# Classification of Basic Handwritten Mathematical Symbols using Conventional Machine Learning Techniques

[1] Vishal Balaji Sivaraman,
Electrical and Computer Engneering,
University of Florida,
Gainesville, United States,
vi.sivaraman@ufl.edu

[1] Bhagya Raj Varadaraju,
Compute and Information Science and Engineering,
University of Florida,
Gainesville, United States,
varadaraju.b@ufl.edu

[1] Ganeson Ravichandran,
Computer and Information Science and Engineering,
University of Florida,
Gainesville, United States,
g.ravichandran@ufl.edu

[1] Sai Pavan Kalyan Munaga,
Computer and Information Science and Engineering,
University of Florida,
Gainesville, United States,
saipavank.munaga@ufl.edu

*Abstract*— **Mathematical characters play a significant part in the development of formulae used during sophisticated computations. Therefore, it is crucial to identify these handwritten mathematical characters. As handwriting styles vary a lot between different individuals, a simple character like '+' can have numerous versions when written in unique ways. This makes the selection of an appropriate algorithm for the classification of handwritten mathematical symbols paramount. So, this paper experiments and compares various machine learning approaches like ensemble algorithms and transfer learning for classifying 10 basic symbols. Furthermore, this comparison will aid in assessing the pros and downsides of each approach based on their accuracy output for our custom handwritten images dataset.**

***Keywords—Data Modeling; Cross Validation; Classification; Computer Vision; Mathematical Symbols; Ensemble learning; Deep Neural Networks; Transfer learning; Metrics; Time Complexity.***

## I. INTRODUCTION

Machine learning [1] is the science of learning patterns in the sample data and predicting the outcomes of future data samples. There are various types of approaches like supervised learning, unsupervised learning, semi-supervised learning, and Artificial neural networks. The goal of supervised learning is to learn the rules that maps example inputs to their targets. In unsupervised learning, there are no labels, the algorithm tries to discover the hidden pattern in the data. Clustering is one of such algorithms. Semi-supervised learning falls between the unsupervised and supervised learning. Artificial neural networks are computing systems with connected units(neurons), inspired by the neural networks inside the brain. Currently neural networks are being used to solve various tasks like speech recognition, medical diagnosis, and computer vision. To classify handwritten mathematical symbols, various techniques like XGBoost, Random Forest and CNN have been used in this paper.

## II. MACHINE LEARNING TECHNIQUES

### *A. Ensemble Methods*

Ensemble learning [2] is a class of algorithms that combine the predictions from different models to improve the performance. There are various types of techniques like bootstrap aggregating, boosting, and stacking, based on how they combine the underlying models. One of the algorithms used in this paper is a random forest which is a popular bagging (bootstrap aggregating) technique. In this algorithm, multiple decision trees are trained on different samples of same dataset and all the predictions are averaged to get the model output. Another ensemble method called gradient boost was also used in our experiments. It is a type of boosting technique, where models are sequentially added to predict and correct the errors of the previous model until no improvement is seen. We have used the XGBoost library for this algorithm.

### *B. Deep Learning*

Deep Learning is an extension of Multi-Layer Perceptrons [3] with different number of units (perceptrons) stacked as layers on top of one other. Each layer learns a feature of input and the successive layers combine those learnings to extract complex features eliminating the need for feature engineering. Transfer learning [4]is a subset of machine learning that focuses on acquiring and applying knowledge learned while solving one issue to a distinct but related topic. This is done by saving the information received from the training on a task to tackle a slightly different but comparable challenge hence lowering the training time. Since the convolutional neural network used for image recognition tasks are more general in the layers near the input and becomes specialized to the dataset at the output layer. As a result, the weights in the previous layers of the pre-trained model are preserved and transferred to a second job for learning the characteristics unique to the new dataset. Furthermore, transfer learning has a number of applications, including but not limited to Image Recognition, Natural Language Processing, and Speech Recognition. ResNets, YOLO, Inception, Xception, the VGG Family of models, and many more are well-known transfer learning techniques used in the business world. The research focuses on explaining the performance of a pre-trained residual network model on the supplied custom dataset using the aforementioned models.

## III. RELATED WORK

The research which examines various feature selection and extraction strategies for analyzing Hand Character images, using a novel classification framework consisting of a feature selection/extraction technique for identifying the most appropriate set of features, followed by a classification algorithm from a list of classification approaches titled statistical classifiers, syntactic classifiers and neural networks [5], yields optimal performance when compared to other methods. In addition, the research that proposes an autonomous framework consisting of image segmentation and model evaluation using a distinct model titled Squeeze and Excitation ResNet [6] for classification of brain tumor from MRI images achieves 89.93% accuracy without data augmentation and 93.83% accuracy with data augmentation, respectively.

## IV. Dataset

The input dataset as is has a uniform distribution across all the 10 symbols [7] with 903 samples per class. Since the dataset is collected through the cumulative manual effort of multiple individuals, it is bound to have some errors. Random assessment of samples revealed considerable portion of the samples were mislabelled. To ease the identification and correction of such samples, the samples, and the target vectors were sorted based on the targets, and samples of each target were split into 9 chunks with 100 - 101 per chunk and were displayed as a 10 * 10 matrix along with their original index. Some images were distorted significantly and some don't belong to any of the labeled classes indicating outliers. The resultant distribution of samples could be approximated to be uniform with partial representing the highest with 909 samples and pi (math constant) representing the lowest with 810 samples. The outliers are regarded as a special class with 234 samples. This forces the requirement for stratification during the splitting of data in train and validation.

## V. Model Implementation

This section of the report briefs the implementation of f the models used for analyzing different classification techniques. The input sample assembled each image as a column. Each image of 300 * 300 pixels was flattened to a single dimension of 90000. The train sample matrix was reoriented to have samples arranged as rows for easy operation with known Machine learning libraries. The dataset was split into train and validation and the built pipeline was trained against just the split train set. No samples from the validation set were exposed to the model before the validation phase. The trained models can be broadly classified into two different types - Machine Learning Models and Transfer Learning models. Machine Learning models comprised of Simple Classification algorithms like K Nearest Neighbours, Fishers' LDA, and Ensemble models like Random Forest and Xtreme Gradient Boosting. Connectionism approaches were also tried out. Artificial Neural Networks with single and two hidden layers were also attempted against the input sample data. Deep Learning models attempted were variations of the Convolution Neural Network (CNN). Since training a CNN from scratch would take a long time, Transfer Learning using pre-trained architectures and parameters was used as lower layers of the CNN architecture. The handwritten digit classification resembles the MNIST, handwritten digit classification, the knowledge of the performance of different CNN architecture against MNIST was used to choose the pre-trained architecture. Different variations of Res Net were trained and their performance against unseen validation set was used to further narrow down the type of ResNet - ResNet 101. These speeds up the training process as the number of parameters that need to be learned is significantly reduced from millions to thousands or hundreds. All layers of ResNet pre-trained models were frozen and an output layer of 11 units was added with the softmax activation function to predict the probability of each of the output classes (10 symbols and one for outliers).

## VI. Model Experimentation

The input resolution of 300 * 300 posed a requirement for a massive space of main memory and thus was resized to 100 * 100 [8]. Also, In order to assert the performance of the model at various phases of the classification process, the input samples were split into Train, Test, and Val samples in the ratio 80: 20: 10 [9]. Though, the input samples were almost uniformly distributed with 800 – 900 samples per class, the outliers were just 234. So, the Train, Test, and Val split was stratified using target labels, so that each set receives a sample distribution of symbols.

Each sample with a simple written text occupies only a portion. Thus, it is efficient to extract the most contribution pixel positions to predict the symbol. It can be achieved through Principal Component Analysis where we extract the principal components that explain the most variants. The number of Components that preserve 85 and 90 percent are 72 and 150 respectively. The images were subjected to pipelines built with models from different strata of the Machine Learning world. Since some models that account for distances were attempted, the flattened input samples were standardized as part of the first step in the pipeline. This report briefs about the experiments that resulted in considerable performance metrics (accuracy). Hyper-parameters play a crucial role in the effectiveness of a model. One of the most common procedures of Hyperparameter tuning GridSearchCV (Grid Search Cross Validation) was used to tune the hyperparameters for models. The best estimator (one with the best hyperparameters) chosen was trained against the samples excluding the validation data set and the time taken for this fit is also captured for comparison later.

Random Forest Classifier (RFC) was attempted with 72 and 150 PCA components and with 400, 800, and 1200 RFC Trees. Assumptions that (higher) 150 PCA components would overfit are verified with the best RFC estimator corresponding to 72 PCA components and 1200 estimators. This result was bolstered by the experiment made with (E)Xtreme Gradient Boosting (XGBoost). XGBoost was attempted with 72 and 150 PCA components and 400, 800, and 1200 estimators. The best estimator used 72 components and 1200 estimators. Simpler ML models were attempted with varied hyper-parameters but did not yield considerable accuracy. Connectionism approaches were explored with Multi-Layer Perceptron. Two MLP modes one single-layered with 40 neurons in the hidden layer and the other with two hidden layers (40, 10) were attempted with PCA components as input. 150 PCA components when pumped to single layered MLP network performed better while the two-layered memorized the train samples leading to poor validation performance. The time taken for the single-layered MLP was also captured.

For experimenting with Res Net-based Transfer learning models, the input image samples in Train, Test, and Validation set were first converted to tensors using a custom data loader and were passed to a diverse array of Residual Neural Network models labeled ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 optimized with a Gradient descent optimizer and a scheduler outfitted with a consistent learning rate of 0.001 and significant step size and batch size and epoch count of 7 and 4 and 26 respectively. Finally, the models were evaluated in terms of reliability and temporal complexity using the given validation dataset.

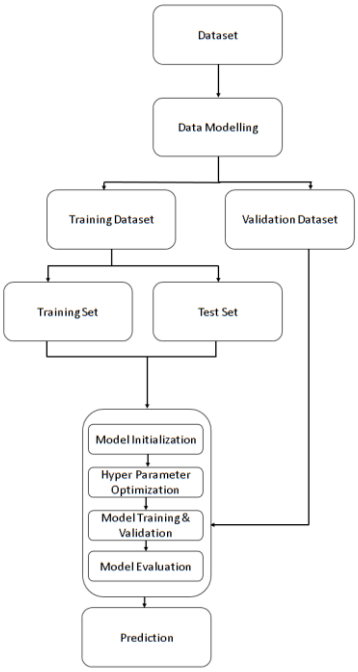


Fig. 1: Architecture of the proposed algorithm

## VII. Model Evaluation

The proposed phases were primarily evaluated using classification metrics. The metrics utilized for evaluating the proposed phases are given by Classification Accuracy Score [10] respectively. The metrics were computed using equation (1).

$$Classification\ Accuracy\ Score = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{-(1)}$$

Furthermore, The accuracy score for the model was obtained by substituting the following components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) into equation (1).

## VIII. Model Results & Discussions

As discussed earlier the proposed frameworks were evaluated using the aforementioned classification metrics.

### A. *Machine Learning Analysis*

This phase compares the performance of a custom classification framework comprising of dimensionality reduction technique titled PCA appended with three distinct classification algorithms individually, random forest, xgboost, and multi-layer perceptron classification optimized using the strategy of grid search cross-validation. Furthermore, the performance of the proposed models was realized through the classification metrics depicted earlier. The results are presented as training and validation outcomes as depicted in Fig 2 & 3.

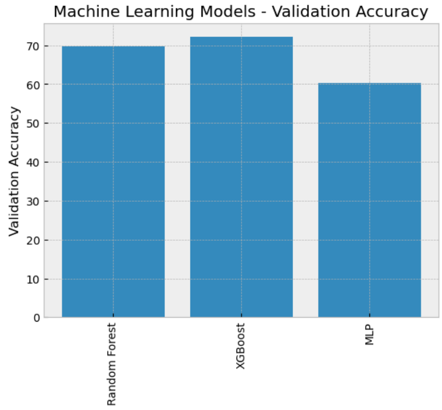


Fig. 2: Machine learning models Accuracy

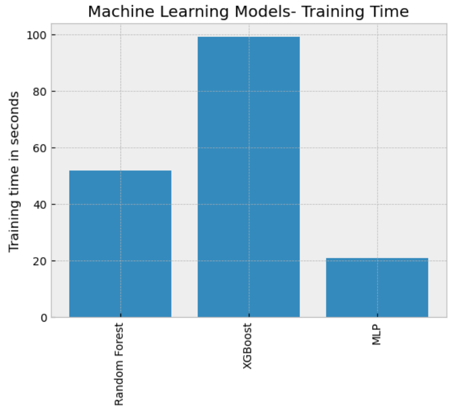


Fig. 3. Machine Learning models Time Complexity Graph

### B. *Transfer Learning Analysis*

This phase examines the performance of the proposed novel transfer learning frameworks. In addition, the performance of the suggested models was determined using classification metric titled accuracy score. The results are presented as training, validation and time complexity outcomes respectively as depicted in Fig 4 & 5.

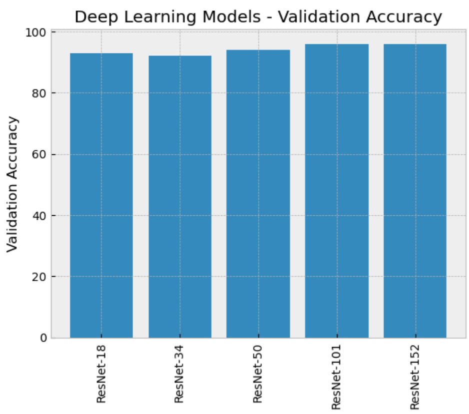


Fig. 4: ResNet Accuracy Graph

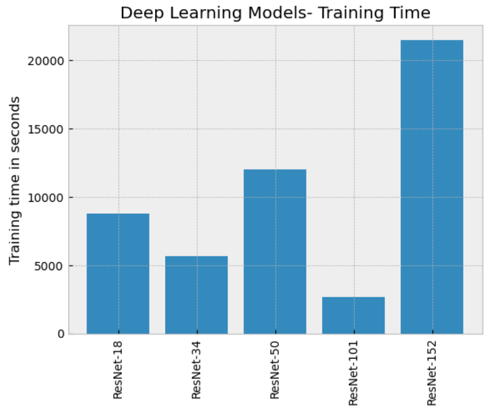


Fig. 5: ResNet Time Complexity Graph

Based on the results presented in the table above, it can be concluded that the transfer learning technique titled ResNet 101 outperforms other transfer learning techniques in terms of accuracy and time complexity. As a consequence, the designated transfer learning method ResNet 101 was employed to analyze the presented dataset. The results are presented as training and validation outcomes as depicted in Fig 6 & 7.

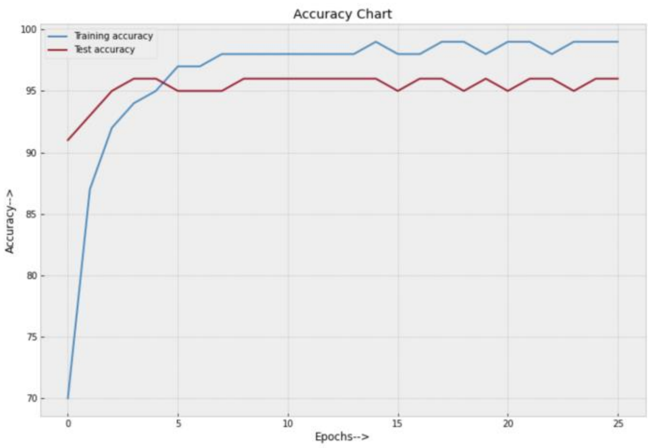


Fig. 6: Resnet 101 Training and Validation accuracy plot

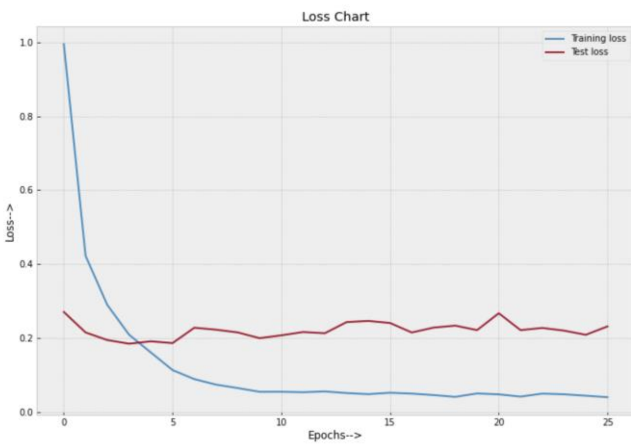


Fig. 7: Resnet 101 Training and Validation loss plot.

## IX. CONCLUSION

This research illustrates the significance of recognizing diverse mathematical symbols using conventional machine learning techniques. Furthermore, the study also indicates that the proposed ensemble models exhibit the phenomenon of overfitting, in which the models attempt to memorize the training samples as a result they underperform on validation data. As a result, it can be concluded that transfer learning techniques are quite compatible with the given dataset in terms of performance and temporal complexity. Finally, based on the results generated among the transfer learning approaches, it can be concluded that the transfer learning model titled ResNet 101 strikes a balance between performance and the temporal complexity thereby making it the best model among all the machine learning models evaluated in this study.

In the future, further innovative classification frameworks could be tested on the given sparse dataset using a similar model architecture.

## REFERENCES


[1] Zhou, Zhi-Hua. "Open-environment machine learning." National Science Review 9, no. 8 (2022): nwac123.

[2] Campagner, Andrea, Davide Ciucci, and Federico Cabitza. "Aggregation models in ensemble learning: A large-scale comparison." Information Fusion 90 (2023): 241-252.

[3] Meshram, Sarita Gajbhiye, Chandrashekhar Meshram, Fateme Akhoni Pourhosseini, Mohd Abul Hasan, and Saiful Islam. "A multi-layer Perceptron (MLP)-fire Fly algorithm (FFA)-based model for sediment prediction." Soft Computing 26, no. 2 (2022): 911-920.

[4] Weiss, K., Khoshgoftaar, T.M., Wang, D. (2016). Transfer Learning Techniques. In: Big Data Technologies and Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-44550-2_3

[5] P. Ghosal, L. Nandanwar, S. Kanchan, A. Bhadra, J. Chakraborty and D. Nandi, "Brain Tumor Classification Using ResNet-101 Based Squeeze and Excitation Deep Neural Network," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), 2019, pp. 1-6, doi: 10.1109/ICACCP.2019.8882973.

[6] Muhammad 'Arif Mohamad, Hassan, H., Nasien, D., & Haron, H.(2015). A review on feature extraction and feature selection for handwritten character recognition. International Journal of AdvancedComputer Science and Applications, 6(2) doi:https://doi.org/10.14569/IJACSA.2015.060230

[7] https://www.kaggle.com/datasets/guru001/hasyv2

[8] Katharopoulos, Angelos, and François Fleuret. "Processing megapixel images with deep attention-sampling models." In International Conference on Machine Learning, pp. 3282-3291. PMLR, 2019.

[9] Kaur, Navdeep, and Ajay Mittal. "CheXVGG: A deep neural network for multi-disease classification of chest-xray images using transfer learning." In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1289-1293. IEEE, 2022.

[10] Pravin, S.C., Palanivelan, M: Acousto-Prosodic Delineation and Classification of Speech Disfluencies in Bilingual Children. In: , et al. Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). SoCPaR 2020. Advances in Intelligent Systems and Computing, vol 1383, 2021, Springer, Cham.