A Deep Learning-Based Framework for At-Risk Student Detection Using Educational Data Mining with a Focus on Learning Health

Reshma B Nair[1] , Dr. G. Naveen Sundar[2] , Dr. D. Narmadha[3]

[1]PhD Scholar, Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India, reshmabalunair2@gmail.com
[2]Associate Professor, Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India, naveensundar@karunya.edu
[3]Assistant Professor, Division of Artificial Intelligence and Machine Learning, Karunya Institute of Technology and Sciences, Coimbatore, India, narmadha@karunya.edu

Abstract:

Detection of at-risk students through Educational Data Mining (EDM) with emphasis on learning health seeks to determine students who are likely to underachieve or drop out by monitoring trends in their learning behavior, performance indicators, and attendance levels. Current models in this Area are plagued by several issues, such as poor capability to capture deep relational features, low generalization through shallow structures, inadequate treatment of imbalanced and high-dimensional educational data, and the absence of adaptive optimization for varied learning patterns. To overcome these constraints, this study introduces a new deep hybrid model named Harris Hawk Optimization based Capsule Deep Residual and Dense Network model (HHO-CapDeReD-Net), which integrates four strong architectures Capsule Network, Deep Neural Network, ResNet-50, and DenseNet-121 to extract multi-level features from encoded student data. The model also incorporates Harris Hawk Optimization (HHO) to fine-tune hyperparameters for optimal learning performance dynamically and minimize overfitting. By concentrating on learning health, the model continuously monitors and predicts educational risk so that early intervention and tailored help can be ensured. The intended HHO-CapDeReD-Net model shows higher prediction accuracy and stability in pinpointing students who are at risk, thus furthering more successful educational interventions and student retention.

Keywords: At-Risk Student Detection, Learning Health Analytics, Student Performance Prediction, Deep Neural Network and Capsule Network.

1. Introduction:

Over the last few years, the value of data-driven decision-making in education has increasingly become spotlighted, particularly as student-related data has become more readily available on digital platforms. One of the new areas that leverage the abundance of data is EDM [1-3]. This discipline concentrates on examining data drawn from educational environments to reveal patterns and insights that can be used to improve learning processes and results [4]. One of the most significant immediate uses of EDM is in the early identification of vulnerable students, who will likely fail academically, drop out of school, or become alienated from the educational process because of a range of cognitive, psychological, socioeconomic, and behavioral issues [5-6]. As the learning environment becomes more complex, educators realize that conventional intervention techniques are reactive, not proactive [7-8]. The inclusion of EDM makes it possible to have a more sophisticated, predictive, and preventative model of academic support, thus making it possible for education stakeholders to pick out struggling students early and shape support mechanisms appropriately [9-10]. The notion of "learning health" presents a new window of opportunity through which students' academic achievement and wellness can be monitored and optimized on an ongoing basis. Learning health, drawn from the learning health systems paradigm in medicine, focuses on the iterative process of data gathering, analysis, feedback, and action in the education system [11-12]. It combines multiple aspects of a learner's experience, cognitive performance, engagement patterns, emotional states, attendance, peer interactions, and others to build a dynamic picture of each student's educational well-being. By focusing more on learning health, researchers and educators can look beyond static performance measures such as grades and test scores and explore deeper, frequently latent factors driving a student's risk status. This more holistic method provides more precision in identifying at-risk students and makes the possibility of targeted interventions more feasible, individualized, timely, and effective [13].

The use of machine learning algorithms and data mining in the education sector enables the discovery of meaningful patterns from large, multidimensional datasets of students [14]. When trained and validated appropriately, these algorithms can forecast risk levels based on a mix of academic background, behavioral indicators, socio-demographic factors, and real-time learning analytics [15-16]. Principal techniques used are classification algorithms like decision trees, support vector machines, random forests, and neural networks that can each approximate a sophisticated relationship between input features and educational risk. Clustering, association rule learning, and deep learning add the capacity to uncover nuanced trends likely to escape the eye of the human observer. Notably, such models' interpretability and ethical use are of prime importance, particularly when decisions regarding student support and intervention can have long-term implications [13]. Furthermore, the COVID-19 pandemic and the worldwide transition to online learning environments have heightened the challenges of student disengagement, further widening the gap between high-achieving and struggling students. The digital divide, absence of social support, mental illness issues, and diminished teacher-student interaction have compounded the urgency for data-driven [14]. These automated tools can alert struggling students early enough to avoid a critical failure. In such a situation, learning health monitoring through ongoing data collection like clickstream data from Learning Management Systems (LMS), virtual classroom participation metrics, and assessments is more important than ever. Educational institutions are now confronted with the double task of ensuring academic achievement as well as fostering emotional resilience and mental health, particularly for those who are at risk of dropping through the system [16-17].

To solve these challenges, this study aims to establish a solid framework for identifying at-risk students through the synergistic application of educational data mining methods and learning health analytics. The structure combines various data sources, such as academic history, attendance records, learning behavior data, and psychosocial evaluations, to develop an inclusive student risk model. The intention is to facilitate early, individualized interventions that enhance academic achievement and promote an inclusive learning environment. By linking predictive analytics to the greater educational purpose of inclusivity and equity, this study helps shape a future in which no student falls behind due to avoidable academic deterioration. At last, the combination of EDM and learning health approaches is a paradigm shift in student support systems. It enables educators and administrators to move away from diffuse, frequently wasteful interventions toward data-driven, student-specific approaches. As schools across the globe seek to enhance retention, graduation rates, and student satisfaction, the contribution of technology and data science in guiding policy and practice cannot be overstated. In addition to investigating the technical application of at-risk student identification systems, this study focuses on their ethical use, scalability, and practical applicability. By supporting an adaptive and responsive learning environment, this strategy allows teachers to recognize and address the individual needs of each learner so that educational achievement is accessible to all.

- **HHO-CapDeReD-Net:** HHO-CapDeReD-Net is a hybrid deep model that integrates Harris HHO with CapsNet, DNN, ResNet-50, and DenseNet-121 to improve identifying at-risk students. HHO tunes important hyperparameters such as learning rate and neuron count to enhance model performance. CapsNet learns spatial relationships from data, and DNN learns deep features. ResNet-50 avoids overfitting through learning residual, and DenseNet-121 enhances gradient flow and reuse of features. The hybrid solution improves accuracy and robustness in identifying at-risk students and offers a stable framework for more effective learning support and resource distribution.

2. Literature review:

This work by Eli Nimy et al. [1] introduced a probabilistic logistic regression model for early detection of at-risk students during the academic year. The model aimed at predicting the probability of students failing future exams based on Moodle engagement, demographic, and performance variables. It facilitated the incorporation of domain knowledge and quantified uncertainty in predictions. The best performance was recorded in week 6 with 92.81% accuracy, and there was a 60% decline in prediction uncertainty between stages 3 and 5. Although performance data enhanced prediction stability, the impact of demographic and engagement data declined as the model continued. Kiran Fahd and Shah J. Miah's study [2] emphasizes predicting at-risk students in higher education using a Multilayer Perceptron (MLP) classifier. The primary objective is to improve early academic intervention by resolving prevalent issues such as small-scale modeling and data imbalance. The suggested deep learning model applies data augmentation to balance the dataset and enhance the precision of prediction. It demonstrates improved performance in recall, precision, and confusion matrix measures. The main strengths include enhanced precision and early identification of underachieving students, whereas a reported limitation is the intricacy and resource usage of deep learning adoption. Balqis Albreiki et al. [3] present a graph-based machine learning approach to identify at-risk students early by representing student data as graphs employing Euclidean and cosine distances. These graphs are enriched with topological attributes and are fed into a Graph Convolutional Network (GCN), enhancing prediction and AUC accuracy. The method beats previous techniques, achieving up to 87.4% accuracy and 0.97 AUC. Though it achieves more insights and higher performance, it is more complex and has higher computation costs than less complex models. Mustapha Skittou et al. [4] work deals with creating an Early Warning System (EWS) based on the K-Nearest Neighbor (KNN) algorithm that can forecast drop-out-risk students. It considers socio-cultural, structural, and educational determinants and achieves above 99% accuracy in the training and test processes. Its highest prediction accuracy is the primary strength, overfitting potential and lower generalizability across other learning environments are possible limitation

2.1 Challenges:

- The majority of existing models available cannot fully include socio-cultural, structural, and education variables, which are essential for explaining student dropout behavior [2].
- Previous models can be limited in the completeness of dropout indicator selection, causing incomplete or lower accuracy predictions [3].
- They tend to rely on general or standard datasets that do not necessarily reflect specific or local dropout patterns, lowering the models' applicability in other settings [4].
- Specific models lack the presentation of results in terms that are intelligible to planners or teachers, thus making it difficult to utilize them practically within actual educational contexts [6].
- High performance on training data, as demonstrated with KNN could result in overfitting, i.e., the model will not perform well on unseen or diverse data [7].

3. Proposed methodology for At-Risk Student Detection using the HHO-CapDeReD-Net model:

The research aims to develop an intelligent deep learning framework, HHO-CapDeReD-Net, for At-Risk Student Detection Using Educational Data Mining with a Focus on Learning Health, aimed at accurately identifying students at academic risk. The methodology begins with collecting and preprocessing a rich student performance dataset [18], incorporating academic, behavioral, and demographic indicators. Data is normalized using Standard Scaler, and key features are extracted using an autoencoder to reduce dimensionality while retaining essential patterns. Feature extraction is achieved via an autoencoder and dimensionality reduction while retaining useful patterns. The strength of the methodology is in the CapDeReD-Net model. This hybrid deep learning architecture blends Capsule Network, DNN, ResNet-50, and DenseNet-121 to efficiently capture spatial, deep, residual, and dense features from educational information. The HHO is utilized to fine-tune hyperparameters to optimize the model, balancing exploration and exploitation towards optimal predictive performance. EDM's preventive, health-related strategy guarantees personalized learning approaches and promotes enhanced student achievement. The architecture for the proposed At-Risk Student Detection is shown in figure 1.

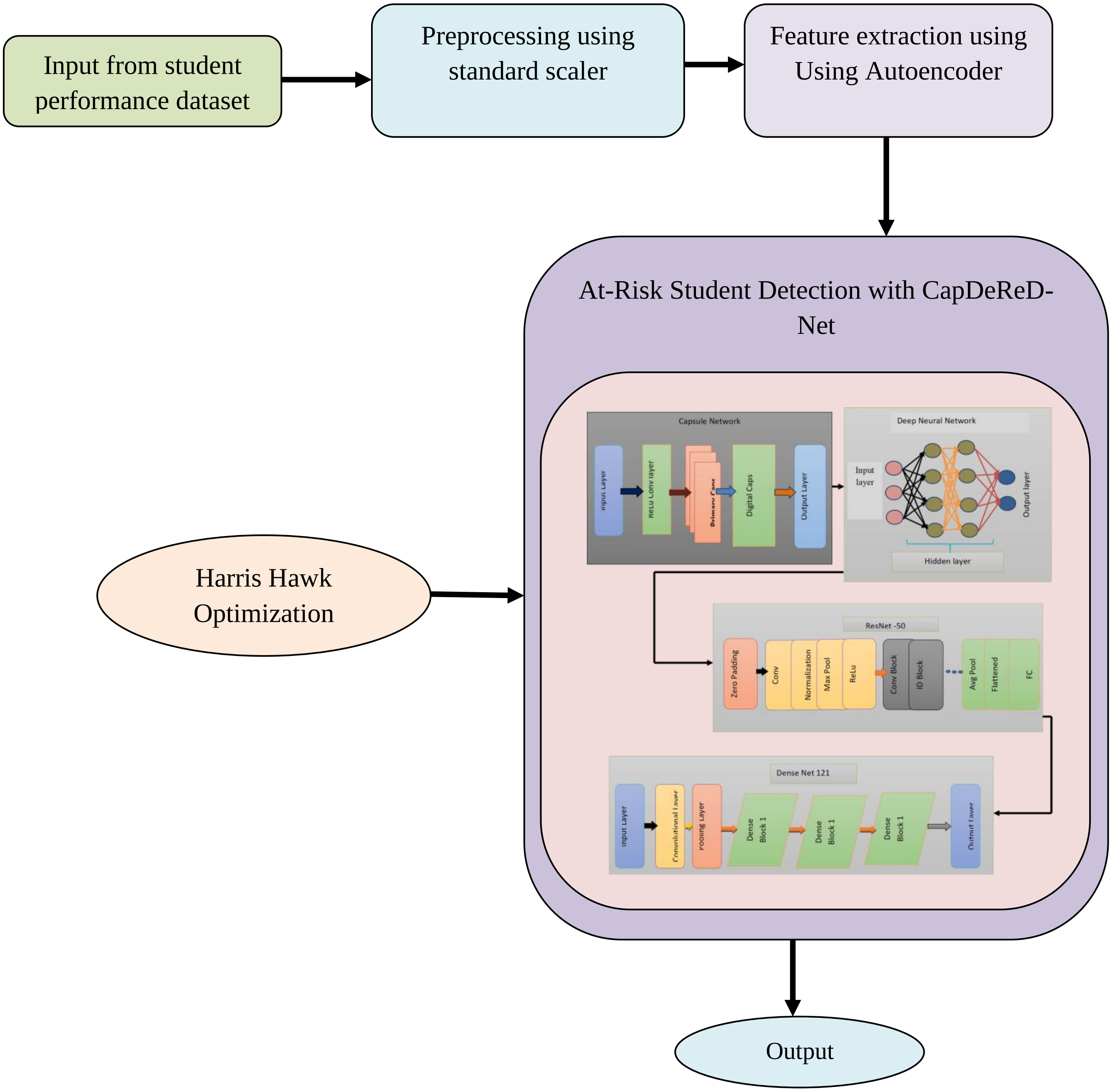

Figure 1: Architecture for the proposed At-Risk Student Detection

3.1 Input collection from student performance dataset: The input data for this study are obtained from a holistic student performance dataset [18], which consists of a range of academic and behavioral measures. These include assignment grades, attendance, exam scores, participation rate, and other related learning measures. Mathematically, the dataset can be defined as a feature matrix:

$$E = \{(y_j, z_j) \mid y_j \in \Re^n, z_j \in \{0,1\}\}_{j=1}^{m} \quad (1)$$

Where D is the complete dataset, y_j is an n-dimensional feature vector representing the academic and behavioral profile of the student j, $z_j \in \{0,1\}$ is the binary class label indicating whether the student is at-risk (1) or not (0), and m is the total number of students in the dataset. Each y_j encapsulates normalized feature values from domains (e.g., coursework, exams, engagement), forming the input space for further analysis.

3.2 Preprocessing Using Standard Scaler: After the student performance dataset is gathered, the second important step is data preprocessing to ensure uniformity in the feature space. As raw input features can widely differ in scale (e.g., attendance in percentage, scores in marks, frequency as counts), normalization is required to scale all features to a common range. This is done through the Standard Scaler, which normalizes every feature by subtracting the mean and scaling to unit variance. The Standard Scaler reshapes every feature dimension y_{jk} in the input matrix $y_j \in \Re^n$ in the following way:

$$y'_{jk} = \frac{y_{jk} - \mu_k}{\sigma_k} \quad (2)$$

Where y_{jk} is the value of the feature for the j^{th} student, μ_k is the mean of the k^{th} feature across all students, σ_k is the feature's standard deviation k^{th}, and y'_{jk} is the standardized feature value. As a result, the preprocessed dataset becomes:

$$E' = \{(y'_j, z_j)\}_{j=1}^{m} \text{, where } y'_j = \left[\frac{y_{j1} - \mu_1}{\sigma_1}, \frac{y_{j2} - \mu_2}{\sigma_2}, \ldots \frac{y_{in} - \mu_n}{\sigma_n}\right] \quad (3)$$

This normalization guarantees that every feature has a zero mean and a unit variance, enhancing the convergence rate and stability of the deeper learning models that follow, particularly the scale-sensitive models like Capsule Networks and Dense networks in the CapDeReD-Net architecture.

3.3 Feature Extraction Using Autoencoder: Following normalization, the normalized dataset derived from preprocessing is referred to as:

$$E' = \{(y'_j, z_j)\}_{j=1}^{m} \quad (4)$$

Each $y'_j \in \Re^n$ is a standardized feature vector with zero mean and unit variance. This preprocessed input data is fed to the next vital step: feature extraction with an autoencoder. An autoencoder is an unsupervised neural network that discovers compact, efficient representations of data by learning to reconstruct it in a lower dimension. It consists of two key components: an encoder function and a decoder function. The encoder maps the input to a lower-dimensional latent representation, and the decoder tries to reconstruct the input from this representation.

3.3.1 Mathematical formulation:

Let $y' \in \Re^n$ be the standardized input vector. The encoder and decoder functions are given by:

Encoder: $$a = f(y') = \sigma(Wy' + b) \quad (5)$$

Here, $W \in \Re^{e \times n}$ it represents the encoder weight matrix, $b \in \Re^{d}$ the encoder bias vector, $\sigma(.)$ an activation function (e.g., ReLU or sigmoid), and $a \in \Re^{e}$ the latent representation (with $e < n$).

Decoder:

$$\hat{Y}' = h(a) = \sigma(W' z + b') \tag{6}$$

Here $W' \in \Re^{n \times e}$ is the decoder weight matrix, $b' \in \Re^{n}$ the decoder bias vector, and $\hat{y}' \in \Re^{n}$ the reconstructed input vector.

Loss function: The autoencoder learns to minimize the reconstruction error between the original standardized input and its reconstruction. The loss function is generally the mean squared error (MSE):

$$L_{AE} = \left\| y' - \hat{y}' \right\|^{2} \tag{7}$$

This goal stimulates the network to retain the key information and reject noise and redundancy. After training, the final latent vector is the extracted feature representation for every sample. So, the feature extraction step yields a transformed dataset:

$$E'' = \{(a_j, z_j)\}_{j=1}^{m}, \text{ where } a_j = f(y'_j) \tag{8}$$

This low-dimensional latent representation a serves as the optimized input for the classification model during the subsequent phase of the framework.

3.4 At-Risk Student Detection with CapDeReD-Net Using Educational Data Mining Techniques:

The CapDeReD-Net model of At-Risk Student Detection uses EDM techniques to learn health analytics. It encodes student data through an encoder to produce a latent feature vector fed into four deep learning branches: Capsule Network, Deep Neural Network, ResNet-50, and DenseNet-121. Each branch learns distinct features from the data, detecting intricate patterns in student performance. The outputs of these branches are combined into one vector, passing through a last classification layer with Softmax activation to predict the student's risk level. This model will identify at-risk students early on so that the intervention can occur quickly.

3.4.1 Capsule Network: The Capsule Network handles features by encoding them into vector outputs that indicate patterns' presence and orientation. It employs routed vectors and dynamic routing to infer how much lower-level features drive higher-level features. These influences are regulated through coupling coefficients. The resultant vectors are normalized via a squashing function to cap their lengths within 0 to 1. The output is a collection of capsule vectors, each dedicated to a specific high-level feature employed for classification in the CapDeReD-Net architecture.

Predicted vector

$$\hat{v}_{k|j} = W_{jk} v_j, \quad \text{Where } v_j = a$$

Coupled output

$$v_k = \sum_j d_{jk} \hat{v}_k |_j \tag{9}$$

Squashing output

$$w_k = \frac{\|v_k\|^2}{1+\|v_k\|^2} \cdot \frac{v_k}{\|v_k\|} \tag{10}$$

CapsNet output

$$F_{caps} = \left[w_1, w_2, \ldots\ldots\ldots\ldots w_n\right] \tag{11}$$

Where, n = number of output capsules.

3.4.2 Deep Neural Network: The DNN of CapDeReD-Net receives the extracted features as inputs and executes them through successive hidden layers. Each layer performs a linear transformation followed by a non-linear activation function to learn the complex patterns in the data. Information transformation from one layer to the next aids the model in learning higher-level representations. The final hidden layer output becomes the DNN's feature representation that gets sent through further fusion with outputs of other networks in the architecture.

$$h^{(m)} = \sigma\left(W^{(m)} h^{(m-1)} + b^{(m)}\right), \quad h^{(0)} = a \tag{12}$$

Final DNN output

$$F_{DNN} = h^{(M)} \tag{13}$$

Where, M = number of hidden layers.

3.4.3 ResNet-50: ResNet-50 in the CapDeReD-Net model employs residual blocks to provide additional feature learning by enabling the network to learn the transformation changes over the input instead of the entire transformation. A residual block consists of convolution, batch normalization, and ReLU activation layers. A shortcut or identity connection avoids these layers and adds the input directly to the transformation output. This method retains valuable information and reduces problems like vanishing gradients in deep networks. The last output of the ResNet-50 module is the learned feature representation, which is later utilized in the fusion process with other network outputs.

Residual block operation:

$$Z = F\left(a, \{W_j\}\right) + a \tag{14}$$

Here, F it consists of Conv-BN-ReLU layers and a is an identity connection.

ResNet output:

$$F_{\mathrm{Re}\,sNet} = Z \tag{15}$$

3.4.4 DenseNet-121: DenseNet-121 promotes feature propagation by interconnecting each layer to all other layers feed-forwardly. Rather than merely taking input from the immediately preceding layer, every layer takes the feature maps of all previous layers. This architecture promotes feature reuse and drastically cuts down the number of parameters while retaining high efficiency. Each transformation in a layer usually involves batch normalization, ReLU activation, and convolution. The first input to the network is propagated through all layers in this highly connected manner, and the ultimate output from the last layer is used as the feature representation of the DenseNet module, which is added to the fusion in CapDeReD-Net.

Each layer takes input from all the preceding layers:

$$y_m = I_{m,}\left(y_0, y_1, \ldots\ldots. y_{m-1}\right) \tag{16}$$

Here, $y_0 = a$ and $I_m = BN - \mathrm{Re}\,LU - Conv$

DenseNet output:

$$F_{DenseNet} = y_m \tag{17}$$

3.4.5 Fusion Strategy in CapDeReD-Net: In a fusion strategy for CapDeReD-Net, the feature outputs of all four networks Capsule Network, Deep Neural Network, ResNet-50, and DenseNet-121 are fused into one unified feature representation. This is achieved by a concatenation operation that combines the multi-modal feature sets, preserving spatial and semantic information. The concatenated features are then fed through a final fully connected layer that learns to integrate these diverse representations. Softmax activation at the end gives the final class prediction probabilities as the model confidence in each output class. This combining method improves the classification accuracy by using the strengths of each architecture.

Concatenate all the features extracted:

$$F_{fusion} = Concat\left(F_{caps}, F_{DNN}, F_{\mathrm{Re}\,sNet}, F_{DenseNet}\right) \tag{18}$$

Apply final fully connected layer + Softmax:

$$z = Soft\max\left(W_f F_{fusion} + b_f\right) \tag{19}$$

Where: W_f denotes the Fusion weights, b_f the fusion bias, and z the Final class prediction probabilities (o/p classes).

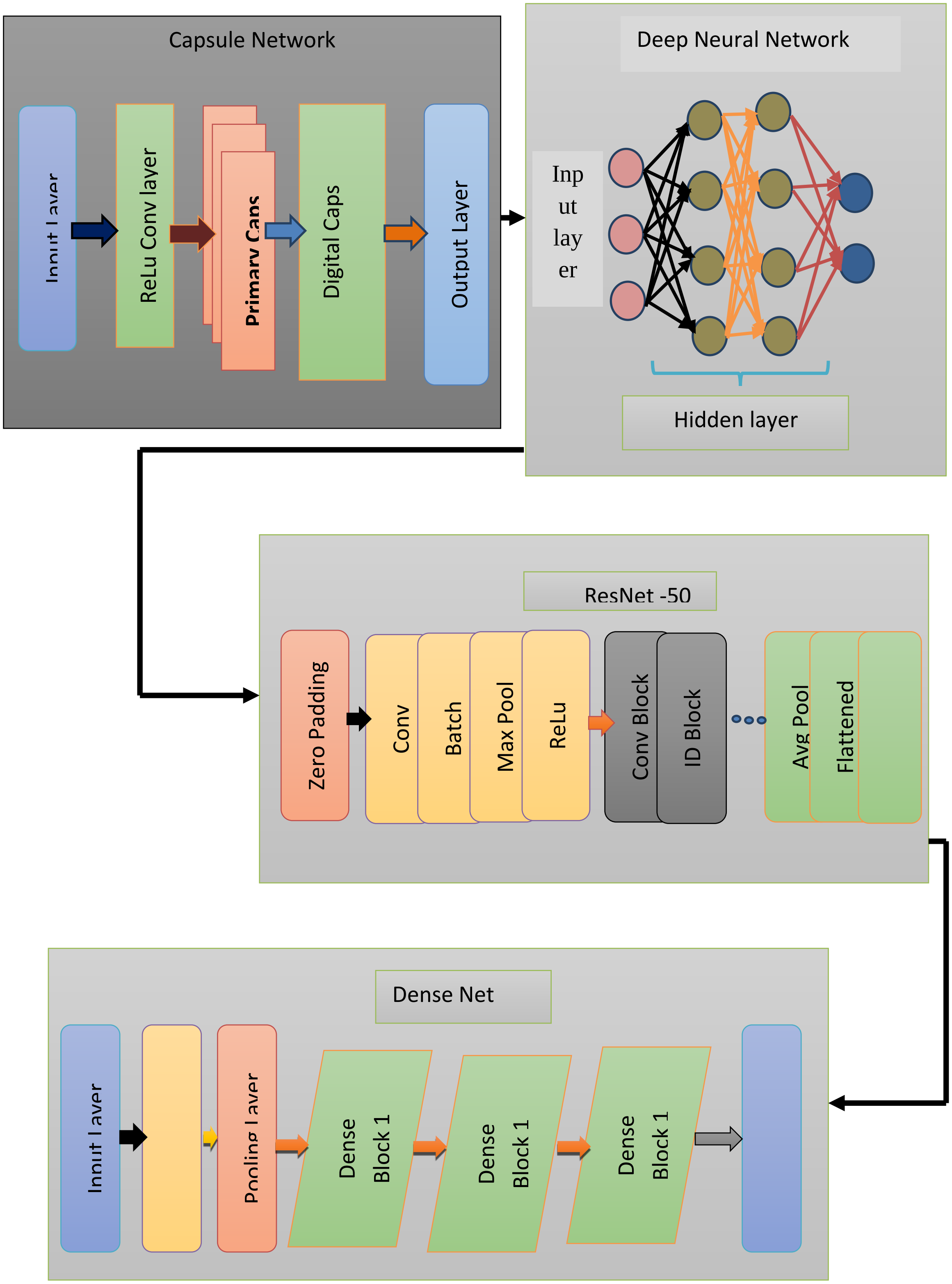

Figure 2: Architecture for the proposed CapDeReD-Net model

4. Harris Hawk Optimization: HHO significantly optimizes the model performance by adjusting its hyperparameters. The HHO algorithm, based on the collaborative hunting mechanism of Harris hawks, is implemented to strike a balance between exploration and exploitation in the hyperparameter search space. The algorithm uses exploration to look through unexplored larger sets of possible solutions, while exploitation uses the search around promising solutions to refine them. Through dynamic tuning of the model's hyperparameters, HHO ensures that the CapDeReD-Net model is optimized to produce the highest possible predictive accuracy in identifying at-risk students, depending on grades, behavior, and socioeconomic information. This optimization process enhances the model's capacity to classify students into correct risk categories and enables more effective intervention strategies.

4.1 Key Phases of HHO Algorithm:

4.1.1 Exploration Phase: During the exploration phase of the HHO algorithm, the hawks, being candidate solutions, try to find promising areas in the search space while moving randomly. Random movement prevents the algorithm from getting trapped in local optima and raises the probability of finding better solutions at the global level. In this step, every hawk updates its location by taking a randomly selected hawk's location and updating its own based on random coefficients. These random values add randomness and guarantee varying exploration patterns. The primary aim of this step is to promote broad coverage of the solution space before the algorithm concentrates more intensely on improving the best-found solutions in subsequent steps.

$$Y(t+1) = Y_{rand}(t) - r_1 . \left| Y_{rand}(t) - 2_{r_2} . Y(t) \right| \tag{20}$$

4.1.2 Exploitation Phase: In the exploitation stage of the HHO algorithm, the hawks concentrate on tuning their search close to the optimal solution discovered until now, referred to the prey. This stage focuses on convergence by altering each hawk's position according to the location of the prey, thereby reducing the search space to utilize high-quality solutions. The movement of every hawk depends on the escaping energy of the prey, representing how hard or easy it is to capture the prey, and the jump strength, representing how hard the hawk will move toward the prey. This mechanism makes the hawks execute adaptive, planned updates on their positions, progressively converging toward the best solution while being able to escape local optima when needed.

$$Y(t+1) = Y_{prey}(t) - F . \left| K . Y_{rand}(t) - Y(t) \right| \tag{21}$$

4.1.3 Fitness Function: Fitness for each solution is calculated based on its performance for classification based on a loss function. The categorical cross-entropy loss function has been employed in this setting for measuring model performance at each step:

$$Fitness(Y) = -\sum_{j=1}^{n} z_j . \log\left(\hat{z}_j \right) \tag{22}$$

Here, the actual label of the j^{th} instance is represented, and the predicted probability for the j^{th} example is determined. The objective is to minimize fitness, which is equivalent to reducing classification loss (categorical cross-entropy).

4.1.4 Hyperparameter Tuning: Throughout the process of optimization, the HHO algorithm tunes the hyperparameters of the model (e.g., learning rate, batch size, and network parameters) to reduce the classification loss and increase the performance of the model. As the process goes on, it tests candidate solutions, adjusts the positions of the hawks based on the exploration and exploitation stages, and finally converges toward an optimal solution set of hyperparameters. This hyperparameter optimization process allows the CapDeReD-Net model to effectively predict at-risk students by determining the best possible configuration of model hyperparameters, thereby improving the overall performance of the detection framework.

5. Result and discussion: The proposed model performs better in identifying at-risk students, with quick convergence in the analysis of fitness value. The confusion matrix reveals high classification accuracy in Low, Medium, and High-risk classes with fewer misclassifications. The ROC curve, with AUC = 1.00 for all classes, reveals perfect classification. The classification metrics (Precision, Recall, F1 Score, and Accuracy) show uniformly high values for all classes, particularly for the Medium class. Comparison with standard classifiers (SVM, Naïve Bayes, Decision Tree, and Neural Network) shows the proposed model's substantial superiority in all the evaluation metrics, highlighting its strength and reliability in educational use.

5.1 Dataset Description:

Student performance dataset [18]: The High School Student Performance Dataset comprises academic and demographic information from two Portuguese high schools, emphasizing student performance prediction. The dataset includes variables such as students' Math, Portuguese grades and socio-demographic variables such as age, gender, family background, and social variables. Of particular interest are the variables of study time, family support, education of parents, and extracurricular activities. The target variable, final_grade, strongly correlates with grades during the first and second periods. The data enable the examination of many factors affecting academic performance and can be applied to predictive modeling to determine which students are at risk. It offers valuable information for planning education and intervention.

5.2 Best Fitness over Iterations: This bar chart shows the optimization algorithm's performance over 100 iterations. The fitness value, which measures how effective a particular model or solution is in detecting at-risk students, exhibits considerable decline during the early iterations, as there is fast convergence towards an optimal or near-optimal solution. As the iterations continue, the fitness value plateaus, reflecting that the model has effectively learned the decision boundaries and patterns in the data. This overlap shows that the optimization algorithm applied is both practical and can locate high-performance solutions in the early stages of training. The graphical representation for Best Fitness over Iterations is illustrated in Figure 3.

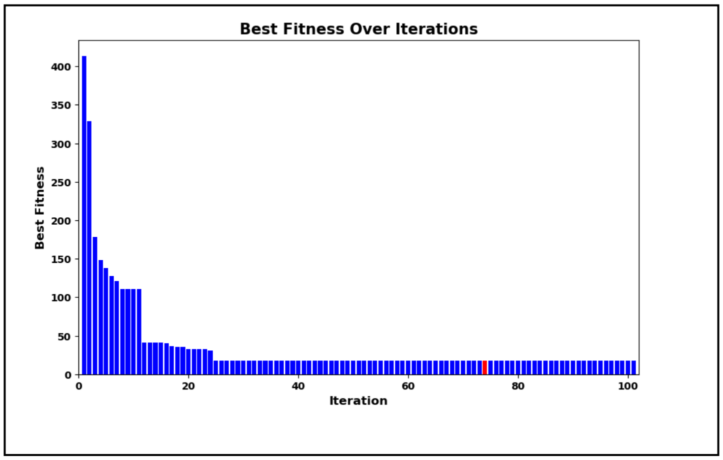

Figure 3: Graphical representation for Best Fitness over Iterations

5.3 Confusion Matrix: The confusion matrix offers information on the classification performance of the introduced model on three risk classes: Low, Medium, and High. The model accurately classified most instances of each class, with just one misclassification for the Low-risk class. The prominent diagonal structure in the matrix testifies to the model's high recall and precision rate for all classes, especially for Medium- and High-risk students. Such reliability is paramount in educational contexts where accurately identifying at-risk learners can result in early intervention and better learning outcomes. A graphical representation of the confusion matrix is depicted in Figure 4.

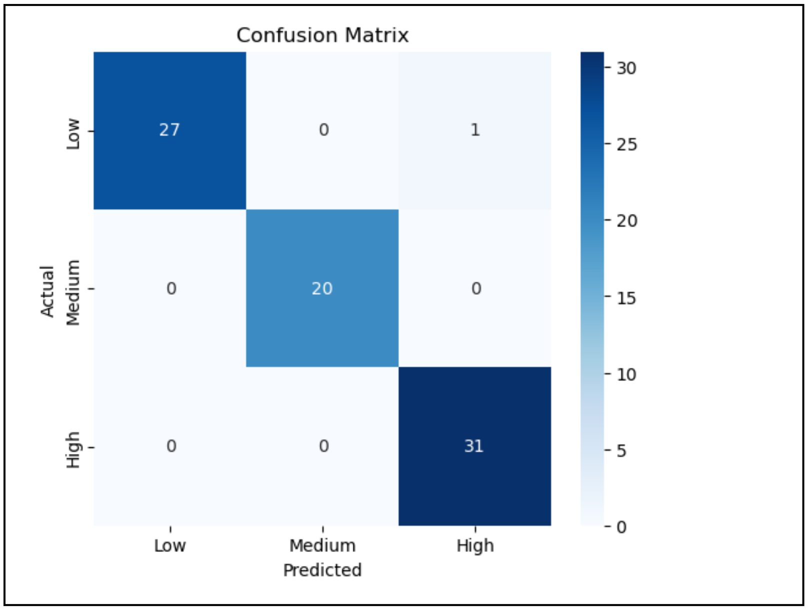

Figure 4: Graphical representation of confusion matrix

5.4 ROC Curve for Each Class: The ROC curve demonstrates the model's classification ability by plotting each class's actual positive rate versus the false positive rate. Interestingly, the Area under the Curve (AUC) for all three classes is 1.00, indicating perfect classification performance. These very high AUC values suggest that the model possesses an excellent capacity to separate at-risk and non-at-risk students, validating the strength of the detection mechanism. Combined with the earlier results, this precision suggests that the system can be effectively used in real-life educational settings where accuracy is essential. A graphical representation of the ROC Curve for Each Class is represented in Figure 5.

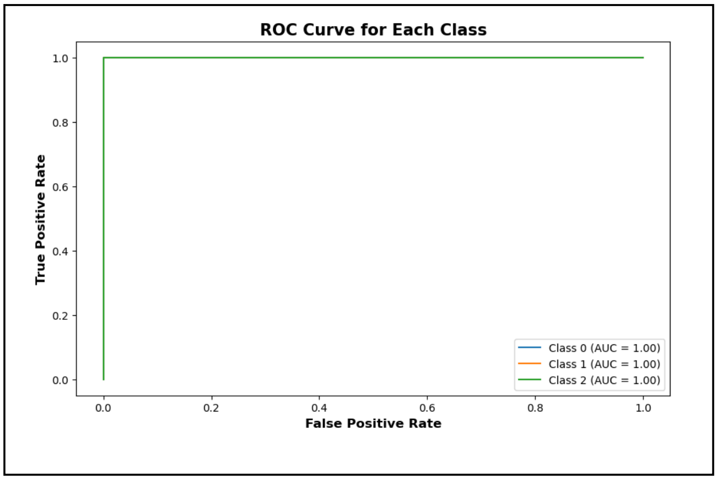

Figure 5: Graphical representation of ROC Curve for Each Class

5.5 Bar Chart – Classification Metrics per Class:

The bar chart presents the model's classification performance for individual classes Low, Medium, and High based on four metrics: Precision, Recall, F1 Score, and Accuracy. The Medium class is highlighted with perfect scores (1.0) in all metrics, indicating superb model performance in predicting this class. The Low and High classes also display good performance, where all the metrics values approach 0.97–1.0, representing a well-balanced classification for all the classes. The above graph reflects how the model has good accuracy and prediction both overall and individually in each class so that no class is left out or misclassified. Graphical representation for Classification Metrics per Class is illustrated in Figure 6.

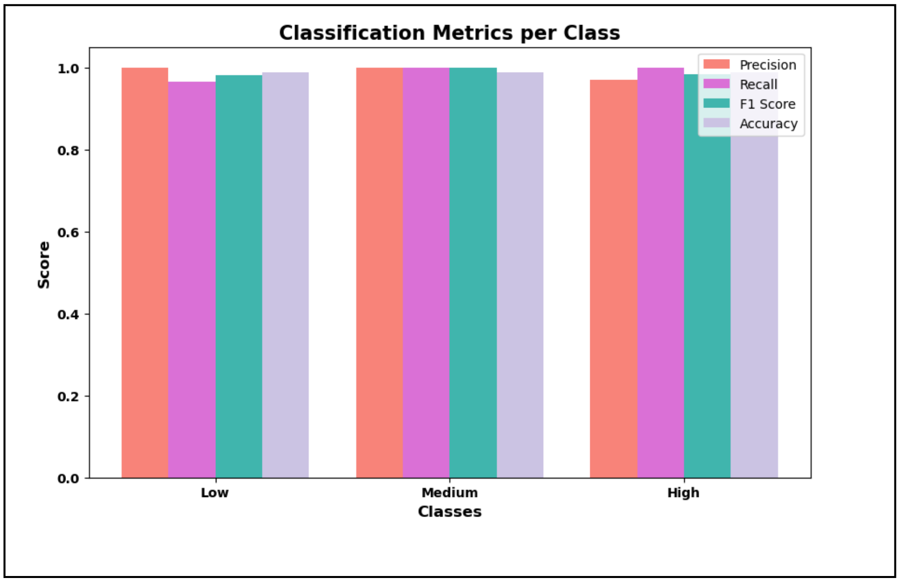

Figure 6: Graphical representation for Classification Metrics per Class

5.6 Horizontal Lollipop Chart – Comparison of Classifier Performance:

This horizontal lollipop chart is a comparative representation of five classifiers, SVM [19], NB [20], DT [21], NN [22], and the developed model based on four performance metrics: Precision (blue), Recall (green), F-Score (red), and Accuracy (peach). The Proposed model outperforms every other model with almost perfect values of 0.99 on all metrics, which speaks of its reliability and stability. Conventional models reflect significantly lower scores, especially on Precision and F-Score, ranging from 0.27 to 0.45. Although NN reflects comparatively better recall and accuracy than traditional models, none even approaches the stability of the proposed model. This chart highlights the superior performance and consistency of the proposed approach over conventional classifiers. Graphical representation for Classification Metrics per Class is shown in Figure 7.

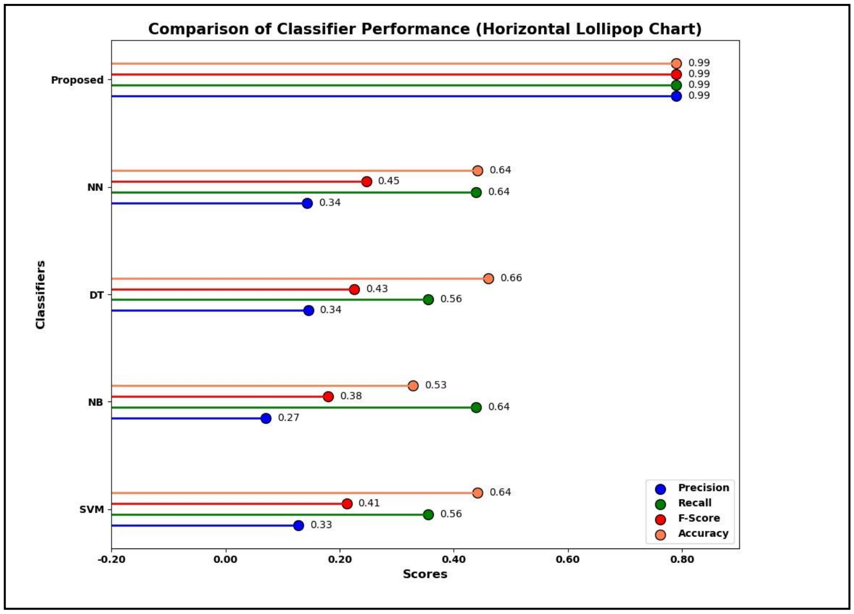

Figure 7: Graphical representation for Classification Metrics per Class

6. Conclusion:

In this research, a novel deep learning framework, HHO-CapDeReD-Net, was developed to enhance the detection of at-risk students by leveraging the principles of EDM with a focus on learning health. The HHO-CapDeReD-Net model efficiently combines four architectures, CapsNet, DNN, ResNet-50, and DenseNet-121, to discover rich and high-level features from intricate educational data. For further performance improvement of the model, the HHO algorithm was utilized for hyperparameter fine-tuning, facilitating adaptive learning and improved generalization on different academic behaviors. Incorporating learning health guarantees active and ongoing monitoring of student performance, enabling early interventions and individualized learning plans. Experimental results show that HHO-CapDeReD-Net is more accurate, robust, and predictive than conventional approaches, making it essential to intelligent educational systems and student success programs. The model can be improved further by adding behavioral and psychological properties, allowing more comprehensive risk categorization. It also targets real-time application in learning environments, broadens applicability to culturally diverse and multilingual datasets, and incorporates explainable AI methods to enhance transparency and assist educators with understanding risk factors.

References:

[1] Nimy, E., Mosia, M., & Chibaya, C. (2023). Identifying at-risk students for early intervention—A probabilistic machine learning approach. Applied Sciences, 13(6), 3869.

[2] Fahd, K., & Miah, S. J. (2023). Effectiveness of data augmentation to predict students at risk using deep learning algorithms. Social Network Analysis and Mining, 13(1), 113.

[3] Albreiki, B., Habuza, T., & Zaki, N. (2023). Extracting topological features to identify at-risk students using machine learning and graph convolutional network models. International Journal of Educational Technology in Higher Education, 20(1), 23.

[4] Skittou, M., Merrouchi, M., & Gadi, T. (2023). Development of an early warning system to support educational planning process by identifying at-risk students. IEEE Access, 12, 2260-2271.

[5] L.You,‘‘Constructionofearly warningmechanismofuniversityeducation networkbasedontheMarkovmodel,’’MobileInf.Syst.,vol.2022,pp. 1–9, Jul. 2022, doi: 10.1155/2022/7302623.

[6] D. Baneres, A. E. Guerrero-Roldán, M. E. Rodríguez-González, and A. Karadeniz, ‘‘A predictive analytics infrastructure to support a trustwor thy early warning system,’’ Appl. Sci., vol. 11, no. 13, p. 5781, Jun. 2021, doi: 10.3390/app11135781.

[7] A.K.Veerasamy, M. J. Laakso, D. D’Souza, and T. Salakoski, ‘‘Predictive models as early warning systems: A Bayesian classification model to iden tify at-risk students of programming,’’ in Intelligent Computing (Lecture Notes in Networks and Systems), vol. 284. Cham, Switzerland: Springer, 2021, pp. 174–195, doi: 10.1007/978-3-030-80126-7_14.

[8] D. Baneres, A. Karadeniz, A. E. Guerrero-Roldan, and M. E. Rodriguez, ‘‘A predictive system for supporting at-risk students’ identification,’’ in Proc. FutureTechnol. Conf. (AdvancesinIntelligent SystemsandComput ing), vol. 1288, 2021, pp. 891–904, doi: 10.1007/978-3-030-63128-4_67.

[9] M. B. Ada and K. Turinicova, ‘‘Developing a dual dashboard early detec tion system,’’ in Proc. IEEE 20th Int. Conf. Adv. Learn. Technol. (ICALT), Jul. 2020, pp. 155–157, doi: 10.1109/ICALT49669.2020.00052.

[10] D. Bañeres, M. E. Rodríguez, A. E. Guerrero-Roldán, and A. Karadeniz, ‘‘An early warning system to detect at-risk students in online higher education,’’ Appl. Sci., vol. 10, no. 13, p. 4427, Jun. 2020, doi: 10.3390/app10134427.

[11] M.Chunqiao,Y.Niefang,andP.Xiaoning,‘‘Methodandsystemconstruct ing for learning situation early warning based on data mining techniques,’’ in Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE), Aug. 2019, pp. 964–969, doi: 10.1109/ICCSE.2019.8845370.

[12] Martin,O.Chapter1: ThinkingProbabilistically. In Bayesian Analysis with Python: Introduction to Statistical Modeling and Probabilistic Programming Using PyMC3 and ArviZ, 2nd ed.; Packt Publishing Ltd: Birmingham, UK, 2018; pp. 18–19.

[13] Akçapınar, G.; Hasnine, M.N.; Majumdar, R.; Flanagan, B.; Ogata, H. Developing an early-warning system for spotting at-risk students by using eBook interaction logs. Smart Learn. Environ. 2019, 6, 1–15. [CrossRef]

[14] Adnan,M.; Habib, A.; Ashraf, J.; Mussadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. IEEE Access 2021, 9, 7519–7539. [CrossRef]

[15] Iatrellis, O.; Savvas, I.K.; Fitsilis, P.; Gerogiannis, V.C. A two-phase machine learning approach for predicting student outcomes. Educ. Inf. Technol. 2021, 26, 69–88. [CrossRef]

[16] Campbell, J.P.; DeBlois, P.B.; Oblinger, D.G. Academic analytics: A new tool for a new era. EDUCAUSE Rev. 2007, 42, 40.

[17] Rao, S.; Poojary, P.; Somaiya, J.; Mahajan, P. A comparative study between various preprocessing techniques for machine learning. Int. J. Eng. Appl. Sci. Technol. 2020, 5, 431–438. [CrossRef]

[18] Student performance dataset: https://www.kaggle.com/datasets/dillonmyrick/high-school-student-performance-and-demographics

[19] Pek, R. Z., Özyer, S. T., Elhage, T., Özyer, T., & Alhajj, R. (2022). The role of machine learning in identifying students at-risk and minimizing failure. IEEE Access, 11, 1224-1243.

[20] Hung, J. L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: A multistage approach. IEEE Transactions on Learning Technologies, 12(2), 148-157.

[21] Pek, R. Z., Özyer, S. T., Elhage, T., Özyer, T., & Alhajj, R. (2022). The role of machine learning in identifying students at-risk and minimizing failure. IEEE Access, 11, 1224-1243.

[22] Yang, Z., Yang, J., Rice, K., Hung, J. L., & Du, X. (2020). Using convolutional neural network to recognize learning images for early warning of at-risk students. IEEE Transactions on Learning Technologies, 13(3), 617-630.