

Identifying Learning Disability Through Digital Handwriting Analysis

S. V. Kedar¹, Purva P. Parab², Akash R. Sharma³, Jaikrishna M. Patil⁴, Rohit T. Wagh⁵

¹JSPM's RSCOE, Pune, India

²Xoriant, Pune, India

³Yardi Software, Pune, India

⁴MSc (CS), TU Kaiserslautern, Germany

⁵TCS, Pune, India

seema_kedar@yahoo.com¹, parabpurva9@gmail.com², akashsharma.comp@gmail.com³,

jaikrishnapatil007@gmail.com⁴, rohitw2698@gmail.com⁵

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

Abstract: Handwriting Analysis is described as a scientific study for the analysis of handwriting. It is a way of interpreting and ability to learn from peculiarities in handwriting. Offline handwriting analysis is a traditional approach that cannot be used efficiently for analysis. Online handwriting analysis, on the other hand, can utilize various aspects like pressure on the pen, timestamp and other factors which help in improving the effectiveness of analysis. Learning disabilities are neurological processing problems which can hamper the learning of the children. Dysgraphia is a learning disability that mainly affects a child's handwriting and motor skills. It is found in 10 to 30% of school-aged children. Dysgraphia can be diagnosed by therapists based on children's handwriting samples and manual evaluation techniques. This method is lengthy and inaccurate. In this work, automatic identification methods for and classification of dysgraphia in children in the age group 7 to 12 is described. The method performs analyzing of the child's writing dynamics via blueprints of the pressure the pen puts on paper with the pen's movements and orientation with the use of a standardized digital writing pad and machine learning algorithms. It basically has two phases, the training phase, and testing phase. In the training phase, handwriting samples of known results are given to the system. Then the model is built using some classifier, Random forest or Support Vector Machine. Once the model is trained, then in testing phase this model is used for classification of unknown samples to predict whether the subject has dysgraphia or not. This is then used to check the accuracy of the designed system.

Keywords: Handwriting Analysis, Dysgraphia, Random Forest, Support Vector Machine

1. Introduction

One of the language production form is writing which is difficult in ranging from the idea of conceptualization to running of motor by hand, leading to handwriting. It is a difficult human movement, measured to be an "overlearned" skill which includes adding rapid sequencing of movements in time, ultimately generating the relationship between plan and generation of product. In detail, 50% of a school days of child is spent performing handwriting tasks. For example, handwriting speed was found to be important for note-taking-recording important information, and handwriting automaticity was correlated with children's composition performance variance. Though the use of computers is widespread, handwriting still serves as a media for communication and is a mandatory life skill. All action, including writing, starts in the brain. Like all other movements, the act of writing depends on the central nervous system. Our brain sends impulses to hand through nervous signals, achieving the motor act. For becoming proficient writers most of the childrens may adjust their handwriting requirements ultimately. As their handwriting might legible, so that bit of a efforts investments can be done by them in the process of handwriting. The process of adopting the letters' form begins in the first grade. Below specified are pair of school years of handwriting, experience leading to unify an automatic proficient way of letters writing, that occurs around age eight. Thus, childrens with the same age that do not succeed in developing proficient handwriting face developmental dysgraphia. Few important aspects of handwriting analysis used for health diagnosis are congestion, the direction of lines, the layout of anomalies, torsion, viscosity, shakiness, slant, movement between stokes, variation in size, alterations in shape, breeze, pressure, periods and accent marks.

2. Literature Survey

In paper [1], the major work is done to understand the characteristics of adolescents with dysgraphia. Authors have used 80 adolescents, 40 with dysgraphia and 40 without dysgraphia. HPSQ-C was used to differentiate between the children with handwriting difficulties and the ones without handwriting difficulties. Wacom Intuos II x-y digitizing tablet (404 × 306 × 10 mm) along with wireless electronic pen was used for taking the writing samples of adolescents. The major focus was on the measures such as the mean stroke performance time in seconds, the mean stroke width in centimetres and the mean stroke height in millimetres. The six traits writing model was used to decide the quality of the written samples. 43% of the variability of mean stroke duration and 63% of the variability of the content quality was predicted by executive functions and developmental and functional measures.

Paper [2] depicts a client-server web-based software system, which is functional as recent devices like tablets and mobilephones which makes use of cutting edge JavaScript APIs and framework and standard algorithms for multiple hand gesture detection, having the name Dynamic Time Warping algorithm that was customized to recognize the complex gestures. The software tools offer the users with the probability of executing collections of dissimilar exercises types, arranged at different levels, from just linking the dots in order to completion of a word writing, ultimately comparison of writing performed with a reference tracking by an expert. The software tool provides swift feedbacks depending upon objective considerations, along with an inclusive collection of data stored both in JSON and INKML structure helps in identifying, studying and rehabilitating dysgraphic handwriting.

In paper [3], the authors Nan-Ying Yua and Shao-Hsia Chang proposed a method for checking the correctness of Chinese handwriting by comparison of each stroke actions amongst template models and a handwritten script. 14 subjects with amnesic mild cognitive impairment (aMCI) and 22 subjects with Alzheimer's Disease (AD) were considered for evaluation. The Mean Squared Error (MSE) of the divergence in the sketchcourse and the goal line was computed to estimate the spatial accuracy. One-way Analysis of Variance (ANOVA) for continuous measures have been used for comparison of handwriting features of three groups of subjects. According to this study pausing on paper (PTS) and the dysfluency of stroke association (RAPL) were better indicators of writing difficulty than the temporal parameter (RAPT).

In Paper [4], an optoelectronic organization allowed three-dimensional acquisition drawing track. The performer's position of the head as well as higher branch actions has been assessed as well. Collectively arguments with association duration, velocity, trace's length, boundaries of action of the greater limb, was calculated and matched in the middle of the 3 groups. Kids with dyslexia tracked the loopswiftly than the other groups. In the cross test, dyslexic participants seen to have a compact execution time and greater than before velocity at drawing the horizontal line. Kids with dyslexia were also faster in sketchsure sides of a rectangle with respect to the other groups. As several children with dyslexia and dysgraphia have specific drawing and writing obstacles because of the varieties of perceptual-motor along with knowledge problems (Mati-Zissi, Zafiropoulou, & Bonoti, 1998; Rosenblum, Dvorkin, & Weiss, 2006), graphic tests are basic for the clinical assessment of these subjects. The main difficulties in writing making were observed incomprehension of differences, organization of parts in an organized whole, spatial movement, size scaling, and classification or distinction of figures.

Paper [5] introduces a right view of parameterizing early Alzheimer, depending upon the assessment of online handwritten cursive loops. Distinct from the survey, they model the loop velocity trajectory (full dynamics) in an unsupervised way. By a timecluster based on K-medoids, in association of active time warping as unsynonymous compute, they introduces clusters which gives new details on the problems. As a characterization, Bayesian formalism can be supposed which averages the task of the clusters in this, by probabilistically combining the discriminative control of each. On a dataset consisting of two cognitive profiles, early-stage Alzheimer disease, and healthy persons, each comprising 27 persons composed at Broca Hospital in Paris, The classification presentation significantly outperforms the state-of-the-art, based on universal kinematic features.

Paper [6] depicts that dysgraphia of sufficient severity for interfering through work of the school has been considered as describing a unique deficit; classification of its actions problems is a fundamental step in order to offer optimized interference. From between kids aged 6–8, 69 children having dysgraphic uniqueness (learn collection) and 69 compared efficient hand-writers (organize cluster) have been consulted in this learning. Four copying tests of varying difficulty levels were monitored using a digital tablet. The acquired data involved a through measure (axial pen tip force) and derived arguments (stroke velocity, pause time, number of velocity peaks and the ratio of In Air to On Paper measures). The main observation is, kids with dysgraphic characteristics had maximized pause time per each strike and a maximized directional changes in velocity. Emphasized differences were also originated amongst student in couple of different grades, mainly in the control group. The values obtained and observed in this study can distinguish as further and classify the handwriting troubles generated by well motor deficits.

In paper [7], Digital script checks have been used to give unique facts of handwriting character. Adolescents with ASD are practical as lower handwriting and lower handwriting constancy (i.e. greater changes in speed and/or pen pressure). The handwriting was assessed by The Computerized Handwriting Speed Test System, Version 2 (CHSTS-2). The participants were tell to copy English words and Chinese characters on A4 sized papers affixed to the outside of a Wacom Intuos Pro L tablet by a wireless electronic pen. CHSTS-2 provided handwriting information on the ground time, airtime, handwriting speed (character per minute), Standard Deviation (SD) of writing time per character, pen pressure, SD of pressure, and readability (number of accurate words or number of

recognized words by the system). Ground time and airtime are behavioral data instead of the length of time when the pen touches the paper and is detained in the air, correspondingly. Four sensorimotor presentation components including motor skills, visual perceptual skills, visual-motor integration, and eye movement were assessed in the study. The objectives of this learn were to evaluate the handwriting presentation and sensorimotor mechanism, together with motor skills, visual perceptual skills, visual-motor integration, and oculomotorbe in charge of adolescents with ASD with typically developing adolescents; and examine the role of sensorimotor mechanism in handwriting presentation in the middle of adolescents among ASD.

In paper [8], the handwriting and understanding of 5 to 8-year-old kids were linked with the manufacture of the ILT, a easy, graphomotor skill learning task. the time-dependent course of motor skill acquisition.) Included (The ILT have been use for study In Phase I of this study, kids handwriting (speed and legibility) was assessed contemporaneously with the ILT. The following year, Handwriting and understanding was been assessed by Phase II. The outcomes clarify that presentation of the ILT emitted well-built connection to handwriting and understanding.

In paper [9], writing performances were recorded using an LCD digitizing tablet (a 21-inch Wacom Cintiq 21UX) connected to a laptop computer (Apple MacBook) piloted by Eye and Pen software ®. Participants wrote on a plastic board placed on top of the tablet screen using a pen (Wacom Ink Pen) with a plastic tip (no ink). They used the Eye and Pen software to (i) record the position and state of the pen tip (with or without pressure) on the tablet in real time, (ii) manage the display of each item to be written, as well as the visual (letter and pseudoword formation) feedback displayed on the screen, and (iii) provide velocity and kinematic (pressure exerted on the pen, pen movement speed, pause duration, distance covered by the pen) at the end of the task. As the plastic board was transparent, participants could see their handwriting in the normal vision condition (vision condition).

In paper [10], a customized handwriting instrumentation system is described, mode of off-the-shelf hardware which simultaneously measures grip forces, normal forces, x–y positions on the tablet, and time information. They have concluded several objective parameters from these measures. For the study purpose A Wacom 9×12 in Intuos3 digital tablet and an instrumented wireless pen were used. A tradition graphite nib for the Wacom pen was developed for better estimated the feel of a pencil. The pen was further instrumented with a TekScan model 9811 pressure sensor array on its barrel (4 elements in azimuth by 8 elements along the axis of the pen). The sensors precise the force applied radially to the barrel by the user’s hand, i.e., the grip force.

In paper [11], there are various handwriting parameters considered such as whole time, per character time used, ‘in-air time’, and speed of handwriting. It is seen that those who faces challenges in normally have a weak performance in writing task, and may require extra ‘in-air time’ i.e. time spent when pen tip without touching paper surface than normal writers.

Paper [12] represents two groups of hand writers, capable and dysgraphic. The school children from 3rd grade were included in the study. The objective of survey is to put a way which may be used for regular diagnosis for this disorder, as well as for judgment of complexity level as set by the screening questionnaire of handwriting proficiency. Standardized and validated ten-item questionnaire for Handwriting Proficiency (HPSQ) helped in identification of dysgraphic hand writers. They used Computerized Penmanship Evaluation Tool (CompPET) for data acquisition. These particular papers have built-in an automatic rating of developmental dysgraphia at population of children’s that includes the state-of-the-art handwriting parameterization techniques and easy intra-writer approach for normalization. Parameterization which is based on 51 features introduced by it and found that those based on altitude/tilt and pressure discriminate well D and C children. Their system was capable of rating developmental dysgraphia along with estimation of entirety score with 10% estimation error of HPSQ. Random forest classifier was used by the system for classifying proficient and dysgraphic.

Paper [13] represents two couple of group of hand writers, proficient and dysgraphic. The school children from 3rd grade were included in the study. Here, a way for automatic identification and characterization of dysgraphia in third-grade children is shown. These ways are depends upon assessment of child’s writing dynamics by sampling the pressure the pen exerts on the paper as well as the pen’s position and orientation by making use of a standard digital writing pad. Dysgraphich and writers were identified through the standardized and validated ten-item questionnaire for Handwriting Proficiency (HPSQ). In this study, handwriting data gathered from 99 writers were considered for training classifiers for a dysgraphia diagnosis. A correctness of 90% was secured by classifier. The fucose of this study is on characterization and identification dysgraphia among Hebrew writing children. Rather than Hebrew script specific (lamedHeight) single feature, the analysis is not based on any specific characteristics of Hebrew letters. SVM was used by the system for classifying proficient and dysgraphic.

Paper [14] represents the concept of handwriting and study of handwriting i.e. graphology. The Graphology science of uses minimum of 300 various features of handwriting investigation process. By following Precise rules for measurement of the writing actions like zones, size, slant, width, pressure, degree of connection, connective forms, end, and start strokes, baselines, space linking words and lines, the page layout, signature. Some of these elements are necessary to the very standard world of the writer and chosen robotically. Some are extra superficial and can change with frame of mind i.e. slant and size.

Paper [15] by Christina T. Fuentes, Stewart H. Mostofsky and AmyBastian, indicates various basics of handwriting causal to impairments observed in kids with autism. A study with case-control of samples of handwriting from children with and without Autism Spectrum Disorders(ASD) was functional using the Minnesota Handwriting Assessment. Samples were scored on an individual letter basis in 5 categories: legibility, form, alignment, size, and spacing. In this, to copy the words on the provided solid lines subjects were instructed (baselines) at bottom half of the test sheet, eventually which makes their letters the like size as the sample and using their better handwriting. It was observed that ASD children do indeed shows whole worse handwriting performance. Results suggested training targeted letter creation, in association with common preparation of well motor organize, and may be the best path for humanizing handwriting presentation in kids with autism.

Paper [16] by Roderick I. Nicolson and Angela J. Fawcett, they have taken an extended tour through dyslexia and other learning disabilities, focusing on the commonalities and differences between dyslexia and dysgraphia. Here, focus is on the developmental disorders of dyslexia (a disorder of reading) and dysgraphia (a disorder of writing), taking in picture their commonalities and differences with a picture of reflection upon theoretical implications. This hypothetical and experiential progress Results in two usages of the expression dysgraphia. One of which considers dysgraphia for referring to errors of writing that are equivalent to errors in reading, one more relates to challenges in handwriting control.

In paper [17], the objective is to develop an occupational therapy screening questionnaire (Handwriting Proficiency Screening Questionnaire [HPSQ]). The content of questionnaire establishment done for validity. Internal consistency, inter-rater and test-retest consistency, and simultaneous and construct validity were initiated. Participants included 7- to 14-year-old (N = 230) naturally developing school-aged kids. The tool established high-quality internal consistency ($\alpha = .90$). Test-retest constancy for the score showing an Intraclass Correlation Coefficient (ICC) of .84 and interrater reliability of ICC =.92 for the whole gain of survey. Also By confirming build and simultaneous validity.

In paper [18], Wacom Inutos digitizing tablet is used for data collection. All data analysis was performed off-line. This software has been programmed to include two independent parts: the data collection software and the data analysis software. Multiple parameters such as In-air time were calculated. This paper shows the potential of using tablets in handwriting analysis because it provides much more information than conventional Offline handwriting data collection technique.

In paper [19] for the gratitude of on-line handwritten characters a novel approach is presented. This combinely indicates Dynamic Time Warping (DTW) and Support Vector Machines (SVM) by integrating DTW into a Gaussian SVM kernel. The experiments are based on the digits, upper- and lower-case characters of the UNIPEN database. They have useful the proposed classification technique to characters of the UNIPEN handwriting recognition database. Training set size was 66 %, the test set size 33% of the UNIPEN database. Experiments have shown superior recognition rate in judgment to an HMM-based classifier for relatively small training sets and comparable rates for larger training sets. SVM GDTW has a low error rate of 4.0%, 7.6%, 13% for digits, upper and lower case respectively.

In paper [20], changes in marks velocity, grip services on top of the pen drum, and standard services on the writing surface, over the course of a 10-min writing job, in a large cohort of 4th-grade kids with and without dysgraphia are examine. Flat stroke speed, grip force and normal force greater than before over time while straight up stroke speed decreased in all kids. Samples of 105 fourth-grade kids (50 male, 100 right-handed) were taken with help of an experienced occupational therapist H. Schweltnus. The kids wrote all components of the assessment with an instrumented stylus on a tablet surface. By using a hierarchical linear model, statistical analysis was performed. probinglikely effects of time and CHES score on grip force, pressure, and speed. These answers point to that important changes in handwriting kinetics and kinematics occur in script tasks as short as 10 min in duration in both kids through and exclusive of dysgraphia.

3. General System

The system basically requires handwriting samples for working. So the first step is to collect data or handwriting samples. For this, we have used a WACOM INTUOS digital tablet. After the collection of data, the next phase is to extract the required features from the data and create the final dataset. Then, the machine learning algorithms are applied to this dataset in order to build a model for classification. The last phase is to test the model for accuracy.

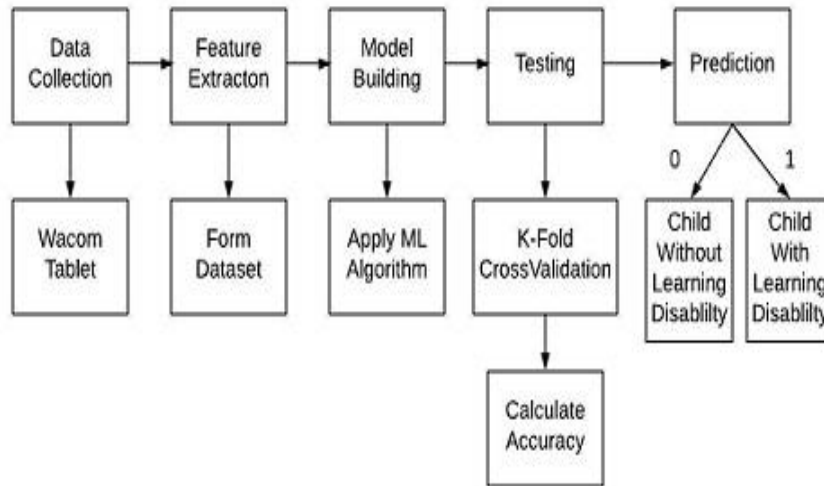


Figure 1. General System

All the 5 phases are explained in brief below:

3.1 Data Collection:

There are two groups of students, normal students who do not have any learning disability and dysgraphic students having some learning disability. Data is collected for both of these categories. Handwriting samples are collected through the WACOM digital tablet. WACOM tablet is provided along with interactive digital pen having a stylus for the tablet screen. A white paper is placed on the tablet which will help subjects to write accurately. Then the next step is to give information to the student about the data to be expected by them on the tablet. The data consists of 5 tasks which are illustrated in Fig 2. Once the data is collected for a student, one separate file is formed for each task.

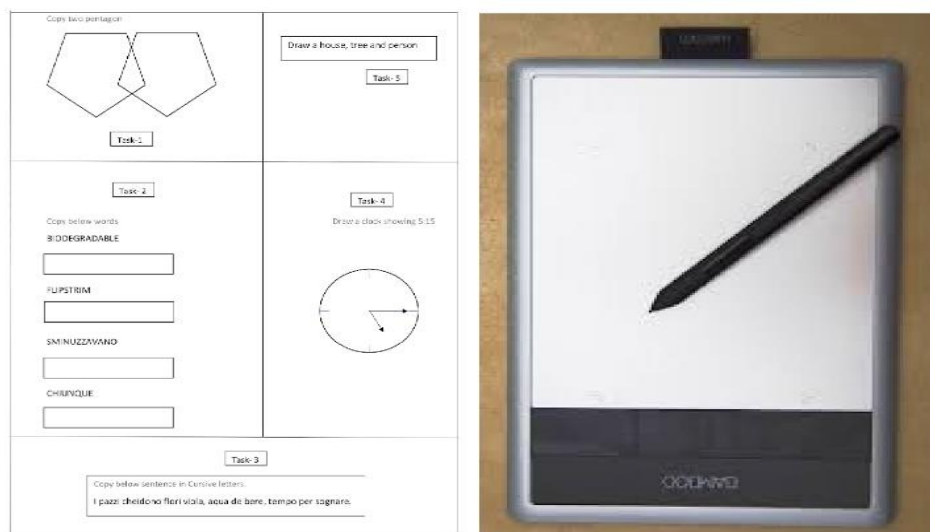


Figure 2. Tasks for online handwriting analysis

For each task, separate txt file is generated containing handwriting parameters. These parameters are timestamp, time in microseconds, x coordinate, y coordinate and pressure. Figure 3 shows snippet of the file containing parameters and its values.

```

01/21/19-17:09:22^1548070762814961^194^227^0 → Time Stamp
01/21/19-17:09:22^1548070762837085^195^227^0
01/21/19-17:09:22^1548070762875021^196^226^0
01/21/19-17:09:22^1548070762896994^196^225^0
01/21/19-17:09:22^1548070762921027^196^224^0 → Time in Micro
01/21/19-17:09:22^1548070762934961^196^223^0 → Seconds
01/21/19-17:09:22^1548070762957021^196^222^0
01/21/19-17:09:23^1548070763011036^196^221^0
01/21/19-17:09:23^1548070763044134^196^221^222 → X Co-ordinate
01/21/19-17:09:23^1548070763048673^196^221^249
01/21/19-17:09:23^1548070763057015^196^221^305 → Y Co-ordinate
01/21/19-17:09:23^1548070763066975^196^221^367
01/21/19-17:09:23^1548070763074008^196^221^404 → Pressure
01/21/19-17:09:23^1548070763080426^196^221^437
01/21/19-17:09:23^1548070763087087^196^221^455
01/21/19-17:09:23^1548070763095263^196^221^468
01/21/19-17:09:23^1548070763102419^196^221^477
01/21/19-17:09:23^1548070763108070^196^221^481
01/21/19-17:09:23^1548070763116311^196^221^486
    
```

Figure 3. File format

3.2 Feature Extraction:

In this phase, required features are extracted in order to form a final dataset. There are number of features that are used in online handwriting analysis. Feature extraction were carried out in Jupyter notebook using python. Features extracted for further analysis are as follows:

- 1) The total number of on-paper segments while completing a task.
- 2) The total number of in-air segments while completing a task.
- 3) Total time that is taken to complete the entire task.
- 4) Total Time spent in-air for completion of a test.
- 5) Total Time spent on-paper for completion of a test.
- 6) Length of a segment.
- 7) Mean pressure applied to the pen.
- 8) The standard deviation in pressure that is applied to the pen.
- 9) The velocity of the pen.

3.3 Building Model:

Once the dataset is formed, building a machine learning model for the system is the next step. This is formed by first dividing the dataset into a training set and testing set. Then various machine learning algorithms are applied to the training set in order to train the system. We have used sklearnlibrary in python to build these machine learning algorithms and achieve required accuracy. Different machine learning algorithms have unique significance. Various machine learning algorithms applied are as follows:

A. Decision tree:

Decision tree make use of tree illustration in order of solving the challenges in which each leaf node corresponds to a class label and attributes are represented on the interior node of the tree. We have used DecisionTreeClassifier()from sklearnfiles in python to build decision tree for the given dataset. Recognizing the attribute for the root node in each level in Decision Tree is the main test. This process is known as attribute selection. We have two accepted attribute selection measures: in order gain and Giniindex. In this model, it was found that attribute selection using Gini index gave best result. Moreover, maximum depth for the tree was set to 5. Futhermore, minimum number of samples required to split a node were set to default value, i.e. 2.

B. Random forest:

Random forest is an ensemble learning method where it generates multiple decision trees with some randomness in each one of them. It may use different subset of features for building decision trees. Here, We have used 'RandomForestClassifier()' from 'sklearn.ensemble' library in python to build model for the dataset.

Number of estimators which gave best result were between 36 and 45. For this system, number of estimators were taken as 42. Here again, Gini index was taken as criterion for selecting attributes. Maximum depth for each tree was set to default value, i.e. until all leaves of the tree are pure. Maximum number of features to be considered while building a tree in random forest was set to 'auto', i.e. $\sqrt{\text{Total number of features}}$.

C. Voting Classifier:

The Voting Classifier for combining machine learning classifiers that are conceptually different and makes use of the average predicted probabilities or majority vote for prediction of the class labels. Here, 'VotingClassifier()' is used from 'sklearn.ensemble' library in python to build a classifier. Voting classifier considered three classifiers for soft voting. SVM, Decision tree and Random forest were given as attributes to the algorithm.

3.4 Testing phase:

As an option to classification evaluation model would be to split the dataset into a smaller training set and a validation set, then train your models against the smaller training set and evaluate them against the validation set. In this, the accuracy of the model is checked using test dataset. Machine learning algorithm providing the highest accuracy for the system is then selected for the final prediction in the real world.

Following figures show the accuracy of the classifiers and other performance measures like precision, recall, F-score:

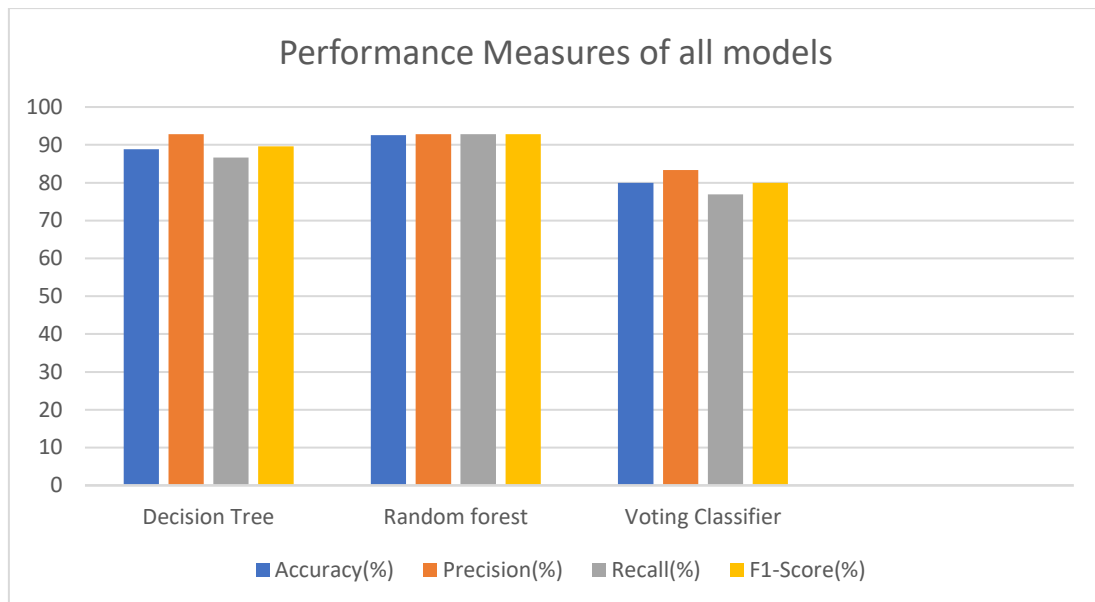


Figure 4. Performance measures for Classifiers (Graph)

Table 1. Performance measures for Classifiers (Tabular)

Classifier	Accuracy	Precision	Recall	F1-Score
Decision Tree	88.88%	92.85%	86.66%	89.64%
Random forest	92.59%	92.85%	92.85%	92.85%
Voting Classifier	80%	83.33%	76.92%	79.99%

The basic measure for evaluating the performance of any classifier model is Accuracy. It gives the ratio of correctly identified observations to the total observations. In our system it was found that Random forest classifier predicted the most accurate results with accuracy of 92.59% followed by Decision tree with 88.88% accuracy and voting classifier with 80% accuracy.

Precision is the other performance measure which identifies the correctly predicted positive comments from total positive observations. In our system, precision gives ratio of correctly predicted observations having learning disability to total number of observations having learning disability. We got maximum precision of 92.85% in Random forest as well as Decision tree followed by 83.33% in voting classifier.

Sensitivity or Recall gives the proportion of correctly predicted positives from all the records. In our system, recall gives ratio of number of correct prediction of records having learning disability to the total number of records having learning disability. We have got pretty good recall as well with highest of 92.85 in Random forest and lowest of 76.92 in Voting classifier. Decision tree showed 86.66% Sensitivity.

F1-Score gives weighted average of precision and recall. In testing, it was found that F1-score of decision tree was 89.64%. F1-score of random forest and voting classifier were 92.85 and 79.99% respectively.

3.5 Prediction:

In this last phase, the unknown sample is provided to the system without output class. The output class 0 (Normal) or 1(Learning disability) is predicted by the system for real world.

3.6 Comparative Result Analysis:

Based on our research, there are very few studies conducted on the identification of learning disabilities through digital handwriting analysis. One such study is conducted by Sara Rosenblum and Gideon Dror. They have used Hebrew handwritten samples for identification of developmental dysgraphia. The details of their work are explained in following table:

Table 2. Details about previous work

Author Names	Publication details	Database size	Algorithm	Accuracy	Sensitivity/ Recall
Sara Rosenblum and Gideon Dror	IEEE Transactions	99 (50 proficient, 49 non-proficient)	SVM	89.9%	90%
S.V. Kedar, PurvaParab, Akash Sharma, JaikrishnaPatil, RohitWagh	-	60	Random forest	92.59%	92.85%

Moreover, we have conducted our research on English handwritten samples instead of Hebrew. The fact that we have got 92.59% accuracy through Random forest classifier better than 89.9% through SVM is success of our research work. We have also got better Sensitivity of 92.85% by the same algorithm.

4. Problem Model

1. Segmentation:

The Pen's status is captured using the pressure field in csv file. If pen tip is in in-air the field has zero value and if it's on-paper, the field shows positive value. Using this field, we divide the writing/drawing into set of segments in which the the pen tip is either continuously in-air or on-paper. Here, we have used on-paper strokes only to extract the features of handwriting. Dividing transactory into continuous On-Paper strokes helps to identify common patters of curve and improves accuracy of results.

2. Feature Extraction:

After segmentation we extract stoke level features using CSV file generated for each task. To get best results we omit segments of shorter length. By doing this we also remove data points which are not actual part of writing/drawing (outlier). Below are the equations used to calculate above mentioned features.

A. On-paper segments(F1):

Let d_1, d_2, \dots, d_n be the set of data frames(records) obtained from CSV file from the tablet such that each record has pressure greater than 0. Then number of on-paper segments(F1) can be calculated as,

$$On_paper_segments = \sum_{i=1}^n 1, \text{ where } p_i > 0$$

B. In-air segments(F2):

Let d_1, d_2, \dots, d_n be the set of data frames(records) obtained from CSV file from the tablet such that each record has pressure equal to 0. Then number of on-paper segments(F1) can be calculated as,

$$In_air_segments = \sum_{i=1}^n 1, \text{ where } p_i = 0$$

C. Total time taken for entire task(F3):

Let d_1, d_2, \dots, d_n be the set of data frames(records) obtained from CSV file from the tablet. Then total time required to complete a task(F1) can be calculated as,

$$dur = d_n(t) - d_1(t)$$

where n is the whole number of data points and $d_i(t)$ is time-stamp recorded for data frame i .

D. Total time spent in air(F4):

Let d_1, d_2, \dots, d_n be the set of data frames(records) obtained from CSV file from the tablet such that each record has pressure equal to 0. Then total time required to complete a task(F1) can be calculated as,

$$In_air_time = \sum_{i=1}^n d_{i+1}(t) - d_i(t), \text{ where } p_i = 0$$

where n is the whole number of data points and $d_i(t)$ is time-stamp recorded for data frame i .

E. Total time spent on paper(F5):

Let d_1, d_2, \dots, d_n be the set of data frames(records) obtained from CSV file from the tablet such that each record has pressure greater than 0. Then total time required to complete a task(F1) can be calculated as,

$$On_paper_time = \sum_{i=1}^n d_{i+1}(t) - d_i(t), \text{ where } p_i > 0$$

where n is the whole number of data points and $d_i(t)$ is time-stamp recorded for data frame i .

F. Mean Pressure(F6):

Let p_1, p_2, \dots, p_n be the set of data frames for which applied pressure is positive. Then average pressure can be calculated as,

$$p_{mean} = \sum_{i=1}^n (p_i(p)/n)$$

where $p_i(p)$ is pressure for data frame i . and standard deviation(F7) in pressure(F6) can be derived using below equation

$$psd = \sqrt{\sum_{i=1}^n (p_i(p) - p_{mean})^2}$$

G. Distance(F8):

The distance travelled between adjacent data frames can be calculated as,

$$pi(d) = \sqrt{(pi(x) - p_{i+1}(x))^2 + (pi(y) - p_{i+1}(y))^2}$$

The average speed(F9) can be calculated by dividing total distance with total time.

$$savg = \sum_{i=1}^n (pi(d)/dur)$$

5. Conclusion

This paper describes a generalized system for identification of learning disability through digital handwriting analysis of the subject. We have taken into consideration 3 machine learning algorithms viz. Random forest classifier, Decision tree, and Voting classifier for classification and correct prediction of subjects having learning disability from the others. We found that handwriting features related to motion, time and pressure are helpful for diagnosis of learning disability. In the actual system it was later understood that features like in air time of a pen, on paper time of a pen and standard deviation in pressure were the most important features for correct prediction of classes and thus increased the accuracy of the model. Random forest classifier showed the highest accuracy of 92.59% while testing the model. The results can further be improved by increasing the volume of dataset.

6. Acknowledgment

We extend this effort for expressing our gratitude to all those who have rendered cooperation and direction that supported for this research. We deeply thankful and express our deepest gratitude to teachers, guides and mentors who helped us to make this study happen. We are thankful to the authorities of SavitribaiPhule Pune University for their constant strategy and support.

References

1. Daniela Giordano and Francesco Maiorana, "Addressing Dysgraphia with a Mobile, Web-based Software with Interactive Feedback", IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, pp. 264-268, 2014.
2. Nan-Ying Yua and Shao-Hsia Changb, "Characterization of the fine motor problems in patients with cognitive dysfunction– A computerized handwriting analysis", Elsevier Human Movement Science, pp. 1-9, 2018.
3. Manuela Gallia, Veronica Cimolina, Giacomo Stellab, Maria Francesca De Pandisc, Andrea Ancillaod and Claudia Condolucie, "Quantitative assessment of drawing tests in children with dyslexia and dysgraphia", Elsevier Human Movement Science, pp. 1-9, 2018.
4. Christian Kahindo, Mounim A. El-Yacoubi, Sonia Garcia-Salicetti, Anne-Sophie Rigaud, and Victoria Cristancho-Lacroix, "Characterizing Early-Stage Alzheimer Through Spatiotemporal Dynamics of Handwriting", IEEE Transactions on Content Mining, pp.1-5, 2017.
5. Shao-Hsia Chang and Nan-Ying Yu, "Research in developmental disabilities - Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics", Research in Developmental Disabilities, Volume 34, Issue 9, pp.2433-2441, 2013.
6. Cecilia W. P. LiTsang¹, Tim M. H. Li¹, Choco H. Y. Ho, Mandy S. W. Lau and Howard W. H. Leung, "The Relationship Between Sensorimotor and Handwriting Performance in Chinese Adolescents with Autism Spectrum Disorder", Springer Journal of Autism and Developmental Disorders, pp.3093-3100, 2018.
7. Mona S. Julius, Rivka Meir, Zivit Shechter-Nissim and Esther Adi-Japha, "Children's ability to learn a motor skill is related to handwriting and reading proficiency", Elsevier - Learning and Individual Differences, pp.1-8, 2017.
8. Jessica Guilberta, Denis Alamargota and Marie-France Morin, "Handwriting on a tablet screen: Role of visual and proprioceptive feedback in the control of movement by children and adults", Elsevier Human Movement Science, Volume 65, Pages 30-41, 2019.
9. Jiri Mekyska, Marcos Faundez-Zanuy, Zdenek Mzourek, Zoltan Galaz, Zdenek Smekal and Sara Rosenblum, "Identification and Rating of Developmental Dysgraphia by Handwriting Analysis", IEEE Transactions on Human-Machine Systems, pp.1-14, 2016.
10. Prof. S. V. Kedar and Dr. D. S. Bormane, "Automatic Emotion recognition", International Journal of Innovations & Advancement in Computer Science, pp. 49-56, 2016.

11. Christina T. Fuentes, BS, Stewart H. Mostofsky, MD, and Amy J. Bastian, "PhD. Children with autism show specific handwriting impairments", AAN Enterprises, Inc., pp.1532-1537, 2009.
12. Rosenblum S. Dyslexia and dysgraphia, "Procedural learning and the cerebellum by Angela Faucet, Roderick Nicolson. Development, Reliability, and Validity of the Handwriting Proficiency Screening Questionnaire (HPSQ)", Elsevier Cortex, Volume 47, Issue 1, pp.117-127, 2011.
13. Rosenblum S, Dvorkin AY and Weiss PL, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting", Elsevier Human Movement Science 25, pp.608-621, 2006.
14. Bahlmann, B. Haasdonk and H. Burkhardt, "Online handwriting recognition with support vector machines - a kernel approach", Eighth International Workshop on Frontiers in Handwriting Recognition, Niagara on the Lake, Ontario, Canada, pp. 49-54, 2002.
15. AzadehKushki, Heidi Schwellnus, FaizahIlyas and Tom Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia", Elsevier Research in Developmental Disabilities 32, pp.1058-1064, 2011.

Authors Profile



Dr. Seema V. Kedar is working as Head of Computer Department in JSPM's Rajarshi Shahu College of Engineering, Pune. More than 25 papers published in various National and International Journals; published various books on database and Object Oriented Programming; Published 7 Patents and 4 Copyrights.
Purva P.



Parabis working in Xoriant, Pune and completed Bachelor of Engineering in Computer Engineering from JSPM's Rajarshi Shahu College of Engineering, Pune.



Akash R. Sharma is working in Yardi Software, Pune and completed Bachelor of Engineering in Computer Engineering from JSPM's Rajarshi Shahu College of Engineering, Pune.



Jaikrishna M. Patil doing MSc in Computer Science from TU Kaiserslautern, Germany. He is completed Bachelor of Engineering in Computer Engineering from JSPM's Rajarshi Shahu College of Engineering, Pune.



Rohit T. Wagh is working in TCS, Pune and completed Bachelor of Engineering in Computer Engineering from JSPM's Rajarshi Shahu College of Engineering, Pune.