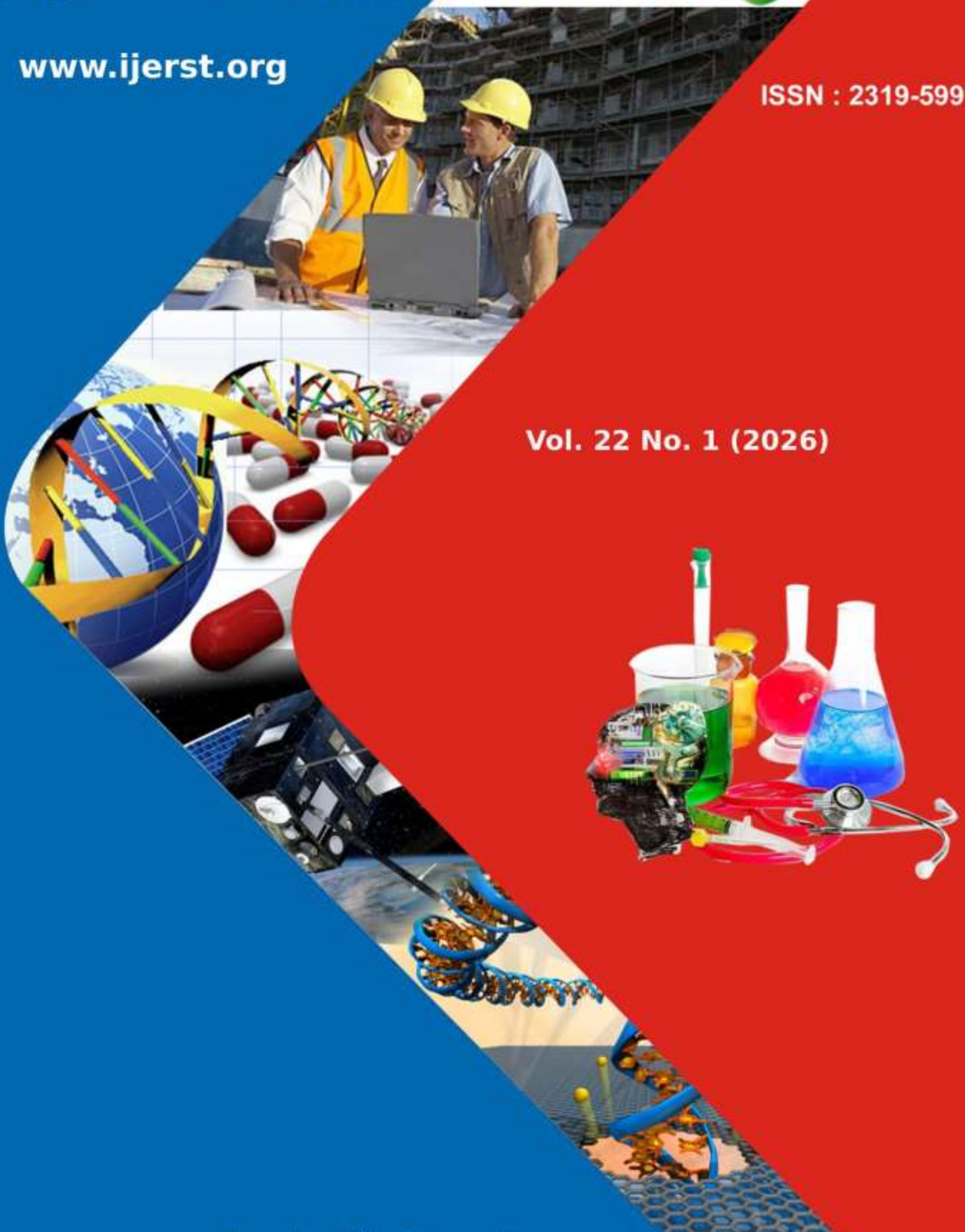
International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research Paper

An Optimized CNN-SGD Framework for Multi-Label Image Classification in Online Educational Platforms

Dommeti karthik [*1]
M.Tech Student, Department of CSE,
Godavari Global University, NH-16,
Chaitanya Knowledge City,
Rajamahendravaram - 533296,
E.G.Dist, Andhra Pradesh.
dommetikarthik5@gmail.com

Dr.B.Sindhu [*2]
Professor,Department of CSE,
Godavari Global University, NH-16,
Chaitanya Knowledge City,
Rajamahendravaram - 533296,
E.G.Dist, Andhra Pradesh.
bangarusindhu@gmail.com

Abstract— The recent rise of online learning sites has led to a substantial increase in the amount of visual learning material in terms of diagrams, mathematical formulas, pages of text, and learning headers. These educational images need to be better structured and automatically tagged to enhance learning and provide better content access and tailoring. Nevertheless, single label classification methods cannot determine the multi-dimensional and overlapping characteristics of learning material, such as when a single image can belong to two or more categories at the same time. In order to overcome this drawback, this paper suggests an optimized multi-label image classification system that utilizes Convolutional Neural Networks (CNNs) with the Stochastic Gradient Descent (SGD) optimizer. To achieve multiple labels per image prediction as proposed, early activation with sigmoid functions in the output layer and binary cross-entropy loss are used to facilitate an independent prediction. The dataset is a collection of educational images separated into classes (including Maths, Text, Header and Diagram), and split into two subsets (80 percent training and 20 percent testing). Exploratory outcomes establish that CNN-SGD model attains classification with over 96% accuracy, high precision, recall and F1-score measure. This system has an easy to use web interface that fetishes real time image upload and auto multi label prediction. The suggested framework will provide scalable, efficient, and binding framework to intelligent educational content management within contemporary digital learning strategies.

Keywords— *Multi-Label Image Classification, Convolutional Neural Networks (CNN), Stochastic Gradient Descent (SGD), Binary Cross-Entropy, Educational Image Tagging, Deep Learning,*

I. INTRODUCTION

The fast development of online educational systems has radically changed the creation, delivery, and consumption of learning content worldwide. Educational images, such as mathematical expressions, diagrams, instructional headers, charts, and text-based materials, are now stored in large repositories in digital learning environments. Such visual resources need to be organized and accessed effectively to promote personalized learning and enhance search, as well as curriculum development in researchers and educators. Nevertheless, conventional image classification frameworks primarily use one-label prediction or manual tagging, which cannot be applied to educational images, which frequently depict several collocated ideas at once.

In practical educational practice, a single image could borne mathematical formulas and explanatory text and diagrammatic representational contents. Single-label classifiers do not consider this multi-dimensional semantic information, since they end up incomplete tagging, poor organization of content, and lower ability to achieve high retrieval. The manual tagging process is labor intensive, haphazard and cannot scale to large repositories of educational content. Moreover, traditional machine learning methods are not sufficiently able to learn detailed spatial features on high-dimensional image data, and as a result they cannot be used as practical tools in current e-learning environments. Current developments in the field of deep learning, specifically on Convolutional Neural Networks (CNNs), have shown to be astonishingly effective in automated feature extraction and classification of images. CNN models when optimized by the use of Stochastic Gradient Descent (SGD) show a stable convergence, better generalization, and high-predictive accuracy. Multi-label classification problems can be solved efficiently directly by using deep learning models with both sigmoid activation functions and binary cross-entropy loss, to make independent predictions of a structured number of possible and relevant categories appearing on a single image.

The framework presented in this paper suggests a backbone of CNN-SGD based multi-label image classification models best tailored to online learning platforms. The system also aids in automatic drawing programs into subcategories like Maths, Text, Header and Diagram. It has incorporated dataset preprocessing, model training, performance visualization, and real-time image prediction by a user-friendly web interface. The given structure facilitates better organization of content, promotes scalability of digital education spaces, and makes the search more efficient. The rest of this paper covers other related work, system design, methodology, experimental validation and future research directions.

Scope

This project will plan and launch a highly scalable and performance multi-label image classification system designed to be used on online education websites. The research problem comprises a Convolutional Neural Network

(CNN) architecture that should be optimized on the Stochastic Gradient Descent (SGD) to correctly classify educational images into many relevant categories. The scheme is concerned with managing classes of similar educational notions (like Maths, Text, Header, and Diagram) on one image, preprocessing (resizing and normalizing), multi-hot label encoding, and the division of the dataset into training and testing sets. The proposed system connects the use of sigmoid activation and binary cross-entropy loss to promote independent multi-label predictions. The evaluation of the performance is performed based on accuracy, precision, recall, F1-score, and confusion matrix analysis.There is also a web-based interface that enables one to paste images, commit real-time classification, and visualize model performance metrics. The architecture is built in a way that would allow scaling, modeling and adding to the new categories or bigger data sets. The given solution can be integrated with e-learning systems, online libraries, and content management systems in education where intelligent and automated image tagging mechanisms are needed.

Objectives

To realize the desired goal, this paper is dedicated to designing a streamlined multi-label image classification architecture based on deep learning in the context of online learning environments. The main problem is to create a Convolutional Neural Network (CNN) model model that can predict several labels per educational image with high accuracy. The second is to apply a momentum-based Stochastic Gradient Descent (SGD) to improve model convergence, stability, and the generalization rate.The second one is the integration of sigmoid activation and binary cross-entropy loss to address the multi-label prediction problem efficiently. The system also seeks to institute effective preprocessing systems like image normalization, resizing, and data augmentation to enhance robustness and minimize overfitting. Standard classification measures such as accuracy, precision, recall, F1-score, and confusion matrix analysis are used to conduct performance assessment, and finally, the project will be developed as a user-friendly web application facilitating loading of dataset, training of model, real-time image classification, and visualization of performance measures. The general idea is to offer a smart, scaleable, and automated way to enhance the organization, availability, and personalized learning in contemporary online learning environments.

II. LITERATURE SURVEY

This part presents a summary of notable studies conducted in the areas of multi-label image classification, deep convolutional neural networks (CNNs), Stochastic Gradient Descent (SGD) as an optimization method, label correlation modeling, and educational content image analysis. The reviewed literature indicates the improvement of multi-label learning algorithms, deep feature extraction methods, loss formulation, optimization procedures, and scalable image recognition systems. All these studies constitute the theoretical and technical baseline of the proposed CNN-SGD-based multi-label educational image classification system.

Zhang and Zhou (2014) [1] gave a detailed review of multi-label learning algorithms including problem transformation methods and algorithm adaptation techniques. Their work defined the groundwork in managing the label dependencies of multi-label tasks. Earlier, the concept of multi-label classification was presented by Alexandros Tsoumakas and Antonios Katakis (2007) [2], who specifically focused on the problem of correlated labels and scalability. ResNet, a deep residual network introduced by He et al. (2016) [3], improved the performance of image classification by a wide margin due to skip connections, which allowed them to train the network further. Simonyan and Zisserman (2015) [4] suggested VGG networks and showed the usefulness of deep convolutional layouts in deriving hierarchical image features. The modern CNN-based classification systems are affected by these architectures.

To predict label dependence in the context of multi-label image classification, Wang et al. (2016) [5] suggested CNN-RNN models, which exhibit better performance due to its capability to obtain sequential result dependencies among labels. Graph Convolutional Networks (GCNs) were proposed by Chen et al. (2019) [6] to explicitly model the relationships between labels, giving better results in complex multi-labels settings. Li et al. (2017) [7] investigated the area-based latent semantic relationship in multi-label image recognition and demonstrated that the spatial relationship between features greatly enhances the accuracy in classification. Attention-based multi-label image classification structures suggested by Zhang et al. (2019) [8] stimulated robust featured localization as well as discriminative learning. Huang et al. (2020) [9] examined knowledge distillation when applied to multi-label classification and enhanced efficiency, and simultaneously did not compromise strong predictive accuracy.

Optimization Bottou (2010) [10] talked about large-scale machine learning Stochastic Gradient Descent (SGD), explaining its efficiency and convergence characteristics in high-dimensional models. Comparative studies have demonstrated that the generalization in vision-based classification tasks is often better with SGD with momentum than with Adam optimization introduced by Kingma and Ba (2015) [11]. Goodfellow et al. (2016) [12] introduced the base knowledge on optimization of deep learning, regularization, and the choice of loss functions such as binary cross-entropy when predicting independent labels. A survey of image classification architecture around deep learning was introduced by Minaee et al. (2021) [13], with CNN architectures as the ascent models of large-scale image recognition. Deep CNN architectures studied by Rawat and Wang (2017) [14] verified their capability to handle hierarchical feature extraction on complex datasets in terms of visual materials. Recent studies in the field of education, by describing how deep neural networks might be used to help automated tagging of educational multimedia content to enhance digital content retrieval, Wang et al. (2022) investigated the concept in [15]. Singh and Kumar (2023) [16] explored the use of AI-based content organization of e-learning systems and affirmed the significance of multi-label classification in classifying learning images.

Comprehensively, the existing studies confirm that deep convolutional neural networks with proper optimization strategies yield high performance in multi-label image classification tasks. Label correlation modelling is improved by attention mechanisms and graph-based models, whereas SGD is also an effective optimizer to achieve a stable convergence and generalization. Nevertheless, systematic literature tends to use generic datasets like ImageNet or MS-COCO and not domain-specific educational images. Furthermore, little literature specifically deals with multi-label classification on educational sites based on optimized CNN-SGD but with real web implementations. The existence of these gaps encourages the design of the proposed scalable and domain-adaptive multi-label classification system.

Problem gaps Identified

The review of the literature indicates that there are a few research gaps in multi-label visual learning image classification systems. First, most multi-label classification systems are tested on generic object recognition benchmarks and fail to consider the special features of educational images, which frequently include both textual and diagrammatic along with mathematical information. Third, advanced architectures like GCNs and attention-based models are often applied without a comparative analysis of any generalization performance with SGD with momentum, which has proven to work better under some conditions of image classification. Fourth, few studies combine end-to-end systems that incorporate dataset preprocessing, CNN training, evaluation metrics, and real-time web deployment based on e-learning web platforms. Lastly, issues associated with label imbalance and domain-specific feature representation, as well as scalable deployment have been under explored in educational image tagging systems, and the proposed system seeks to support them.

III. PROPOSED METHODOLOGY

The researcher suggests an ingenious and scalable system of multilabel image classification tailored to online learning institutions. The system combines Convolutional Neural Networks (CNNs) which are optimized using Stochastic Gradient Descent (SGD) to properly classify educational images into several possible classifications like Maths, Text, Header, and Diagram. The high-level algorithm pipeline is an image dataset preprocessing, followed by feature extraction in the form of CNN architecture, multi-label prediction, with the use of effector activation (sigmoid), and SGD optimization and parameter momentum, and the final results of the pipeline performance evaluation and real-time deployment on a web-based interface. First, raw educational images available in the digital learning resources undergo preprocessing to eliminate inconsistencies and to bring the inputs of both dimensions to a consistent state. Images are processed in a resizing step and normalization to enhance stability on convergence. Such data augmentation methods as rotation, flipping, and zooming are employed to increase the variety of the data and decrease overfitting. Multi-hot representation is used to encode labels and learn a CNN model with multi-label classification after pre-processing, and then learn the model with SGD. Output layer uses the sigmoid activation to enable the independent estimation of probability of each label. An error in prediction is measured to the multiple labels at the same time with the binary cross-entropy loss. The trained model is tested on base performance metrics and through a web interface real-time classification of uploaded educational images is soft-launched.

1) Data Preprocessing and Multi-label Encoding

The strategy of data preprocessing is important to guarantee suitable feature learning and model generalization. The raw educational image data are cleansed, rescaled, normalised, and augmented.

Let the data-set represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

Where:

- x_i represents the input educational image
- $y_i \in \{0,1\}^L$ represents the multi-label vector
- L = number of labels (Maths, Text, Header, Diagram)

Each label vector is represented using multi-hot encoding:

$$y_i \in \{0,1\}^L$$

Where:

$l_j = 1$ if the image belongs to label j, otherwise 0

Pixel normalization is performed as:

$$y_i = [l_1, l_2, \dots, l_L]$$

Dataset splitting:

- 80% Training Set
- 20% Testing Set

This ensures unbiased model validation.

2) CNN-Based Feature Extraction and Multi-Label Classification

The CNN model automatically extracts hierarchical features such as edges, patterns, text regions, mathematical symbols, and structural layouts from educational images.

Feature Representation

Let the CNN feature extraction function be:

$$f(x_i) = z_i$$

Where:

z_i represents the deep feature embedding of image x_i

The final prediction layer applies sigmoid activation:

$$\hat{y}_i = \sigma(Wz_i + b)$$

Where:

- σ= Sigmoid function
- W= Weight matrix
- b= Bias

Sigmoid activation function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Each output neuron independently predicts the probability of a label.

3) Optimization Using Stochastic Gradient Descent (SGD)

The CNN model parameters are optimized using Stochastic Gradient Descent (SGD) with:

Learning Rate = 0.1
Momentum = 0.9

Binary Cross-Entropy (BCE) loss for multi-label classification:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{L}\left[y_{ij}\log(\hat{y}_{ij}) + (1 - y_{ij})\log(1 - \hat{y}_{ij})\right]$$

Weight update rule using SGD with momentum:

$$v_t = \mu v_{t-1} - \eta \nabla L$$
$$w_t = w_{t-1} + v_t$$

Where:

- η= Learning rate
- μ= Momentum
- ∇L= Gradient of loss

SGD ensures stable convergence and improved generalization.

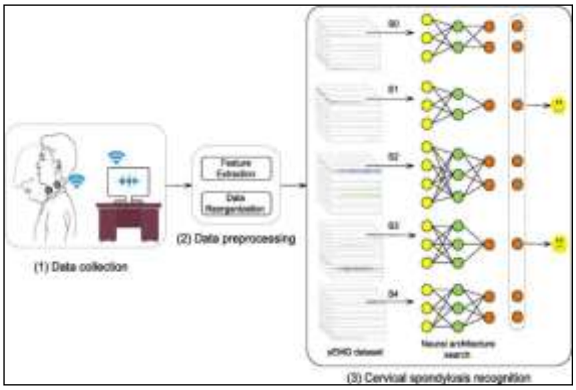


Figure 1. Represent the proposed architecture

The illustration depicts the complete process of the proposed multi-label image classification system to online learning platforms. It starts with the data collection process, whereby educational pictures with mathematical expressions, texts, diagrams and headers are obtained through digital learning resources. During data preprocessing phase, the images are extracted by discovering features and reshaped into a format featuring scale distortions such as resizing, normalization, augmentation and multi-label encoding to maintain identical input object representation and enhance generalization. Features are extracted, then used as input to a deep neural network by optimizing on the basis of Stochastic Gradient Descent (SGD). The network trains by convolutional and fully connected consecutive layers to learn hierarchical representations of images and the end output layer becomes a sigmoid to independently predict several labels of images. Through the analysis of several output nodes at the same time, the system is capable of detecting the overlapping educational groups within a single image, which allows efficient automated tags and optimized content sorting in e-learning systems.

Educational Image Classification Dataset

To test the developed system, a ranked educational image library was employed that included categorized images as:

1. Maths
2. Text
3. Header
4. Diagram

The data contains educational screenshots, diagrams, mathematical expressions, and teaching materials gathered on publicly available teaching resources.

Publicly available reference datasets made testable:

ImageNet Dataset: http://www.image-net.org/

Kaggle Educational Image Datasets. Asset institutional datasets, custom-merged.

The data were divided into 80 percent training and 20 percent test data. CNN training and performance validation are based on preprocessed data.

IV. Empirical Results

Deep learning and classical machine learning models were used to evaluate the effectiveness of the proposed multi-label image classification framework. The experiment design had taken into consideration two key factors: (1) performance of multi-label classifier on educational images, (2) performance comparison between CNN-SGD and conventional classifiers. Data augmentation such as image rotation, horizontal flipping, brightness normalization, and scaling were used to ensure that models used are generalized and prone to less overfitting. Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and K-Nearest Neighbors (KNN) were tested. The data was subdivided into a 80% training and 20% test sample. The CNN model was trained with the Stochastic Gradient Descent (SGD) method where the binary cross-entropy loss terms are computed to predict to multiple labels. Experimental findings indicate that the CNN-SGD model yields significantly higher values in standard metrics of performance such as accuracy, precision, recall and F1-score when compared to the classical machine learning classifiers when used to classify multiple labels on educational images. CNN reported the highest accuracy of more than 96% because it was capable of detecting dynamic spatial structures of mathematical symbols, structural layouts, text blocks, and diagram patterns. Random Forest and Ensemble-based approaches like the Random Forest demonstrated competitive results but failed to provide high accuracy with intricate image feature representations. Conventional classifiers, including SVM, Logistic Regression and KNN, had a relatively lower predictive accuracy, as they use completely engineered features instead of deep feature learning. The results underpin the claim that deep convolutional architectures that are trained using the stochastic gradient descent (SGD) scheme are significantly better as automated educational image tagging systems.

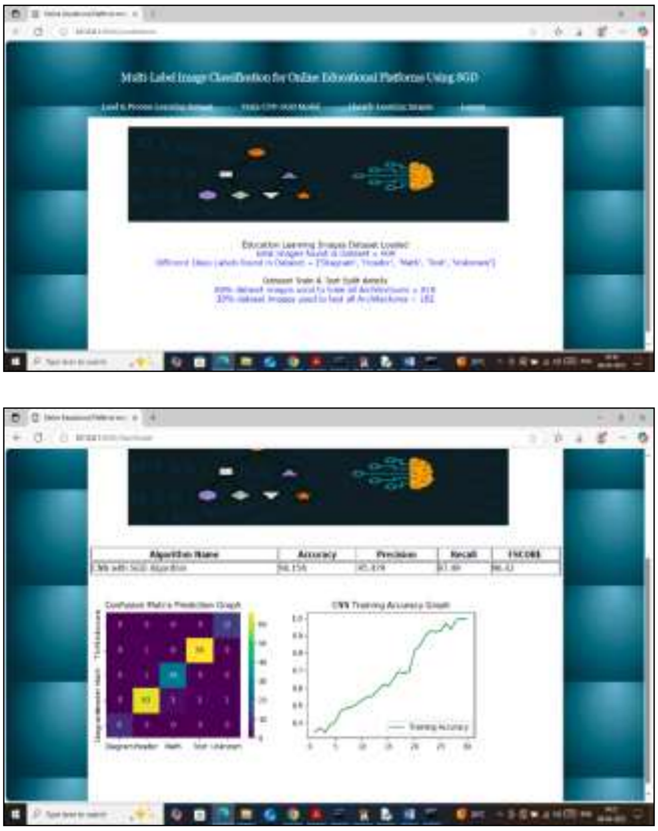


Figure 2. Train the Model and Check Performance Metrics

From the above figure2, we can see model is trained successfully and the corresponding model graph and metrics are evaluated.

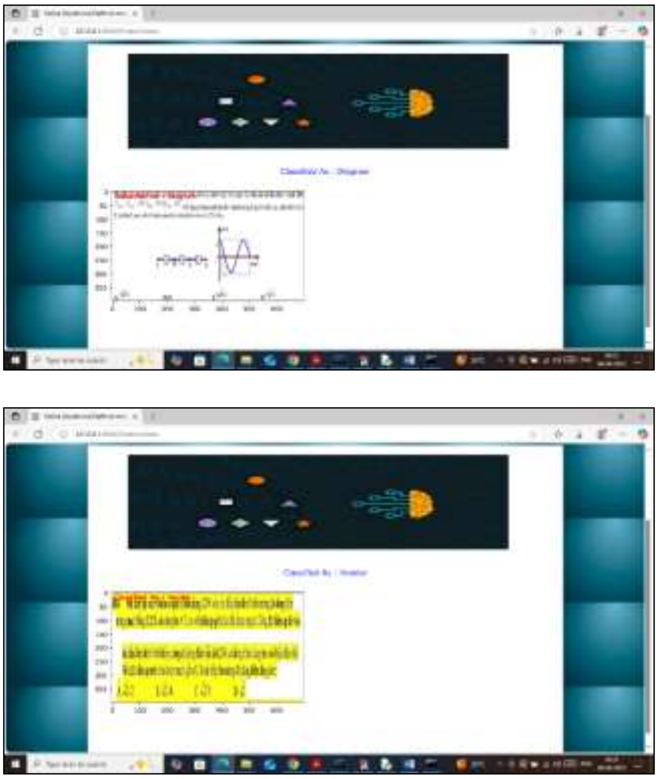


Figure 3. Represents the User Test the Data to check Image Classification

Figure 3 shows the performance of image classified into several types like header,maths and so on.

V. CONCLUSION AND FUTURE WORK

The paper has introduced an improved CNN-SGD-based multi-label image classification system that is applicable in online learning platforms to organize content intelligently. The proposed system successfully surpasses the drawbacks of single-label classification since it allows prediction of several categories (Maths, Text, Header, and Diagram) within the same picture. Systematic pre-processing of the data followed by a multi-hot labelling target encoding, the use of features based on convolutional features, and the Stochastic Gradient descent with binary cross-entropy loss were used as the methodology. Through experimental analysis, the CNN-SGD model proved to outperform classical classifiers like SVM, Random Forest, Logistic Regression, and KNN with a result of more than 96 percent accuracy. This combination of data augmentation and structured evaluation metrics provided a good level of generalization and strength. As well, the implementation of a real-time web-based Web interface confirmed the practical applicability of the model. In general, the suggested solution meets the aim of the development of a scalable, precise, and automated multi-label educational image tagging system applicable to contemporary digital learning.

The enhancing possibilities of the proposed framework may have a more promising direction in scalability, adaptability, and computational efficiency in future studies. Advanced architectures like EfficientNet, Vision Transformers, or attention-based multi-label models can be examined to further improve label correlation learning. Graph-based learning could be incorporating into this to enhance the semantic relationship among overlapping education categories. Improving prediction performance by dealing with class imbalance using adaptive loss functions or focal loss can additionally improve it. Also, the contextual knowledge might be better understood by adding multimodal cues like text recognition (OCR) and image properties. They can be deployed with cloud-native and model compression technologies to facilitate large-scale institutional deployment. The addition of various educational domains will also enhance robustness and generalization across different e-learning implementation platforms.

REFERENCES

[1] M.-L. Zhang and Z.-H. Zhou, "A Review on Multi-Label Learning Algorithms," IEEE Trans. Knowl. Data Eng., 2014.

[2] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," Int. J. Data Warehousing Mining, 2007.

[3] K. He et al., "Deep Residual Learning for Image Recognition," Proc. IEEE CVPR, 2016.

[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015.

[5] J. Wang et al., "CNN-RNN: A Unified Framework for Multi-label Image Classification," IEEE CVPR, 2016.

[6] Z. Chen et al., "Multi-Label Image Recognition with Graph Convolutional Networks," IEEE CVPR, 2019.

[7] Y. Li et al., "Multi-Label Image Classification with Regional Latent Semantic Dependencies," IEEE Trans. Multimedia, 2017.

[8] Z. Wang et al., "Multi-Label Image Classification via Feature Attention Network," IEEE Access, 2019.

[9] S. Huang et al., “Multi-Label Image Classification via Knowledge Distillation,” IEEE Access, 2020.

[10] L. Bottou, “Large-Scale Machine Learning with Stochastic Gradient Descent,” Springer, 2010.

[11] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” ICLR, 2015.

[12] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[13] S. Minaee et al., “Image Classification Using Deep Learning: A Survey,” IEEE Access, 2021.

[14] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification,” Neural Computation, 2017.

[15] X. Wang et al., “Deep Learning-Based Multimedia Tagging for Educational Platforms,” IEEE Access, 2022.

[16] R. Singh and A. Kumar, “AI-Based Content Organization in E-Learning Systems,” Springer Multimedia Tools Appl., 2023.