Improving Student Grade Prediction Using Hybrid Stacking Machine Learning Model

Saloni Reddy[1] and Sagar Dhanraj Pande[2, *]

[1,2, *]School of Computer Science & Engineering, VIT-AP University, Andhra Pradesh, India.

Abstract

With increasing technical procedures, academic institutions are adapting to a data-driven decision-making approach of which grade prediction is an integral part. The purpose of this study is to propose a hybrid model based on a stacking approach and compare its accuracy with those of the individual base models. The model hybridizes K-nearest neighbours, Random forests, XGBoost and multi-layer perceptron networks to improve the accuracy of grade prediction by enabling a combination of strengths of different algorithms for the creation of a more robust and accurate model. The proposed model achieved an average overall accuracy of around 90.9% for 10 epochs, which is significantly higher than that achieved by any of the individual algorithms of the stack. The results demonstrate the improvement of prediction results but using a stacking approach. This study has significant implications for academic institutions which can help them make informed grade predictions for the improvement of student outcomes.

*Corresponding author. Email: sagarpande30@gmail.com

1. Introduction

The world is built on the promise of constant personal and societal development which is only possible through good education by the virtue of knowledge and skills it imparts. Moreover, data has lately become a driving force in all sectors of the world. Education is one such sector where data plays an important role in determining all sorts of variables [1]. But in the past few years, there has been a rise in dropout rates [2] and the first step in preventing this is predicting the possibility of dropouts followed by a need to recognize the potential of students to provide for a better system and ensure they adequately acquire the required knowledge and skills necessary for the success and development of mankind. This might be one of the main reasons why academic institutions are driven towards adopting data-driven decision-making approaches by using various technical procedures. One application of this approach is a prediction of the future grades of various students based on their previous marks [3]. This calls for the development of new models that can accurately predict the grades. But in cases where the testers

might want to use pre-existing models, due to such a vast number of algorithms and techniques available, it can be difficult to determine the most effective approach to use. This is where this study comes in to help by comparing not only various algorithms but also the various approaches to prediction.

The main purpose of this study is to aid any academic institution seeking to make informed decisions about student outcomes. There have been various studies listed in the forthcoming sections that imply the need for not only providing timely support to struggling students but also for providing further support and recognition to high-performing students [4]. All these can be done quite effectively by analyzing patterns in the obtained marks and using them for grade predictions. Additionally, it can help the institutions adopt a more targeted approach rather than a general one which in turn can lead to more efficient use of resources. This offers practical applications in the field of education by rigorous comparison between the different models deployed, whilst providing a high overall accuracy score for the stacked model.

This study compares the results of a stacking model with the base algorithms, namely K-nearest neighbors [5], Random forests [6], XGBoost [7] and multi-layer perceptron networks [8] to study the effectiveness of each of them and the stacking approach used to predict and analyze grade and methods that can be implemented for the improvement of the same. The process starts with the main steps of the knowledge discovery process which include data collection, data pre-processing, and data mining process among others. By doing so, we aim to create a hybrid model which aims to enable a combination of strengths of different algorithms [9] for the creation of a more robust and accurate model while reducing the weakness of each individual approach. The results of the study are evaluated using three robust metrics discussed in the forthcoming sections and show that the proposed hybrid model achieved an average overall accuracy of around 90.9%, which is significantly higher than that achieved by any of the individual algorithms of the stack, as elaborated in Section 4.

The research was conducted in the following procedure:

- Data collection and pre-processing for the effective performance of models
- Application of base machine learning models namely K-nearest neighbor (KNN), random forest (RF), Multi-layer perceptron networks (MLP) and XG Boosting (XGB) on the processed dataset,
- Performance evaluation based on metrics like Accuracy, Matthew's Correlation Coefficient (MCC) and F1 score.
- Construction of a hybrid stack model after experimentation with stacking order for best accuracies throughout several epochs
- Comparison of the performance of base classifiers with the proposed hybrid model.

As for the remainder of the paper, Section 2 provides a literary overview of relevant research related to this study, Section 3 goes over the proposed methodology, Section 4 elaborates on the experimental results and discussion, and to finish, the conclusion and future work is discussed in Section 5.

2. Literature Review

The use of machine learning techniques for educational data mining has been gaining a considerable amount of attention in the past few years. Predicting student grades is one of the major applications of the same. However, the major concentration of the research has been on assessing the implications of the grades obtained to facilitate degree planning or to determine dropout risk [10].

A 2021 study by Namoun [11] analyzed 47 studies published between 2010 and 2020 and concluded that decision trees, followed by logistic regression, neural networks, and support vector machines were the most used for academic predictions. These are however omitted since a lot of work has already been done using the above models in the field of academic-related research. They also identified several challenges and limitations in the existing literature,

such as the lack of standardization in data collection and analysis, the limited use of advanced machine learning techniques, and the lack of generalizability of the predictive models. This has been a starting point in building and analyzing our dataset.

In a paper by (Jayaprakash et al., 2020), an improved random forest classifier has been engaged to get prior consultation about the performance of students and curate a plan for success in the future [12]. Another more comprehensive study by Y.K. Salal, M. Hussain, and T. Paraskevi focuses on using machine learning techniques including Support Vector Machines (SVM), Random Forest (RF), and Naive Bayes (NB) to predict the next assignment submission of a student and managed to achieve an accuracy of 85.5% [13]. Although the study was conducted dataset of relatively small size, the results implicate its effective extension to larger datasets as well.

The main motivation behind this paper, however, was a 2022 study by Kanetaki et al. [14] which illustrated the development of a hybrid machine learning model for grade prediction in online engineering education which used a combination of decision tree, random forest, and gradient boosting algorithms. The performance of the hybrid model came to be vastly higher than the individual algorithms in terms of accuracy and precision which was an important factor for the flourishing of the academic industry. Ghosh et al.'s 2023 study on machine learning for [15] water quality analysis, 'Water Quality Assessment Through Predictive Machine Learning', explores predictive analytics for water parameters. In 2023, Rahat and Ghosh's 'Unraveling the Heterogeneity [16] of Lower-Grade Gliomas' discusses the use of deep learning in brain MR image analysis for medical insights. The 2023 work [17] by Ghosh, Rahat, and their team, 'Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks', demonstrates the use of neural networks in detecting potato leaf diseases. Mandava, Vinta, Ghosh, and Rahat's 2023 research, 'An All-Inclusive Machine Learning and Deep [18] Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population', integrates [19] AI for health forecasting. The study 'Identification and Categorization of Yellow Rust Infection in Wheat through Deep Learning Techniques' by Mandava et al. in 2023, applies deep learning to wheat disease detection. Khasim, Rahat, Ghosh, and others' 2023 article, 'Using Deep [20] Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh', explores AI in rice-leaf disease diagnosis. In 2023, Khasim, Ghosh, Rahat [21] and colleagues' 'Deciphering Microorganisms through Intelligent Image Recognition' discusses machine learning for microorganism identification. Mohanty, Ghosh, Rahat, and Reddy's 2023 study, 'Advanced [22] Deep Learning Models for Corn Leaf Disease Classification', focuses on deep learning for classifying corn leaf diseases. Alenezi and team's 2021 research [23]'Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light' investigates CNN methods for underwater image enhancement.

3. Methodology

Fig. 1 represents the flow of work that had been followed to conduct the research. The dataset was pre-processed using cleaning techniques and Label Encoding followed by feature reduction using a threshold for Pearson's correlation and min_max scaling. The processed dataset was then divided into training and validation sets. The base models were built on the training set and subsequently acted as building blocks for the hybrid stack model. This information, along with the application of these models on the validation set, was used for evaluation across three metrics. To improve the performance, hyperparameter tuning for each model was conducted until the optimal performance was reached. The optimised models were then used for comparison across one another and also for predicting grades for new test data.
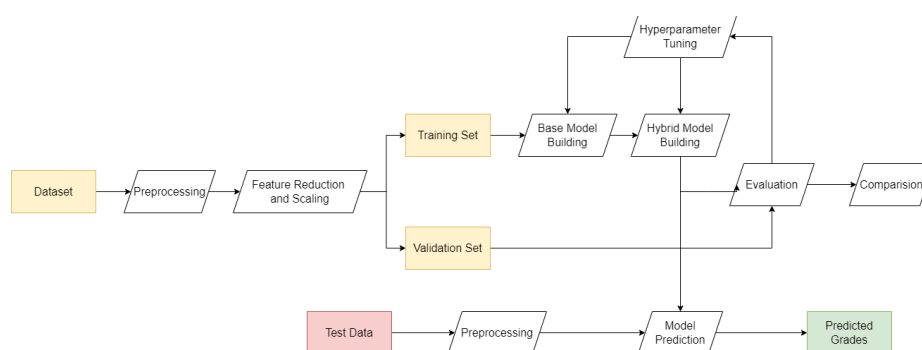

Figure 1. Proposed flow of work

3.1. Dataset

The data used in prediction is a collection of marks and grades of 40 subjects (Table 1) manually collected over the course of 6 semesters by using different test and evaluation techniques, including Continuous Assessment Test-1 (CAT1), Continuous Assessment Test-2 (CAT2), Final Assessment Test (FAT), Lab Component and internal assessment test for 500 students from an engineering background. The dataset consists of 213 columns. For each subject, we are provided with information about the performance of the students, the number of attempts taken to clear the course, the status of completion and the final grades achieved by each student. The dataset consists of real-world data since it has been sourced from our own institution and is hence original and unexplored. While most of the columns are numeric, some columns such as 'Status' and 'Grade contain categorical representations. The dataset is expected to provide valuable insights into the performance of students in different subjects.

Table 1. Different subjects included in Dataset.

DSA(CSE2001)	OT(MAT2003)	OOP(CSE2005)
DMS(MAT1003)	NP(CSE3011)	OS(CSE2008)
FEEE(ECE1002)	MAD(CSE4002)	SE(CSE1005)
ED(MEC1004)	C(ONL1002)	DBMS(CSE2007)
CG(CSE2006)	AFE(MGT1008)	DBMS(CSE2007)
FSM(MGT1020)	STAT(MAT1001)	TBL(MGT1027)
DAA(CSE3004)	NLP(CSE3015)	WT(CSE4004)
LSM(MGT1002)	IML(CSE3008)	FRENCH-I(FRL1001)
ADDA(MAT1002)	OPTO(PHY2004)	LA(MAT2005)
CSDF(CSE2011)	NANO(PHY2005)	EVS(CHY1001)
SNA(CSE4008)	ITC(CSE2013)	NME(MAT2001)
FOM(MGT1003)	COA(ECE2002)	CN(CSE3003)

3.2. Preprocessing

The dataset has a rather huge number of attributes. Hence, to deal efficiently with the dataset, there was an initial split that separated the data into two new datasets- one containing the marks across all the different subjects, and the second having all the target attributes such as Grade, Total marks, Percentage, etc. This was followed by basic text cleaning and transformation techniques. One final preprocessing technique, label encoding with custom labels, was applied on categorical columns, mainly Grade and Status columns, to enable training and testing of machine learning models [15].
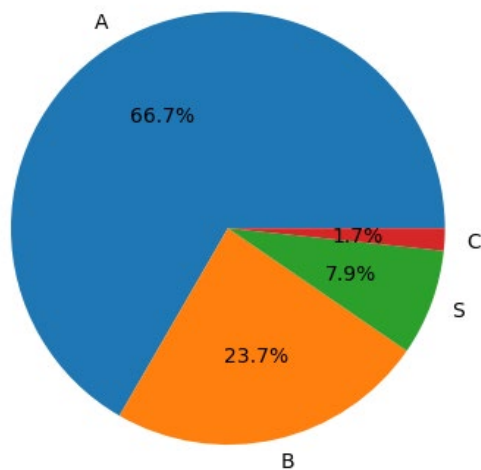
Fig. 2. Distribution of grades in the dataset

As can be seen from Fig. 2, the target variable is highly imbalanced, i.e., more than 50% of the target attributes belong to one single class. Despite this fact, we opted to not perform data balancing since domain knowledge, in this case, suggests that such a skewed distribution is expected. Moreover, instead of removing bias of classification, in this case, balancing may additionally introduce bias in the data which would in turn result in overfitting and as a result, the model would perform poorly on any test data.

Analysis of the data posed two main difficulties. Firstly, the number of observations was rather small in relation to the number of feature/attribute variables, which greatly increased the possibility of overfitting. Second, strong correlations were expected within the different features which would lead to redundancy in computations and an unfair and unnecessary weight advantage in the evaluating models. To combat these difficulties, there was a need to include a feature reduction. We used Pearson's Correlation metric (1) to evaluate the correlation between each of the features which can be seen in Fig. 3 [16]. From any pair having a correlation coefficient of over 90%, the first feature was dropped since its impact on the target attribute remained zero since changes in the said column would be reflected in its pair anyway [17].

$$r = \frac{n(\sum xy) - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} \qquad (1)$$
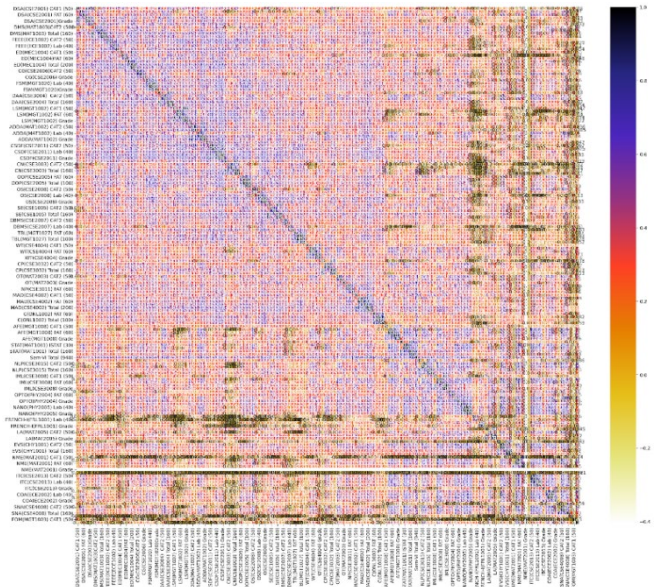

Figure 3. Pearson's correlation matrix

We end the preprocessing stage with a Min-Max scaling procedure to scale the range of data between 0 and 1 described in equation (2) below which allows the models to train for more iterations and assists in increasing the training efficiency while running multiple epochs [18]. This resulted in a much more concise dataset with around 150 features instead of the initial 213 on which we further develop our model.

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

Since some of the base algorithms being used, namely KNN and MLP, are sensitive to the scale of the input features, there arises a need to ensure that all features being given as input to the models are on the same scale, in the absence of which, features with a greater range, especially outliers may induce some bias in the outcome prediction process leading to possibly poor performance. Apart from this, min-max scaling has also been used to quickly reach the convergence of optimization algorithms during training.

3.3. Predictive Models

In order to classify marks of different students into grades, we use four base models, namely K-nearest neighbours (KNN) [5], Random forests (RF) [6], XGBoost(XGB) [7] and Multi-layer perceptron networks(MLP) [8] and end with a stack-based hybrid model [19] built on these base models. All the base models, described below, are implemented using the scikit-learn library. The same library was also used for most of the pre-processing of the data.

KNN calculates the distance of the test tuple with all instances of the training tuple using the Euclidean Distance measure stated in equation (3) and assigns to it the most common label of the K nearest instances. Since the major amount of distribution is covered by 3 grades, the hyperparameter K was set to 3 which was found to give the best results without being sensitive to noise or resulting in underfitting.

$$D(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \qquad (3)$$

RF randomly extracts subsets of data from the given dataset which it uses to train by constructing a decision tree for each of this subset and combining the predictions. The number of subsets to be extracted is determined by the hyperparameter called "n_estimators" which was set to 100 for this dataset, which is considered the most generally optimal.

Like RF, XGB works by learning via predictions of multiple decision trees. But the approach differs vastly since XGB builds an initial decision tree model and proceeds to work on the same tree by using a gradient descent algorithm to correct errors of the initial tree by minimizing the objective function (4), instead of incorporating predictions from decision tree classifiers for different subsets. Apart from this, XBG automatically includes L1 and L2 regularisations and a weighted quantile sketch algorithm to prevent overfitting and to efficiently determine optimal split points respectively [18].

$$obj(\theta) = \sum_{i}^{n} l\left(y_{i,y_j}\right) + \sum_{k=l}^{k} \omega(f_k) \quad (4)$$

MLP receives data from the input layer, process and transforms the data in the hidden layers into prediction-ready features with the help of neurons which apply linear transformation followed by a ReLu activation function (5) which is then fed into an output layer for the final prediction. This process is repeated for 1000 iterations and Adam (Adaptive Moment Estimation) solver is applied to obtain a network with optimal weights.

$$f(x) = \max(0, x) \qquad (5)$$

KNN, despite being one of the simplest algorithms, provides one of the best performance rates of all classifiers. MLP can effectively learn complex relationships between data. RF handles large datasets in high-dimensional spaces while XGB is considered one of the fastest algorithms based on its complexity and computational performance.

The hybrid model works on the principle of combining all the best attributes and strengths of the base algorithms into one. The hybrid model is based on a stacking system that combines the four algorithms listed above. To implement this hybridisation, we use the stacking approach. StackingClassifier from the sklearn library. The model list is fed into the classifier, along with the final estimator, which in this case was taken to be Logistic Regression since it performed the best. The test data was then fed into this regressor to evaluate the overall performance of the hybrid model.

3.4. Performance Evaluation Metrics

Instead of the whole classification report as a summary for performance evaluation, we are going to use three explicitly imported evaluation metrics namely Accuracy(acc) (6), Matthew's Correlation Coefficient (MCC) (7) [20] and F1 score(f1) (8) [21] which calculate scores based on true positives, true negatives, false positives, and false negatives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (6)$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad (7)$$

$$F1 = \frac{2TP}{2TP+FP+FN} \qquad (8)$$

Accuracy is the most used evaluation metric among all which measures the percentage of correctly classified grades. Since the target grade class is balanced, at least for the higher grades, accuracy can give a good estimate of the classification success. Although Matthew's Correlation Coefficient works best on binary classifications, it can be extended to a multiclass model like this one by using a confusion matrix-based extension. MCC measures the balance of true and false positives and true and false negatives. An imbalance in the data occurs since lower grades are considerably less than the higher grades due to which MCC and F1 scores act valuable as evaluation metrics. F1 score, a harmonic mean of precision and recall, ranges from 0 to 1, 1 naturally indicating a perfect classification. F1 score is a good choice considering its importance in evaluating unbalanced data. Using these three evaluation metrics provides a good balance between performance and interpretability. Owing to the virtue of being straightforward, accuracy tends to be easier to interpret than the others, while MCC and F1 scores despite being a bit more complex provide a more balanced view.

4. Results and Discussions

4.1. Experimental Environment

The proposed model was implemented on a laptop using 11th Gen Intel(R) Core (TM) i5-1135G7 @2.40GHz with 8GB RAM and 512 GB SSD memory and Windows 11 on Jupyter Notebook. The same was used for training and evaluation of the said model.

4.2. Experimental Performance

We proposed a stack hybrid model based on soft voting with four base models namely K-nearest neighbours, Random forests, XGBoost, and multi-layer perceptron networks. Table 2 depicts the performance in 1 epoch as

evaluated by test statistics of all the algorithms applied in our study.

With an average accuracy of 90.9% across 10 epochs, the proposed hybrid model performs significantly better than any of the individual algorithms of the stack which demonstrate an improvement in grade prediction. The average individual accuracies of base algorithms came out to be 74.736% for KNN which was the highest among the base models, followed by RF (70.526%), XGB (68.421%), and MLP (63.157%).

Table 2. Evaluation metrics of models

	ACCURACY	MCC	F1
KNN	0.747368	0.391810	0.709692
XGB	0.684211	0.356020	0.680666
RF	0.705263	0.371095	0.696579
MLP	0.631579	0.192648	0.618414
STACK	0.909091	0.815234	0.897322

As can be inferred from the results, KNN predicts the grades of this dataset most accurately among the based models. We see a significant difference in the performance metrics of the hybrid model as compared to the base models. There is a consistent bump of at least 20% across all three metrics. This is because in some form, the hybrid model extracts and combines all the best features of the individual base model to extract the maximum accuracy.

As mentioned above, the Logistic Regression is used as the final estimator since it causes an overall maximisation of the performance evaluation metrics. The optimal estimator may however change for a different dataset. Similarly, for each individual model, optimal hyperparameter values may differ from what were applied during testing. This statement is backed by the underlying result of the experiment that concluded different optimal estimators for each epoch with a distinct training and test set. Hence, for individual epoch(s), there might be a better estimator. Similarly, the overall accuracies might also differ with each dataset owing to the differences in feature distribution and categorisation. However, since the data we have is collected from a real-world source, it is highly likely that the overall distribution of most of the real-time data would be similar. But even with a different estimator, the model always has an accuracy higher than that of the individual base models. Hence, we can proceed with the same weights as discussed in this study.

A new dataset with five new data points was also collected, of which we did not have the final grade prediction. On passing the data points through the different models, including the stack model, we find the results as shown. Table 3 is the result of prediction using KNN, Table 4 for RF, Table 5 for XGB, Table 6 for MLP and Table 7 for the hybridised stack model. On checking with manual predictions and estimations, we find that the results of the

hybrid model, KNN and RF are correct. XGB predicts just one error in grade, whereas MLP, despite being the most structurally complex, results in the most number of errors.

Table 3. Prediction of KNN

	ENCODED	LABEL_DECODED
0	5	A
1	5	A
2	4	B
3	4	B
4	4	B

Table 4. Prediction of RF

	ENCODED	LABEL_DECODED
0	5	A
1	5	A
2	4	B
3	4	B
4	4	B

Table 5. Prediction of XGB

	ENCODED	LABEL_DECODED
0	6	S
1	5	A
2	4	B
3	4	B
4	4	B

Table 6. Prediction of MLP

	ENCODED	LABEL_DECODED
0	6	S
1	6	S
2	4	B
3	4	B
4	5	A

Table 7. Prediction of Stacked Model

	ENCODED	LABEL_DECODED
0	5	A
1	5	A
2	4	B
3	4	B
4	4	B

5. Conclusion and Future Scope

This study is intended to search for a method of developing an improved grade prediction model which performed better than the ones which already exist. The results suggest that the proposed stack of models is highly effective in improving the prediction performance by enabling a combination of the strengths of the various base models used. The accuracy of the proposed hybrid model turned out to be much higher than any of the base models, implying that the proposed hybrid model turned out to be more robust and accurate than the individual base models. This study has significant implications for academic institutions, as it can help them make informed grade predictions for improving student outcomes, as shown via testing on new and real-world data, which can, in turn, facilitate informed decision-making approaches to teaching and evaluation methods among academic institutions to maximise the knowledge extraction process among the young learning generations.

The opportunities for further research in this area, however, are still quite abundant. One such path of research could be the expansion of the proposed model by exploring various other classifiers and/or approaching a different method of hybridisation. We can also explore the effect of different variables on the final grade and suggest an improvement for highly impactful features. It would also be interesting to explore how these predictions, with the help of a few additional constraints, would help recommend different career paths based on the performance trends in each subfield and help guide the student in the right direction. But most importantly, the integration of the said hybrid model into the existing academic structure would have significant implications for the academic world and contribute to improving student outcomes.

References

[1] Raja, R. & Nagasubramani, P.. (2018). Impact of modern technology in education. Journal of Applied and Advanced Research. 3. 33. 10.21839/jaar.2018.v3iS1.165.

[2] Richiţeanu-Năstase, Elena-Ramona & Stăiculescu, Camelia. (2018). University dropout. Causes and solution. 1. 71-75.

[3] Baradwaj, B.K.; Pal, S. Mining educational data to analyze students' performance. Int. J. Adv. Comput. Sci. Appl. 2012, 2, 63–69

[4] Bain, S.; Fedynich, L.; Knight, M. The successful graduate student: A review of the factors for success. J. Acad. Bus. Ethics 2011, 3, 1.

[5] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification.

[6] Cutler, Adele & Cutler, David & Stevens, John. (2011). Random Forests. 10.1007/978-1-4419-9326-7_5.

[7] Chen, Tianqi and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016): n. pag.

[8] Popescu, Marius-Constantin & Balas, Valentina & Perescu-Popescu, Liliana & Mastorakis, Nikos. (2009). Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems. 8.

[9] Kurz, S., De Gersem, H., Galetzka, A. et al. Hybrid modeling: towards the next level of scientific computing in engineering. J.Math.Industry 12, 8 (2022).

[10] Daniel, B. Big data and analytics in higher education: Opportunities and challenges. Br. J. Educ. Technol. 2015, 46, 904–920.

[11] Namoun, A.; Alshanqiti, A. Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Appl. Sci. 2021, 11, 237.

[12] Jayaprakash, Sujith & Krishnan, Sangeetha & V, Jaiganesh. (2020). Predicting Students Academic Performance using an Improved Random Forest Classifier. 238-243. 10.1109/ESCI48226.2020.9167547.

[13] Salal, Yass & Hussain, Mushtaq & Paraskevi, T.. (2021). Student Next Assignment Submission Prediction Using a Machine Learning Approach. 10.1007/978-3-030-71119-1_38.

[14] Kanetaki, Zoe & Stergiou, Constantinos & Bekas, George & Troussas, Christos & Sgouropoulou, C.. (2022). A hybrid machine learning model for grade prediction in online engineering education.

[15] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6

[16] Rahat IS, Ghosh H, Shaik K, Khasim S, Rajaram G. Unraveling the Heterogeneity of Lower-Grade Gliomas: Deep Learning-Assisted Flair Segmentation and Genomic Analysis of Brain MR Images. EAI Endorsed Trans Perv Health Tech [Internet]. 2023 Sep. 29 [cited 2023 Oct. 2];9. https://doi.org/10.4108/eetpht.9.4016

[17] Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. EAI Endorsed Scal Inf Syst [Internet]. 2023 Sep. 21 https://doi.org/10.4108/eetsis.3937

[18] Mandava, S. R. Vinta, H. Ghosh, and I. S. Rahat, "An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population", EAI Endorsed Trans Perv Health Tech, vol. 9, Oct. 2023. https://doi.org/10.4108/eetpht.9.4052

[19] Mandava, M.; Vinta, S. R.; Ghosh, H.; Rahat, I. S. Identification and Categorization of Yellow Rust Infection in Wheat through Deep Learning Techniques. EAI Endorsed Trans IoT 2023, 10. https://doi.org/10.4108/eetiot.4603

[20] Khasim, I. S. Rahat, H. Ghosh, K. Shaik, and S. K. Panda, "Using Deep Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh", EAI Endorsed Trans IoT, vol. 10, Dec. 2023 https://doi.org/10.4108/eetiot.4579

[21] Khasim, H. Ghosh, I. S. Rahat, K. Shaik, and M. Yesubabu, "Deciphering Microorganisms through Intelligent Image Recognition: Machine Learning and Deep Learning Approaches, Challenges, and Advancements", EAI Endorsed Trans IoT, vol. 10, Nov. 2023. https://doi.org/10.4108/eetiot.4484

[22] Mohanty, S.N.; Ghosh, H.; Rahat, I.S.; Reddy, C.V.R. Advanced Deep Learning Models for Corn Leaf Disease Classification: A Field Study in Bangladesh. Eng. Proc. 2023, 59, 69. https://doi.org/10.3390/engproc2023059069

[23] Alenezi, F.; Armghan, A.; Mohanty, S.N.; Jhaveri, R.H.; Tiwari, P. Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light. Water 2021, 13, 3470. https://doi.org/10.3390/w13233470

[24] Gong, Jing and Tao-An Chen. "Does Configuration Encoding Matter in Learning Software Performance? An Empirical Study on Encoding Schemes." 2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR) (2022): 482-494.

[25] Samuels, Peter & Gilchrist, Mollie. (2014). Pearson Correlation.

[26] Senthilnathan, Samithamby. (2019). Usefulness of Correlation Analysis. SSRN Electronic Journal. 10.2139/ssrn.3416918.

[27] Muhammad Ali, Peshawa & Faraj, Rezhna. (2014). Data Normalization and Standardization: A Technical Report. 10.13140/RG.2.2.28948.04489.

[28] Abdollahi, J., Nouri-Moghaddam, B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J Comput Sci 5, 205–220 (2022).

[29] Zhu, Qiuming A.. "On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset." Pattern Recognit. Lett. 136 (2020): 71-80.

[30] Grandini, Margherita et al. "Metrics for Multi-Class Classification: an Overview." ArXiv abs/2008.05756 (2020): n. pag.