Microsystem Technologies
https://doi.org/10.1007/s00542-024-05741-9

TECHNICAL PAPER

Vision transformer-based model for early detection of dysgraphia among school students

Prateek Sharma[1] · Basant Agarwal[2] · Gyan Singh Yadav[1] · Sonal Jain[3]

Received: 29 December 2023 / Accepted: 5 August 2024

Abstract

Learning disorders, an umbrella term for a range of learning difficulties, impair a person's capacity to learn new skills. Dysgraphia is one of the prevalent learning disorders among children all over the globe. It is defined as a child's functional restriction in establishing accurate letter or word construction, inadequate speed, and readability of written text. Lack of availability of experts who can diagnose and high diagnostic cost, makes it important to discover a diagnostic approach for dysgraphia that is accurate, accessible and simple to use. Analyzing handwriting is the most common technique to detect dysgraphia which can be automated through image processing techniques. The use of deep learning algorithms has become increasingly widespread in image processing over the course of the last few decades and analysis. However, the effective classification of these handwritten images presents a number of challenges like low accuracy, inadequate availability of labelled data for training purposes. Considering the notable efficacy demonstrated by the Vision Transformer (ViT) in image classification, we proposed a ViT-based classification model in this paper. This model splits handwriting images into patches and then process those through a transformer. Just like word embedding, these input image patches are passed in a sequence to the transformer. To compare performance of proposed model, we also applied transfer learning techniques VGG16, VGG19, ResNet50 and InceptionV3. After comparing the results, it was found that Vision Transformers are best suitable for the classification. Vision transformer has outperformed with macro average F1 score value of 0.92 for the classification. Out of all pretrained models VGG16 performed best with the macro average F1 score value of 0.90. The findings of this study indicate that ViT-based model has the potential to assist experts in the early detection of dysgraphia.

1 Introduction

Learning problems, especially dyslexia and dysgraphia, are fairly common in classrooms all throughout the world. Learning disorders or impairments, a blanket term for a variety of learning challenges, impair a person's capacity to learn new skills. The main definition of dysgraphia is a problem of written expression. In addition to the handwriting elements, it can also have an impact on the spelling, grammar, organization, etc. (Chung et al. 2020). Just like other learning disabilities dysgraphia is also caused by neurological problem. If dysgraphia develops in childhood, the cause is primarily a problem with orthographic coding (Devi and Kavya 2023). Orthographic coding basically enables you to recall written words and the movements required to write them.

The diagnosis of dysgraphia typically involves a multidisciplinary team comprising medical professionals, and mental health specialists, such as licensed psychologists or other professionals with expertise in addressing learning difficulties (Agarwal et al. 2023a). Because dysgraphia predominantly affects handwriting, the majority of traditional diagnostic methods involve analyzing handwriting and identifying particular patterns. An occupational therapist or a certified psychologist assesses the student's

✉ Prateek Sharma
127.0.0.17@gmail.com

Basant Agarwal
basant@curaj.ac.in

Gyan Singh Yadav
gyansingh.cse@iiitkota.ac.in

Sonal Jain
sonaljain@spuvvn.edu

[1] Indian Institute of Information Technology, Kota, Rajasthan, India

[2] Central University of Rajasthan, Kishangarh, Rajasthan, India

[3] PG Department of Computer Science and Technology, Sardar Patel University, Vallabh Vidyanagar, Anand, India

Published online: 12 August 2024

numerous capabilities, such as constructional abilities, working memory, writing and spelling skills, executive function, and so on, to identify the presence of dysgraphia. The "Beery Visual Motor Test of Integration – Sixth Edition (VMI6)" (Beery 2004) is one of the most well-known conventional tests used by professional psychologists to measure constructional competence. Executive function abilities can help in planning, focusing, recalling orders, and arranging many tasks. For writing, the individual must have all executive functions proper (Chung and Patel 2015). Rey Complex Figure Test (Meyers and Meyers 1995) and Behavior Rating Inventory of Executive Function (BRIEF) (Roth et al. 2013) are the tests that are most often used for evaluating executive function. It is also known, dysgraphia may arise when there is some level of difficulty with the working memory (Crouch and Jakubecy 2007). The two most common tests that are used to evaluate whether or not an individual's working memory has been impaired are Test of Memory and Learning – 2 (TOMAL-2) and the Wide Range Assessment of Memory and Learning – 2 (WRAML-2) (Hartman 2007). A popular manual handwriting analysis method is the Concise Evaluation Scale for Children's Handwriting (BHK), which is used to evaluate the speed and quality of the writing under the guidance of a physiotherapist or counselor. BHK tests are typically conducted in individual healthcare settings. It was first created to evaluate handwritten collections of second and third grade kids. BHK scores are now being utilized in research to build the benchmark of data needed to train and evaluate computer learning-based dysgraphia diagnosis systems. Primary dysgraphia evaluation technique is based on handwriting analysis. There are also some prominent automated dysgraphia screening solutions that are based on digital technologies. TestGraphia (Dimauro et al. 2020) is a software that has been developed for detecting dysgraphia in the initial phase. The well-known BHK test is the basis for TestGraphia. They made a software device for automated diagnostics by following the normal BHK method. There is also a tablet-based application called Play Draw Write (Dui et al. 2020) developed for checking pre literacy children's handwriting abilities. The suggested tablet software would detect the presence of dysgraphia indicators by quantifying three handwriting laws.

As was mentioned above, the majority of psychological diagnosis methods for dysgraphia are assessment oriented. Most of these are carried out in clinical settings, and the interpretation of the findings requires the expertise of a trained therapist or counselor. In addition, the whole procedure takes close to a year to finish and is quite costly both financially and time-wise (Agarwal et al. 2023b). It is quite unlikely that a person who comes from a financially struggling household would be able to go through a diagnostic procedure that is not only costly but also time demanding. Considering above-mentioned facts, it is of the utmost importance to discover a diagnostic approach for dysgraphia that is accurate, accessible and simple to use.

This research introduces an innovative method for the detection of dysgraphia at an early stage by a vision transformer (Dosovitskiy et al. 2020) based model, which offers a cost-effective solution for handwriting analysis. To compare our results with other studies we also performed classification using well-known convolutional neural networks (CNNs) such as VGG16, VGG19, ResNet50, and InceptionV3.

The primary findings and contributions of our research are outlined as follows:

- A dataset of handwriting images that have been labeled as normal and dysgraphic by trained professionals.
- The current literature on dysgraphia classification focuses on CNN-based models. We made a vision transformer-based model that could learn from a wide range of features and be more accurate.
- Four well-known CNN networks, VGG16, VGG19, ResNet50, and InceptionV3, were also deployed, and their results were compared to the results from the proposed vision transformer-based model.

The remaining part of the paper is arranged in the following manner: In Sect. 2, related work in relevant field is explained. We discussed the dataset and proposed methodology in Sect. 3. The specifics of the experiment are presented in Sect. 4. Section 5 offers a discussion and comparison of the results with other models. Section 6 concludes the paper by outlining its limitations and prospective research directions.

2 Related work

In the area of image processing, machine learning methods, specifically deep convolutional neural networks (CNN), have shown superior results (El-Gayar et al. 2020). Numerous researchers have effectively used machine learning methodologies to develop a dysgraphia detection system (Kunhoth et al. 2024). This section has examined recent advancements in machine learning models for detection of dysgraphia.

The pipeline of automated diagnosing dysgraphia is quite similar to other machine learning pipelines. It starts with data collection, pre-processing of images before feeding those into the pipeline, followed by feature extraction and relevant feature selection and then training the classifier model (Fig. 1).

Extraction of relevant features from the data is an essential step in the training process for machine learning

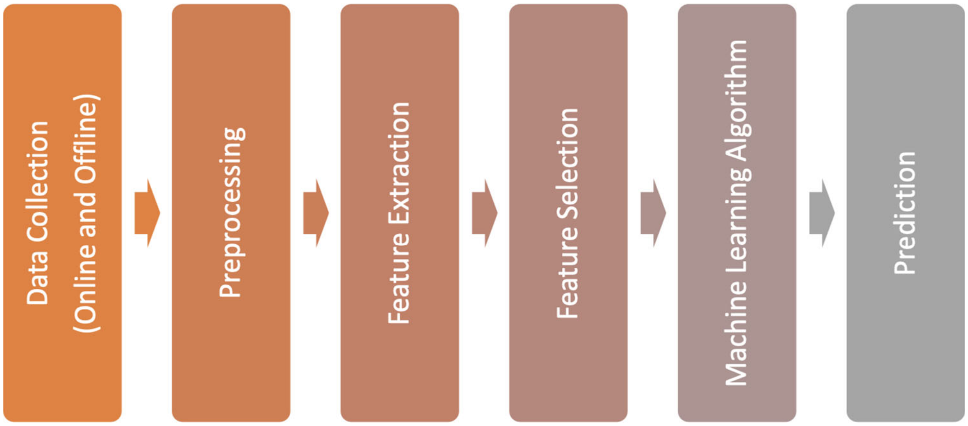


Fig. 1 Standard workflow of machine learning models for dysgraphia detection

models. Numerous features can be captured from the collected images of handwritten text in order to conduct in-depth research using the existing models. The image below illustrates some of the most significant handwriting features that can be used by image processing models for the detection of dysgraphia (Fig. 2).

All of these features are important to consider while screening for dysgraphia since children who have the condition are more likely to produce these errors when writing than children who do not have the condition (Table 1).

Yogarajah and Bhushan (2020) proposed an automated dysgraphia diagnosis method based on the handwriting images. Based on the graded difficulty level, fourteen Hindi words and three conjoined consonants were selected. For the purpose of this research, a total of 267 handwriting images were gathered, 164 coming from children diagnosed with dyslexia- dysgraphia and the remaining 103 coming from age-matched normal group. These pictures were scaled down to a uniform height of 113 pixels while maintaining a variety of widths that varied according to the aspect ratio of the original image. A Convolutional Neural Network (CNN) using Keras and Tensorflow was implemented which produces accuracy of (86.14 ± 1.02) %.

Mekyska et al. (2019) suggested a technique in which handwriting data is collected through a digital tablet. Experiment was conducted on a total of 54 students, consisting of 27 normal and 27 dysgraphic individuals. During the training process for the models, the Random Forest classifier and the Linear discriminant analysis technique were deployed. The results suggested that the Random Forest classifier is capable of classifying people as either dysgraphic or normal with a sensitivity and specificity of 96%.

Asselborn et al. (2018) proposed an approach where students were given an assignment to write on a paper, that was attached to a Wacom Intous tablet, for approximately five minutes. In this study a total of 298 participants, including 56 people who were dysgraphic, participated. Following the completion of the handwriting exercise, a total of 54 features of the handwriting were retrieved for further processing. The extracted feature belongs to three categories Static features, kinematic features and dynamic features. Random Forest classifiers was trained using the

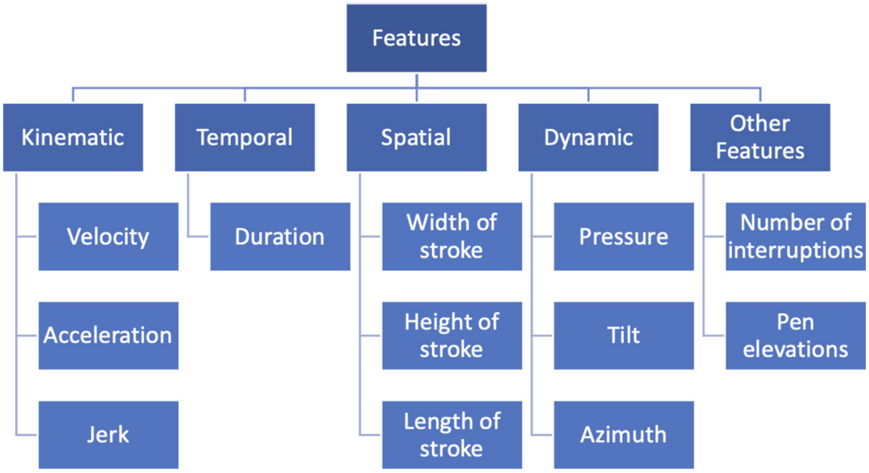


Fig. 2 Handwriting features relevant for dysgraphia detection

Table 1 Features Used By Different Studies Proposed In Literature

References	Features
Yogarajah and Bhushan (2020)	Spatial
Mekyska et al. (2019)	Kinetic, Temporal, Spatial, Dynamic, Other
Asselborn et al. (2018)	Kinetic, Temporal, Spatial, Dynamic
Drotar and Dobes (2020)	Kinetic, Temporal, Spatial, Dynamic, Other
Asselborn et al. (2020)	Kinetic, Temporal, Spatial, Dynamic
Devillaine et al. (2021)	Kinetic, Spatial, Dynamic
Deschamps et al. (2021)	Kinetic, Temporal, Spatial, Dynamic
Dankovicova et al. (2019)	Kinetic, Temporal, Dynamic
Rosenblum and Dror (2016)	Temporal, Spatial, Dynamic

extracted features. This approach yielded a 91% accuracy rate.

Drotar and Dobes (2020) proposed a machine learning based dysgraphia detection method based on an approach to collect a news data set and analyze it through ensemble learning models like Random Forest, Adaboost. 120 students from different schools took part in the research project. On the piece of paper that was affixed onto the display of the WACOM tablet, the participants were asked to write certain words and sentences through a normal pen. The WACOM tablet was able to record 5 distinct signals, including the movement of the pen in X or Y direction, pressure of the tip, as well the pen's azimuth and altitude when writing. Based on the findings, it seems that the AdaBoost algorithm was successful in attaining the greatest classification accuracy of close to 80%. It is also observed that pressure and pen lifts are the most important features.

In Asselborn et al. (2020), authors presented a methodology to measure various scales of difficulties related to handwriting. In this study 448 children participated. Participants were instructed to write down 5 sentences taken from the BHK test on an iPad device. For collecting handwriting raw data Dynamico software was used. Total 63 features were extracted from data collected by dynamic. These features were categorized under kinematic and dynamic categories. To project 63 features in 3D space principal component analysis (PCA) was used. Proposed study employed K-Mean unsupervised learning.

In Devillaine et al. (2021), graph motor tests were used to gather raw data, which was evaluated using machine learning algorithms. Initially, a total of 305 students were invited to participate in two distinct assessments, namely the graph motor test and the BHK test. The BHK test is conducted with the purpose of determining the presence of dysgraphia, which then used for training purpose of machine learning models. No feature of BHK test was used for training of models. Several classifiers, such as Random Forest, Ada Boost, Extra Tree, SVM, and Gaussian Naive Bayes, are used to train the model in this work. Linear Support Vector Machine (LSVM), obtained the highest accuracy of 73.4% among all machine learning techniques that were put into use.

Deschamps et al. (2021) ensured that the dysgraphia diagnosis system should be independent of the data gathering tool, several different models of tablets were utilized for the data collection process. There were 580 participants, which made this the largest dataset for this problem. Children from grade two through five were candidates for participation in the project. 122 of these have been prerecognized as children with dysgraphia. The participants were instructed to follow the French version of BHK using the digital tablet. There was a total of 100 different features extracted from collected data. Training and testing dataset size ratio was 80:20. The SVM machine learning classifier was used which produced specificity of 81% and a sensitivity of 91% for a given problem.

Dankovicova et al. (2019) suggested a machine leaning model which was based on Random Forest classifier for dysgraphia diagnosis. For data collection 78 participants were asked to perform various handwriting related activities on a Wacom Intous Pro tablet. 36 participants were suffering with dysgraphia. For the training purpose, various temporal and kinematic features, such as stylus angle, stylus lifts etc. were extracted from each sample. Three distinct machine learning classifiers Random Forest, AdaBoost and SVC were trained from extracted features. According to the findings, the Random Forest classifier exhibited best classification performance.

An online, handwritten feature-based assessment tool was presented by Rosenblum and Dror (2016) for the identification of dysgraphia in students enrolled in the third grade. 99 children have taken part in the research, of which 50 students have handwriting skills that are considered to be competent. In order to collect the online handwriting data, participants were asked to perform multiple assignments on a paper that is set on top of a digital tablet. The ComPET tool is used for collecting data. The participants were given the task of writing a Hebrew phrase consisting of six words that was taken from a well-known children's book. Extracted features were used to train a SVM linear Model. The results of the classification indicated that the suggested method is effective in classification with a sensitivity and specificity of 90%.

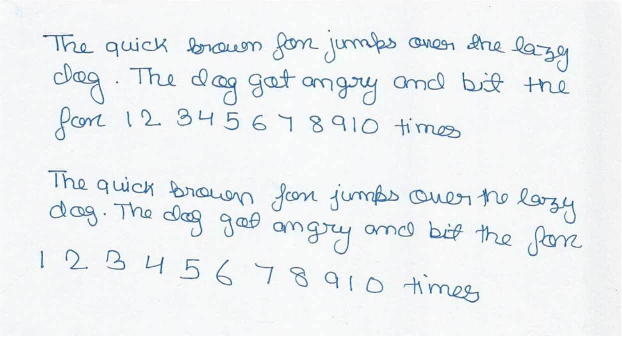
3 Materials and methods

3.1 Dataset

Research in medical data classification is increasing tremendously. Imaging techniques are widely used in the detection of many diseases. Major problem with medical data is the small size of the dataset. Although many data sets are available for different diseases, but for dysgraphia, finding a suitable data set was a challenge, so we collected our own dataset for the experimentation. In our study to train classification models we need labelled data for children without dysgraphia and children with dysgraphia. Data for dysgraphic children were collected through various sources like trained professionals, counselors, Internet and previous studies. A total of 112 images are collected under the dysgraphia category.

Handwriting images of children without dysgraphia are collected from various schools (Fig. 3). A total 2100 students participated in the data collection and 2881 samples were gathered, out of which 2473 were found suitable for the training (Fig. 4). The data was obtained through the implementation of two activities involving students. Table 2 shows how the samples distributed by grade.

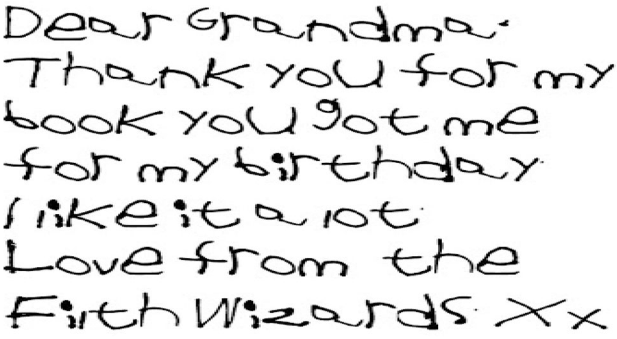
The first activity involved creative writing exercises for students. Each student was allotted a topic appropriate to their class and given a certain amount of time to write their thoughts on lined paper. During the second task, participants were requested to transcribe a sentence onto a blank sheet of paper. Individuals with hand injuries or those who were physically incapable of writing were excluded. Each subject's eyesight was either normal or adjusted for it. Dysgraphia must be diagnosed in order to qualify. Subjects with additional developmental problems, which frequently co-occur with dysgraphia, were not excluded from the study. Our goal was to provide widely applicable diagnostic tools for identifying dysgraphia.



The quick brown fox jumps over the lazy
dog. The dog got angry and bit the
fox 1 2 3 4 5 6 7 8 9 10 times

The quick brown fox jumps over the lazy
dog. The dog got angry and bit the fox
1 2 3 4 5 6 7 8 9 10 times

Fig. 3 Sample handwriting image of a child without dysgraphia



Dear Grandma:
Thank you for my
book you got me
for my birthday
I like it a lot
Love from the
Firth Wizards Xx

Fig. 4 Sample handwriting image of a child with dysgraphia

Table 2 Grade wise data collection

Grade	Number of Samples
1	304
2	168
3	204
4	652
5	464
6	573
7	516

3.2 Proposed Methodology

This section provides a description of the data preprocessing stages and the proposed model which are shown in Figs. 5 and 6, respectively. This approach comprises two main stages: data preprocessing and training the ViT model. First of all, preprocessing is performed to obtain similar size and clean images. After that, The image dataset is separated into train, validation, and test sub-sets. The model was trained through the training dataset and then the prediction on validation and test data is analyzed and performance of the classifier to correctly classify images into normal and dysgraphia is calculated.

In our research, we have deployed Vision Transformer and some popular CNN architectures namely VGG16 (Simonyan and Zisserman 2014), VGG19 (Simonyan and Zisserman 2014). ResNet50 (He et al. 2016), and InceptionV3 (Szegedy et al. 2016).

3.3 Vision transformer (ViT)

The field of natural language processing is the one where the concept of transformer is primarily used. Vision Transformer (ViT) is application of transformer over image data. In ViTs, images are converted into a sequence of

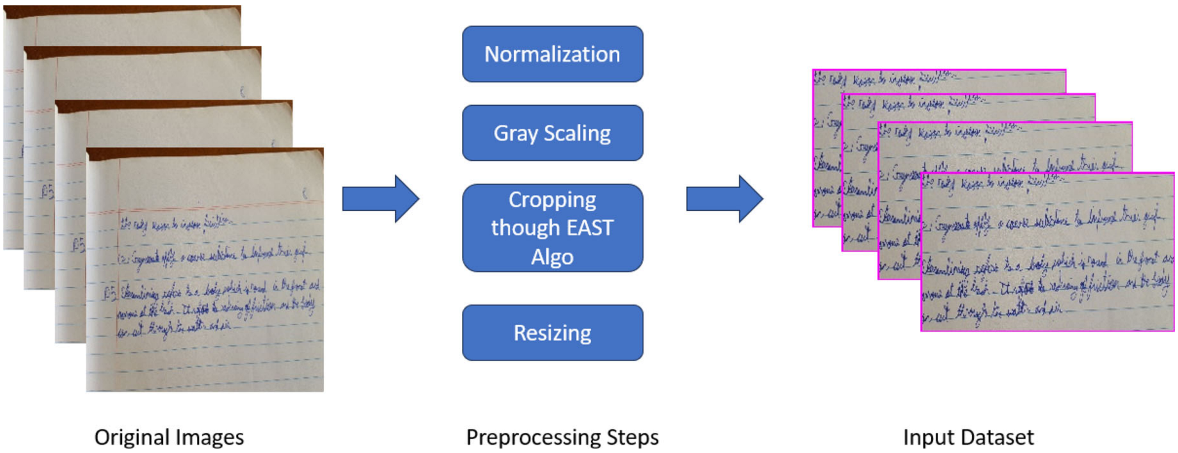
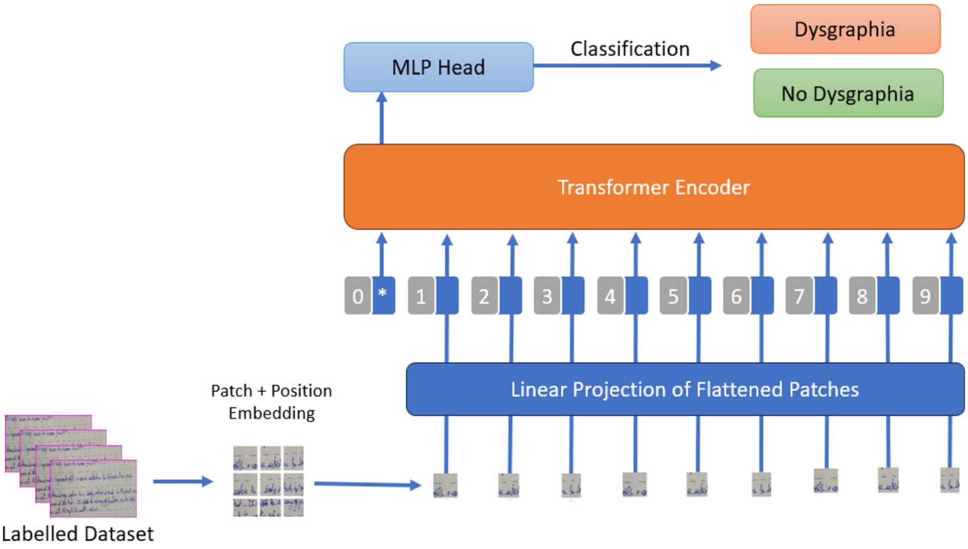


Fig. 5 Preprocessing Steps

Fig. 6 Proposed ViT-based Classification Model



patches, which allows models to learn image structure independently through parallel processing (Dan et al. 2022). To understand ViT, the most important thing to understand is attention and self-attention. The attention mechanism emphasizes the critical aspects of the input data while downplaying the less significant details. Think about the situation where you are captioning an image. In order to produce captions that are significant, there was a need to focus on the relevant portion of the image. This is the function of the attention mechanisms. Attention is an interface that connects the encoder and the decoder. It supplies the decoder with information from every hidden state of the encoder. The model is able to choose to focus on crucial parts of the sequence with the help of this framework, and as a result, it can learn the relationship between those components.

In the case of a typical encoder decoder based recurrent neural network, given that the encoder's hidden state is $h = h_1, h_2, \ldots h_n$, and the decoder's previous state is y_{i-1}. We can represent attention like

$$\alpha_i = \text{attention}\ (y_{i-1},\ h) \tag{1}$$

So basically, we calculate a score through the decoder's previous hidden state and encoder's current hidden state. So, for each hidden state which is denoted by j, we can calculate a scalar

$$\alpha_{ij} = \text{attention}(y_{i-1}, \text{h}_j) \tag{2}$$

One kind of attention mechanism is known as the self-attention mechanism. This form of attention mechanism enables each component of a sequence to interact with one another and choose who they need to focus their attention on more. Now, rather than searching for an input–output sequence relationship or alignment, we are searching for scores that correspond to each element in the sequence. The functioning of a self-attention module involves comparing each word in the phrase to each other word in the sentence, including itself, and then reweighing the word embeddings of each word to take into account the significance of the context in which the word is being used.

4 Experimentation

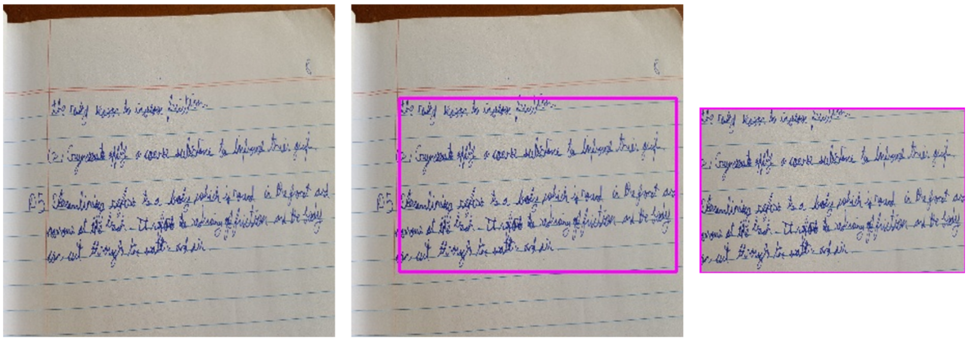
In this section, a comprehensive explanation is given regarding the experimentation of the proposed approach. The pre-processing of the data is an important and crucial stage in machine learning that serves to improve the quality of the data to be utilized. The primary issue with dataset was white space around the images. This was the most crucial part as most of the images had lot of empty white space around written content which was dominating the learning process and disturbing model learning. So, it was very important to remove the white space around the handwritten text. We modified the existing Efficient and Accurate Scene Text detector (EAST) algorithm (Zhou et al. 2017) to eliminate the white space. EAST algorithm uses a fully convolutional network (FCN) model. This model is generally used for generating word or line level boundaries by excluding the intermediate processing steps which are redundant and slow. We modified existing EAST algorithm as its designed to crop individual word and letters only (Fig. 7). Instead of using individual bounding boxes, we created a larger bounding box that encompasses all predicted boxes for a text instance. For this purpose, we calculated the X_{min}, X_{max}, Y_{min} and Y_{max} from all the bounding boxes predicted by EAST algorithm. An enlarged bounding box is created using these coordinates. This enlarged bounding box ensured that the complete text region is captured, reducing the risk of incomplete cropping.

While performing training of the classification model, the same size of the input is required. It was discovered that many of the input images varied in size, so they were all resized to 224 × 224 pixels. If images are reduced to very small dimensions, there are the chances that some important information of images may also get lost. Along with size standardization data augmentation was also performed to obtain new images with zooming and width-height range parameters (Table 3).

In this research, vision transformer and transfer learning approach is used for the detection of dysgraphia. For data augmentation *Keras ImageDataGenerator* library is used. *zoom_range* is used for a random zoom of 10% on training images. *width_shift_range* and *height_shift_range* are functions which create shifted version images of training data. Keras provides functions for horizontal and vertical random shifting of training data artificially. Here 10% of horizontal and vertical shift were used.

In this research ViT is used for classification purposes. The handwriting images have been partitioned into multiple smaller patches of a fixed size. Subsequently, each patch is augmented with position embeddings. Finally resulting sequence of vectors is fed to a standard Transformer encoder. Figure 8 shows the normal text image and image divided into patches while using Vision Transformer.

The other approach used for classification is transfer learning. Transfer learning is a strategy in which knowledge learnt from one task is reused to increase performance on a different but related task. It is a very common and popular method in deep learning, in which pre-trained models are used as the beginning point for different applications like computer vision and NLP tasks. We preferred to use transfer learning models for our comparison for two reasons. First, our local training dataset was small, and transfer learning allowed us to make the most of the limited target training data. Second, transfer learning enhances model generalizability, as the pre-trained models can incorporate learning from the new dataset as well. In this classification problem, due to the availability of fewer dysgraphic images, a transfer learning approach is used and VGG16, VGG19, ResNet50 and InceptionV3 are the CNN architectures that were used for pre-training. These CNN architectures use convolutional layers which are used for


Original image

Calculated Text Region

Cropped Image

Fig. 7 Image Cropping through modified EAST Algorithm

Table 3 Data Preprocessing Methods

Technique	Value
Resize	224 × 224
Zoom Range	0.1
Width_shift_range	0.1
Height_shift_range	0.1

Table 4 Architecture of CNN models

CNN model	No. of layers	Model description
VGG16	16	13 convolutional layers + 3 FC layers
VGG19	19	16 convolutional layers + 3 FC layers
ResNet50	50	49 convolutional layers + 1 FC layers
InceptionV3	48	47 convolutional layers + 1 FC layers

feature extraction, pooling layer for reducing the parameters and fully connected layers for classification tasks. Every architecture has a different number of convolutional and pooling layers. With the stack of layers higher level features are extracted. Table 4 shows the description of the models used for transfer learning.

Transfer learning is very popular in the deep learning area. It has already proven its worth in medical image classification applications, as to get large amounts of labeled data in the medical field is a cumbersome task so transfer learning helps in getting pretrained models which are already trained over large amounts of data and then using this knowledge with our own model which has less data for the training.

In present work, ViT and Transfer Learning approaches are used for the classification purpose and their performance is also evaluated. The evaluation of all the models is performed using different measures. In this research work, the following performance measures are used for the evaluation:

Precision: it tells proportion of positive predictions by the model are actually correct.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Recall: it tells what proportion of actual positives is detected correctly by the model.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

F1 Score: F1 score is calculated by taking harmonic mean of precision and recall, it is used to maintain the balance between Precision and Recall.

$$\begin{aligned}\text{F1 Score} &= 2 * (\text{Precision} * \text{Recall})/\text{Precision} \\ &\quad + \text{Recall}\end{aligned}$$

Accuracy: Accuracy shows the number of correctly classified data samples out of the total data samples.

$$\text{Accuracy} = \text{TN} + \text{TP}/(\text{TN} + \text{FP} + \text{TP} + \text{FN})$$

where FN is false negative, FP is false positive, TN is true negative and TP is true positive.

Here macro average scores are used for the Precision, Recall and F1 scores. The macro average is the arithmetic

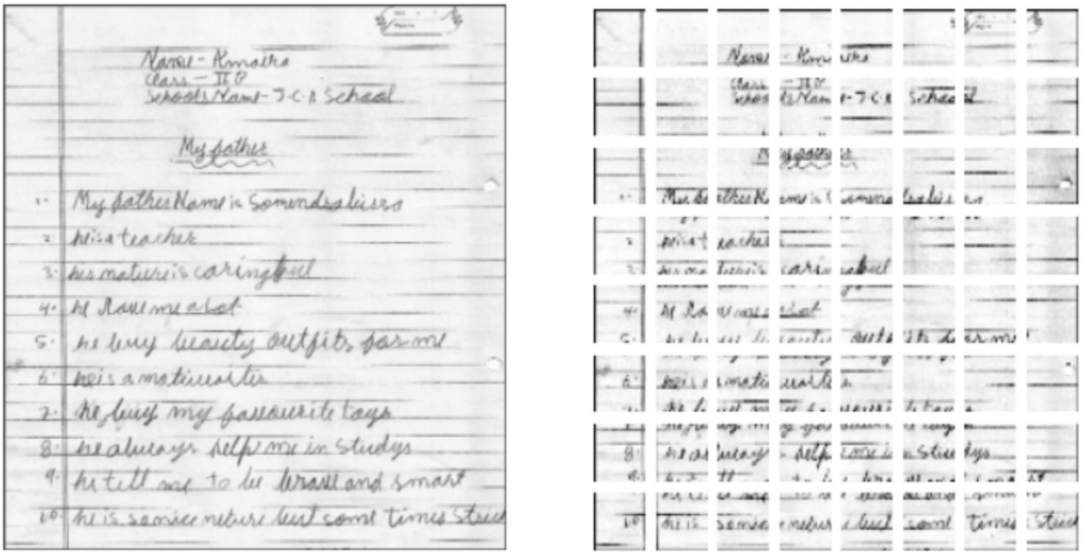

Name - Rmaitha
Class - II P
Schools Name - J.C.A School
My father
1. My father Name is Somendsalires
2. he is a teacher
3. his nature is caringfull
4. he Love me alot
5. he buy beauty outfits for me
6. he is a motivaiter
7. he buy my favourite toys
8. he always help me in studys
9. he tell me to be brave and smart
10. he is sanier neture but some times Stric

Fig. 8 Normal text image and text image with patches created by Vision Transformer

Table 5 Validation loss of vision transformer

Epoch	1	2	3	4	5	6	7	8	9	10
Loss	0.6943	0.1692	0.2187	0.1696	0.1434	0.0661	0.0970	0.0126	0.0296	0.0916

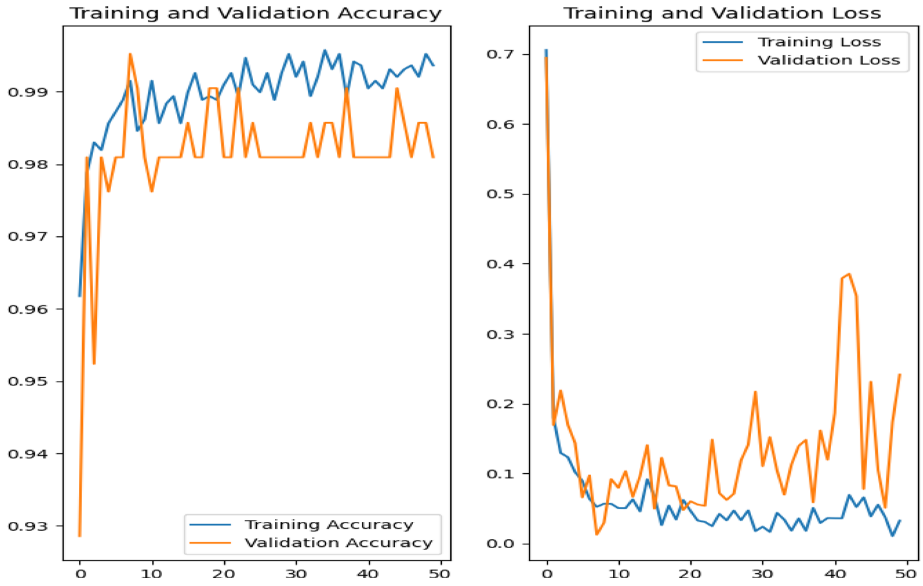


Fig. 9 Validation vs. training accuracy and loss of ViT

Table 6 Performance of applied classification models

Model	Precision average	Recall average	F1 score average
ViT	0.99	0.86	0.92
Transfer learning			
VGG16	0.94	0.88	0.90
VGG19	0.91	0.84	0.87
ResNet50	0.90	0.82	0.85
InceptionV3	0.88	0.81	0.84

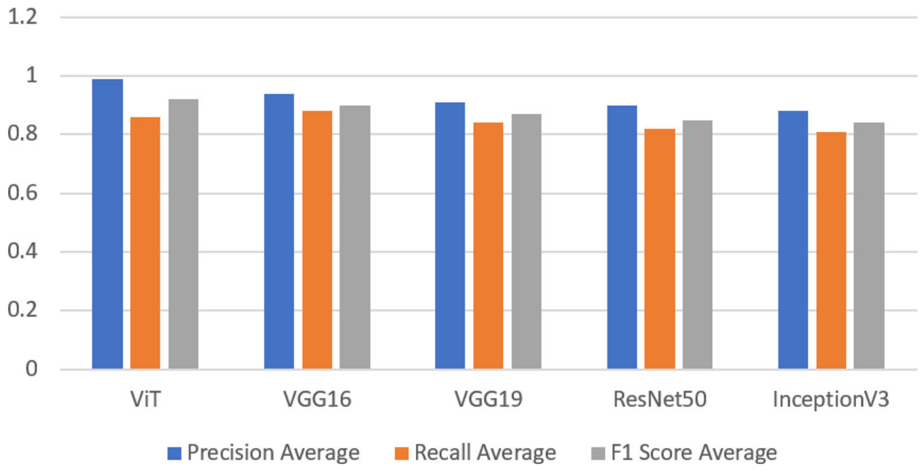


Fig. 10 Performance of classification models

mean of each class related to Precision, Recall, and F1 score. Macro Average scores are used when all classes need to be evaluated equally to evaluate the overall performance of the classifier against the most common class labels.

5 Results and discussion

In this study a dataset containing handwritten samples was collected from school children. The implementation was performed using Python on Google Colab. NumPy, Pandas, SciPy, Scikit-Learn, TensorFlow, and Keras were used for the purpose of implementation. The results of different tasks are discussed in the following paragraph.

In this research, ViT is used for the classification of handwritten text images in two categories, normal and dysgraphia. Table 5 shows the validation loss of ViT for the first 10 epochs. Figure 9 demonstrates the training and validation accuracy of the model. As the epochs increased, there is improvement in training and validation accuracy. In addition to this, significant decrease in training and validation loss was also observed. The training and validation loss started with 0.7054 and 0.6943 values in first epoch and at the end it decreased to 0.0327 and 0.2414 respectively.

In this work, we have also used the power of transfer learning. In research, transfer learning has already shown its worth. Here four different models of transfer learning are used. Table 6 shows macro average values of precision, recall and F1 score for all the classification models.

Figure 10 presents macro average values of precision, recall and F1 score in graphical format. The performance of Vision Transformer is compared with pretrained models. Data augmentation helped to generate new data as small dataset was available.

In terms of the macro average F1 score, Fig. 10 illustrates that the Vision Transformer surpassed other models, achieving a value of 0.92. In terms of the macro average F1 score, InceptionV3 attained the lowest value at 0.84.

6 Conclusion and future work

Machine learning algorithms have already proven their importance in detection of various medical conditions which needs expert's attention. This paper describes a framework for dysgraphia classification of hand-written text images of children. In this study various approaches of ML are explored for the detection of dysgraphia. Data is collected from various schools at Jaipur. Children of age group 6–12 are taken as the target group. Here, ViT-based framework is used for the detection of dysgraphia. Since dataset was small in size, so data augmentation techniques are also used. For the comparison purpose transfer learning approach is also used, different transfer learning based pre-trained models VGG16, VGG19, ResNet50 and InceptionV3 are used. After comparing the results, it was found that Vision Transformers are best suitable for the classification task. Due to class imbalance in our dataset, accuracy cannot be considered a reliable performance indicator. The Vision Transformer exhibited superior performance, boasting precision, recall, and F1 score values of 0.99, 0.86, and 0.92, respectively. Among all pretrained models, VGG3 stood out as the top performer with precision, recall, and F1 score values of 0.94, 0.88, and 0.90. In summary, our findings suggest that the proposed Vision Transformer-based model surpasses current state-of-the-art techniques for dysgraphia detection.

This study also has some limitations as it suffers from less amount of data for children with dysgraphia. In the future, more data can be collected especially dysgraphia subjects. Secondly, we can also extract some important features like pressure and angle of pen at the time of writing, skew angle of the text etc. to get more precise results and for this hand-crafted feature extraction can also be explored. Another aspect is incorporating explainable artificial intelligence to enhance trust of user in system's decision.

Acknowledgements I am also thankful to my colleagues Dr. Manisha at Indian Institute of Information Technology, Kota for her collaboration and constructive discussions. The exchange of ideas and perspectives greatly enriched the quality of this work.

Author contributions Prateek Sharma implemented the proposed methodologies and wrote the main manuscript text. Basant Agarwal, Gyan Singh Yadav and Sonal Jain guided the research and helped in data collection, methodology preparation and result evaluation. All authors reviewed the manuscript.

Funding This research received financial support from Department of Science & Technology, Government of India. The funding source had no involvement in the study design, data collection, analysis, interpretation, or decision to submit the manuscript for publication.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest I, Prateek Sharma, declare that there are no conflicts of interest that could influence the research findings or the interpretation of the data presented in this manuscript. Any potential conflicts of interest among the authors have been disclosed and managed appropriately.

References

Agarwal B, Jain S, Bansal P, Shrivastava S, Mohan N (2023a) Dysgraphia detection using machine learning-based techniques:

a survey. In International Conference On Emerging Trends In Expert Applications & Security. Springer, Singapore. pp 315–328

Agarwal B, Jain S, Beladiya K, Gupta Y, Yadav AS, Ahuja NJ (2023b) Early and automated diagnosis of dysgraphia using machine learning approach. SN Comput Sci 4(5):523

Asselborn T, Gargot T, Kidziński Ł, Johal W, Cohen D, Jolly C, Dillenbourg P (2018) Automated human-level diagnosis of dysgraphia using a consumer tablet. NPJ Digit Med 1(1):42

Asselborn T, Chapatte M, Dillenbourg P (2020) Extending the spectrum of dysgraphia: a data driven strategy to estimate handwriting quality. Sci Rep 10(1):3140

Beery KE (2004) Beery VMI: the Beery–Buktenica developmental test of visual-motor integration. Pearson, Minneapolis, MN

Chung P, Patel DR (2015) Dysgraphia. Int J Child Adolesc Health 8(1):27

Chung PJ, Patel DR, Nizami I (2020) Disorder of written expression and dysgraphia: definition, diagnosis, and management. Transl Pediatr 9(Suppl 1):S46

Crouch AL, Jakubecy JJ (2007) Dysgraphia: how it affects a student's performance and what can be done about It. Teach except Child plus 3(3):n3

Dan Y, Zhu Z, Jin W, Li Z (2022) S-Swin transformer: simplified swin transformer model for offline handwritten Chinese character recognition. PeerJ Comput Sci 8:e1093

Dankovičová Z, Hurtuk J, Feciľak P (2019) Evaluation of digitalized handwriting for dysgraphia detection using random forest classification method. In: 2019 IEEE 17th International Symposium on Intelligent Systems and Informatics (SISY). IEEE. pp 000149–000154

Deschamps L, Devillaine L, Gaffet C, Lambert R, Aloui S, Boutet J, Jolly C (2021) Development of a pre-diagnosis tool based on machine learning algorithms on the BHK test to improve the diagnosis of dysgraphia. Adv Artif Intell Mach Learn 1(2):114–135

Devi A, Kavya G (2023) Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches. Prog Neuro-Psychopharmacol Biol Psychiatry 120:110647

Devillaine L, Lambert R, Boutet J, Aloui S, Brault V, Jolly C, Labyt E (2021) Analysis of graphomotor tests with machine learning algorithms for an early and universal pre-diagnosis of dysgraphia. Sensors 21(21):7026

Dimauro G, Bevilacqua V, Colizzi L, Di Pierro D (2020) TestGraphia, a software system for the early diagnosis of dysgraphia. IEEE Access 8:19564–19575

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Houlsby N (2020) An image is worth 16 × 16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929

Drotár P, Dobeš M (2020) Dysgraphia detection through machine learning. Sci Rep 10(1):21541

Dui LG, Lunardini F, Termine C, Matteucci M, Stucchi NA, Borghese NA, Ferrante S (2020) A tablet app for handwriting skill screening at the preliteracy stage: Instrument validation study. JMIR Ser Games 8(4):e20126

El-Gayar OF, Ambati LS, Nawar N (2020) Wearables, artificial intelligence, and the future of healthcare. AI and big data's potential for disruptive innovation. IGI Global, pp 104–129

Hartman DE (2007) Wide Range Assessment of Memory and Learning-2 (WRAML-2): WRedesigned and WReally Improved

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778

Kunhoth J, Al-Maadeed S, Kunhoth S, Akbari Y, Saleh M (2024) Automated systems for diagnosis of dysgraphia in children: a survey and novel framework. IJDAR. https://doi.org/10.1007/s10032-024-00464-z

Mekyska J, Galaz Z, Safarova K, Zvoncak V, Mucha J, Smekal Z, Faundez-Zanuy M (2019) Computerised assessment of graphomotor difficulties in a cohort of school-aged children. In: 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE. pp 1–6

Meyers JE, Meyers KR (1995) Rey complex figure test under four different administration procedures. Clin Neuropsychol 9(1):63–67

Rosenblum S, Dror G (2016) Identifying developmental dysgraphia characteristics utilizing handwriting classification methods. IEEE Trans Hum-Mach Syst 47(2):293–298

Roth RM, Isquith PK, Gioia GA (2013) Assessment of executive functioning using the Behavior Rating Inventory of Executive Function (BRIEF). Handbook of executive functioning. Springer, New York, NY, pp 301–331

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2818–2826

Yogarajah P, Bhushan B (2020) Deep learning approach to automated detection of dyslexia-dysgraphia. In: The 25th IEEE international conference on pattern recognition

Zhou X, Yao C, Wen H, Wang Y, Zhou S, He W, Liang J (2017) East: an efficient and accurate scene text detector. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp 5551–5560

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.