using the top 10 features that are highly correlated to the output to train the model, hybrid CNN-RNN still having the highest classification accuracy. However, the improvement is less significant for the hybrid CNN-RNN model compared to other machine learning and deep learning method, this demonstrate the feature extraction ability in the hybrid model, thus feature selection steps can be skipped by using this model. Table 12 illustrates the accuracy comparison for all features and selected features.

Table 12: Accuracy comparison for all features and selected features

	Accuracy (All features)	Accuracy (Top 10 correlated features)	Difference
RNN	73.07	78.46	5.39
CNN	76.15	80.76	4.61
CNN-RNN	79.23	82.31	3.08
ANN	76.15	81.53	5.38
KNN	75.69	80.00	4.31
SVM	74.15	78.46	4.31

In conclusion, hybrid CNN-RNN perform the best in our experiment, due to CNN filters assist in extracting the right and relevant feature from the input data and RNN captures the sequential information present in the input data, with the capability of both CNN-RNN the model will have a build in feature selection method thus it able to archive a higher accuracy.

4. Conclusion

Young generation is the future of a nation, having a good education system raise the overall skill level of the country. With a strong early prediction system, student performance may be correctly predicted, and an early plan or action can be taken to assist the student. Although there are several studies utilized machine learning in student performance prediction, but there are room of improvement for the prediction accuracy. For the current visualization method, chart plotting will be straightforward for univariate features, yet it will be challenging to analyze and discover the underlying data pattern for multivariate charts. Machine learning is one of the methods that is easier to tune and use. Nevertheless, it may suffer from the curse of dimensionality; as more and more features become available in the future, the sparsity of the data increases, degrading the performance of tasks which are dependent on the input space [36]. The proposed CNN-RNN model is targeted to improve the accuracy of the prediction model and reduce the problem of curse of dimensionality. With the convolution layer, the number of input parameter for fully connected network can be reduced and significantly reduce the complexity of the model. The hybrid model performs better than traditional machine learning classifiers in the experiment. The proposed model also performed well when using all the features compare to ANN, this demonstrates it ability of extracting meaningful features from the data set automatically.

Future studies on the current topic are recommended such as more experiments in hybrid model for instance meta-heuristic hybrid with deep learning, using meta-heuristics in tuning the deep learning model. Additionally, research different methods in selecting the features and perform experiment using the selected features. This also can identify the main factor that affecting the student performance. Furthermore, obtaining more features by using automated tracking device.

Acknowledgements

This work is supported by the Universiti Sains Malaysia, Short Term Grant [Grant Number: 304 / PKOMP / 6315616], with the project entitled “New Coefficient of Variation Control Charts based on Variable Charting Statistics in Industry 4.0 for the Quality Smart Manufacturing and Services”.

References

[1] J. Sultana, M. Usha, and M. A. H. Farquad, “Student’s performance prediction using deep learning and data mining methods,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 1018–1021, 2019.

[2] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, “Predicting student performance in higher educational institutions using video learning analytics and data mining techniques,” Appl. Sci., vol. 10, no. 11, 2020.

[3] Kadhim, R. R., and M. Y. Kamil. “Evaluation of Machine Learning Models for Breast Cancer Diagnosis Via Histogram of Oriented Gradients Method and Histopathology Images”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 36-42, doi:10.17762/ijritcc.v10i4.5532.

[4] R. Hasan, S. Palaniappan, S. Mahmood, K. U. Sarker, and A. Abbas, “Modelling and predicting student’s academic performance using classification data mining techniques,” Int. J. Bus. Inf. Syst., vol. 34, no. 3, pp. 403–422, 2020.

[5] R. Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarker, “Student Academic Performance Prediction by using Decision Tree Algorithm,” in 2018 4th International Conference on Computer and Information Sciences: Revolutionising Digital Landscape for Sustainable Smart Society, 2018.

[6] B. Oancea, R. Dragoescu, and S. Ciucu, “Predicting students’ results in higher education using a neural network Predicting students’ results in higher education using a neural network,” Appl. Inf. Commun. Technol., no. 72041, 2013.

[7] L. M. Abu Zohair, “Prediction of Student’s performance by modelling small dataset size,” Int. J. Educ. Technol. High. Educ., vol. 16, no. 1, pp. 1–18, Aug. 2019.

[8] Y. Bengio and Y. Lecun, “Scaling Learning Algorithms Towards

AI," Large-Scale Kernel Mach., vol. 1, pp. 1–41, 2017.
[9] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," Comput. Sci., 2008.
[10] S. Natek and M. Zwilling, "Student data mining solution-knowledge management system related to higher education institutions," Expert Syst. Appl., vol. 41, no. 14, pp. 6400–6407, Oct. 2014.
[11] A. A. Saa, M. Al-Emran, and K. Shaalan, "Mining Student Information System Records to Predict Students' Academic Performance," in Advances in Intelligent Systems and Computing, 2020, vol. 921, pp. 229–239.
[12] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," Comput. Educ., vol. 143, p. 103676, Jan. 2020.
[13] Chaudhary, D. S. . (2022). Analysis of Concept of Big Data Process, Strategies, Adoption and Implementation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 05–08. https://doi.org/10.17762/ijfrcsce.v8i1.2065
[14] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," Cybern. Inf. Technol., vol. 13, no. 1, pp. 61–72, 2013.
[15] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: An application of data mining methods with an educational web-based system," in 33rd Annual Frontiers in Education, 2003, vol. 1, p. T2A13-T2A18.
[16] E. Osmanbegovic and M. Suljic, "Data Mining Approach for Predicting Student Performance," J. Econ. Bus., vol. 10, no. 1, pp. 3–12, 2012.
[17] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," Appl. Sci., vol. 11, no. 1, pp. 1–28, Jan. 2021.
[18] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," EUROSIS-ETI, 2008.
[19] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," in Procedia Computer Science, 2015, vol. 72, pp. 414–422.
[20] A. Daud, M. D. Lytras, N. R. Aljohani, F. Abbas, R. A. Abbasi, and J. S. Alowibdi, "Predicting student performance using advanced learning analytics," in 26th International World Wide Web Conference 2017, WWW 2017 Companion, 2017, pp. 415–421.
[21] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," Procedia Technol., vol. 25, pp. 326–332, Jan. 2016.
[22] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacob, and N. M. Sobri, "Supervised data mining approach for predicting student performance," Indones. J. Electr. Eng. Comput. Sci., vol. 16, no. 3, pp. 1584–1592, 2019.
[23] R. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," J. Educ. Data Min., vol. 1, no. 1, pp. 3–17, Oct. 2009.
[24] A. Kumar, R. Pandi Selvam, and K. Sathesh Kumar, "Review on prediction algorithms in educational data mining," Int. J. Pure Appl. Math., vol. 118, no. Special Issue 8, pp. 531–537, 2018.
[25] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Syst. Appl., vol. 33, no. 1, pp. 135–146, Jul. 2007.
[26] Q. Liu et al., "Fuzzy cognitive diagnosis for modelling examinee performance," ACM Trans. Intell. Syst. Technol., vol. 9, no. 4, Jan. 2018.
[27] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," IEEE Access, vol. 8, pp. 55462–55470, 2020.
[28] Y. Li, J. Gou, and Z. Fan, "Educational data mining for students' performance based on fuzzy C-means clustering," J. Eng., vol. 11, pp. 8245–8250, 2019.
[29] J. T. Heaton, Programming Neural Networks with Encog3 in Java, 1st ed. Heaton Research, 2011.
[30] A. Ethem, Introduction to Machine Learning. The MIT Press, 2014.
[31] N. K. Rotich, J. Backman, L. Linnanen, and P. Daniil, "Wind resource assessment and forecast planning with neural networks," J. Sustain. Dev. Energy, Water Environ. Syst., vol. 2, no. 2, pp. 174–190, Jun. 2014.
[32] Kiran, M. S., & Yunusova, P. (2022). Tree-Seed Programming for Modelling of Turkey Electricity Energy Demand. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 142–152. https://doi.org/10.18201/ijisae.2022.278
[33] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," IEEE Trans. Nucl. Sci., vol. 44, no. 3 PART 3, pp. 1464–1468, 1997.
[34] O. Abdelrahman and P. Keikhosrokiani, "Assembly Line Anomaly Detection and Root Cause Analysis Using Machine Learning," IEEE Access, vol. 8, pp. 189661–189672, 2020.
[35] A. Sharma, A. Ram, and A. Bansal, "Feature Extraction Mining for Student Performance Analysis," in Proceedings of ICETIT 2019, 2020, vol. 605, pp. 785–797.
[36] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," IEEE Access, vol. 8, pp. 24713–24720, 2020.
[37] D. P. Kingma and B. Jimmy, "Adam: A Method for Stochastic Optimization," in 3rd International Conference for Learning Representations, 2015.
[38] N. A. Libre. (2021). A Discussion Platform for Enhancing Students Interaction in the Online Education. Journal of Online Engineering Education, 12(2), 07–12. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/49
[39] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," Educ. Inf. Technol., vol. 26, no. 1, pp. 205–240, 2021.
[40] N. Venkat, "The Curse of Dimensionality: Inside Out," ServerProcessor, 2018. [Online].