

Comparison of Classification Methods used in Machine Learning for Dysgraphia Identification

Sarthika Dutt¹, Neelu Jyothi Ahuja²

^{1,2} University of Petroleum and Energy Studies Dehradun, India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract. Dysgraphia is a disorder that affects writing skills. Dysgraphia Identification at an early age of a child's development is a difficult task. It can be identified using problematic skills associated with Dysgraphia difficulty. In this study motor ability, space knowledge, copying skill, Visual Spatial Response are some of the features included for Dysgraphia identification. The features that affect Dysgraphia disability are analyzed using a feature selection technique EN (Elastic Net). The significant features are classified using machine learning techniques. The classification models compared are KNN (K-Nearest Neighbors), Naïve Bayes, Decision tree, Random Forest, SVM (Support Vector Machine) on the Dysgraphia dataset. Results indicate the highest performance of the Random forest classification model for Dysgraphia identification.

Keywords: Dysgraphia, Learning difficulties, Machine Learning, Motor ability

1 Introduction

Dysgraphia is a learning difficulty caused by a neurobiological disorder, Dysgraphia symptoms can be analyzed through problematic skills related to Dysgraphia difficulty. Dysgraphia in some cases can be due to Dyslexia when a learner is incapable to pronounce properly and write incorrect spellings. [1] Dysgraphia in learners can show symptoms of weak motor skills, space knowledge, or incorrect spellings. [2] [3]

Visual-motor integration is required to perform tasks related to writing.[4] Weakness in Motor skills affects the movements of hands, which as a result cause illegibility in writing, [5] lack of Motor skills indirectly adds to the Dysgraphia symptoms. Another Reason that causes Dysgraphia symptom is lack of space knowledge. [3], [6] When the learner writes improperly out of line maximum time, this indicates the learner has no space knowledge. The lack of cognition can result in Dyslexia and this indirectly contributes to Dysgraphia difficulties. [7]

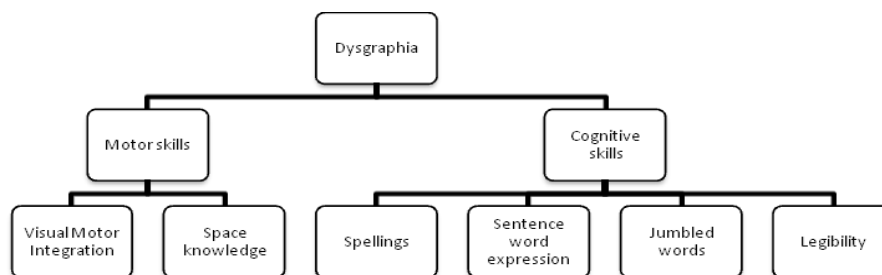


Figure 1. Symptoms related to Dysgraphia difficulty

Figure 1 explains the skills that are related to Dysgraphia identification. Dysgraphia disorder can be determined by analyzing skills related to motor ability and cognitive skills. Activities that are focused on visual-motor integration skills can be used to detect motor capabilities. Space knowledge can also be used to determine motor skills. In case, if Dysgraphia is due to Dyslexia cognitive skills can be observed through spelling, sentence word expression, jumbled words, and handwriting legibility. [21] [23] Learners are different and can show single or combine symptoms of Dysgraphia. [6] Therefore, it is important to analyze which features highly contribute in Dysgraphia identification. These features provide learner characteristics and behavior to identify Dysgraphia. Dysgraphia identification for providing learner specific environment in learning is important. [28]

The feature selection method Elastic Net has been used on the Dysgraphia dataset for feature selection. The proposed study solved Dysgraphia identification problem by using classification algorithms. Classification models are compared and the model with high performance is selected for Dysgraphia identification. So, The main focus of this work is to select features that can be used to identify Dysgraphia Difficulty. Secondly, all

these factors will be analyzed and highly contributing features will be selected using the feature selection method EN(Elastic Net). Thirdly, these features will be trained on KNN, Linear regression, Naïve Bayes, decision tree, and Random forest models. The model with the highest performance will be used for Dysgraphia identification.

2 Related Work

Learning difficulties and its severity in India has been discussed in many studies. Learning Difficulty specifically, Dysgraphia is a disorder that makes it difficult for learners to write as normal learners [8] Dysgraphia is a neurobiological disorder and causes the learner to have poor writing skills. This disorder can be a result of weak motor skills, no space knowledge, or low cognitive skills.[9][21]

The development of motor skills is required at elementary school for better handwriting skills. [10] In early childhood, motor skills development using dot-connecting exercise help to improve the underdeveloped muscular activities. [5] Space knowledge of learners with Dysgraphia is a problematic skill, it results in illegible handwriting. [10] the weak visual-spatial response also makes writing difficult, and learners often get confused with left-right direction. [11] Sentence structure and word expression are other problematic skills that contribute to Dysgraphia symptoms. Developed motor skills can improve the writing capabilities of a learner. How motor skills affect Dysgraphia severity is not explored in previous studies.[12]

Data mining has been used to understand the fact and plan actions with that data. [14]. The data analysis process will generate a clear idea about the fact related to the data and correlation among data present. This will help in maintaining the effectiveness of the data for prediction. Abundant data of scholarly need to be visualized properly for fact generation from data. Feature selection has been extensively used for data knowledge and to understand the relationship among features. [14] [15] [24] [25] The data of learners with Dysgraphia along with normal learners data need to be properly explored before using machine learning and deep learning methods.

Dyslexia, Dysgraphia, and Dyscalculia Identification has been solved using the fuzzy k-mean clustering approach. [13] The deep learning approach is used for solving writing difficulties, cognitive disabilities, [16], and spoken language understanding [17]. Machine learning models SVM, KNN, and Random Forest are compared for Dyslexia, Dysgraphia, and Dyscalculia prediction. The input used in these Machine Learning models is all extracted from the game based screening of these disabilities. [19] Machine learning models SVM, KNN results are evaluated and compared on accuracy performance metrics, these models with high accuracy are then used in ensemble machine learning models for final results. The performance of each machine learning model is individually compared with the ensemble machine learning model. [18] The author in this study has proposed the identification of Dyslexia, Dysgraphia, and Dyscalculia through a mobile application. Handwriting samples and audio samples are analyzed for prediction. [20] Dysgraphia has been predicted using 52 extracted handwriting attributes (velocity, acceleration, etc...). PCA has been used to visualize attributes from handwriting.[21] Dysgraphia is predicted based machine learning model for third-grade children developmental Dysgraphia prediction. Pen pressure, pen position, and pen lifts are taken as input by using a digital writing pad, input in the machine learning model is given from 99 samples for prediction. [22]

3 Methodology

2.1 Dataset

The Data of 240 learners have been used for analysis 142 were learning disabled. Out of 142 learners only 45 learners were having Dysgraphia problem, 18 learners were with Dyslexia and Dysgraphia problem, 14 with Dysgraphia and Dyscalculia, 7 with all Dyslexia, Dysgraphia and Dyscalculia symptoms, 36 with only Dyslexia difficulty, 21 with Dyscalculia difficulty.

The dataset used for Dysgraphia prediction includes data related to dot-connecting exercise to analyze the motor skills of learners with Dysgraphia. Writing samples to analyze the legibility and space knowledge of learners. Other parameters for dysgraphia prediction are based on skills related to sentence structure, sentence word expression, visual-motor integration, and visual-spatial relation. The Pretest and handwritten content were taken as input in form of a questionnaire in a computer-based test. Score and time data of learners with a learning disability and non-learning disability of 240 learners are analyzed. Boolean input is taken through sentence structure, word formation, and visual-spatial response, handwritten content is analyzed using image processing technique Structural Similarity Index Measure (SSIM), spellings check through a spelling checker. Our previous study explains the extraction of features, subtype, and their mapping with Dysgraphia difficulties in detail. [23]

3.2 Feature Selection and Classification

Features have been selected using EN, feature selection technique. This technique is selected based on its performance in previous researches. [24] –[26] The most important features are selected and are trained and tested over classification models KNN, Naïve Bayes, Decision tree, Random Forest, SVM. Some of these classification models are used for Dysgraphia prediction in previous studies. [18] [19]

The 80 % of data is used in this study as training data and the rest 20% as the testing data. These five classification models are compared for their performance, accuracy, and AUC/ROC curve is used as the performance metrics used for comparing these models. Equation 1 represents the accuracy, TP- True Positive and are the instances which have been correctly predicted positive instance, TN –True Negative are the instances which are correctly predicted negative instances, FP- False Positive is the instance which are incorrectly predicted positive instance, FN- False Negative is the instance which is incorrectly predicted negative instance

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \tag{1}$$

AUC curve is another performance metric used in this study to compare different machine learning models on the Dysgraphia dataset. The equation used for the AUC curve is given as Equation 2, 3, 4, 5. True positive rate (TPR) is the used to measure actual positive instances which are predicted correctly. False Negative Rate (FNR) is used to measure actual positive instances that are predicted correctly. True Negative Rate (TNR), to measure the actual positive instances which are predicted incorrectly. False Positive Rate (FPR) , is used to determine the actual negative instances which are predicted incorrectly. [27]

$$Sensitivity/True\ Positive\ Rate = \frac{TP}{TP + FP} \tag{2}$$

$$FNR = \frac{FN}{TP + FN} \tag{3}$$

$$Specificity/True\ Negative\ Rate = \frac{TN}{TN + FP} \tag{4}$$

$$FPR = \frac{FP}{TN + FP} = 1 - specificity \tag{5}$$

Feature selection and all classification models are implemented using scikit learn package in Python (version-3.6).

4 Result and Discussion

Feature Selection is the only way to cut through a dataset, Here it is identified to pick data points like literacy skill, Phonological Awareness, Visual-Spatial Relation, Visual-Motor Integration, Spellings, Handwriting Legibility, Short /Long term memory. Still, they are more than 20 data points from single datapoints like literacy skill, phonological awareness, sentence word expression, word formation, addition, subtraction, reasoning, place value, direction, rhyming, basic mathematic skills, word problem, decoding, random naming, and spellings to derived data points such as visual motor integration, handwriting legibility, reading analysis.

Feature Extraction, feature scaling, feature transformation, are the techniques for improving the accuracy of a data-based model many techniques. Starting from feature selection, there are various Feature selection techniques used for feature analysis. In this study Elastic Net (EN) is used, It is one of the effective feature selection method used in data exploration. The coefficient values of the features analyzed are depicted in Figure 2. The highly contributing factors for Dysgraphia prediction are found to be Legibility, motor skills (VMI_MS), Visual-spatial response (VSR_LR_1), basic Reading skills, Literacy skills (LS_LI), and spellings. It has been revealed by the feature selection process that motor skill and space knowledge are majorly contributing features with high significance for Dysgraphia identification. The dimensionality of dataset is reduced by 15.97% using EN feature selection method. Sixteen Features are selected after dimensionality reduction of the dataset. These Sixteen Features are sorted from lowest to highest significance in Dysgraphia identification in Figure 2. These sixteen features are trained using classification models KNN, Naïve Bayes, Decision tree, Random Forest, SVM.

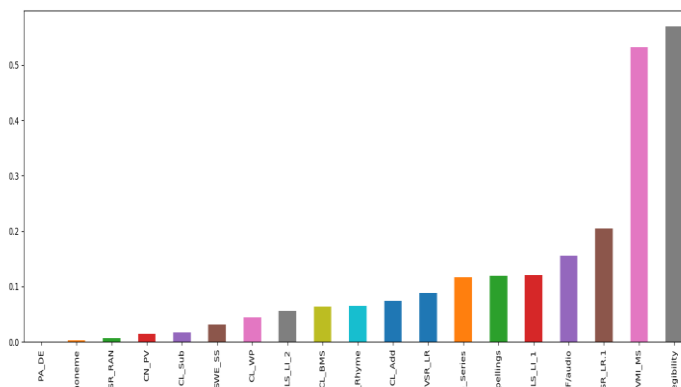


Figure 2. Features significance using Elastic Net

Table 1 represents the accuracy of KNN, Naïve Bayes, Decision tree, Random Forest, SVM models the Random Forest has been found to have the highest accuracy. The accuracy is influenced when these models are trained after using feature selection methods. This shows how feature selection can improve the overall accuracy of the trained models. The Accuracy and AUC/ROC value when compared after using EN Feature Selection method improved significantly.

models metrics	KNN	NB	DT	RF	SVM
Accuracy	97.21%	90.00%	97.00%	99.00%	89.23%
ROC (area)	0.97	0.90	0.97	0.99	0.91
Accuracy with (EN FS)	99.00%	91.58%	99.00%	99.03%	91.00%
ROC (area) with (EN FS)	0.99	0.91	0.99	0.99	0.91

The accuracy of all classification methods has increased by some decimal points when only selected features are trained over all Machine Learning models. Feature selection before implementing a classification model hence proved to improve overall efficacy of classification models. The accuracy of Random Forest, KNN, and SVM are comparatively the same with 99.03%, 99.00%, and 99.00% accuracy score. Naïve Bayes and SVM yielded 91.58% and 91.00% accuracy score.

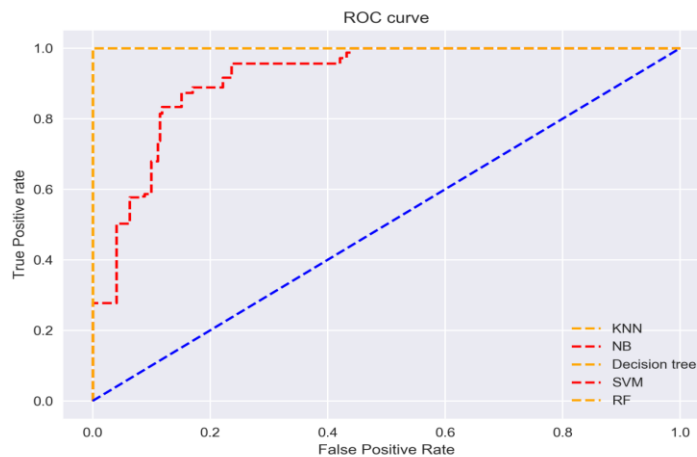


Figure 3. AUC/ROC Curve of classifier models

It is evident from figure 3 that AUC for the KNN, Decision tree, and RF ROC curve is higher than that for the ROC curve of the Decision tree and SVM. The performance of KNN, NB, and RF is found to be comparatively the same for Dysgraphia prediction, NB and SVM performance yielded the low performance when compared with KNN, Decision tree, and RF classification methods.

The results indicate that the performance of most classification algorithms is comparatively same on dataset for dysgraphia identification. KNN, Decision tree and Random Forest performance was significant in predicting the correct output when compared with performance of Naïve Bayes and SVM classification algorithm. When, the models were trained on selected features with high significance using Feature selection method Elastic Net. It has slightly influenced the accuracy metrics of classification models. The limitation of this study is that no deep learning approach is discussed and compared in this study. The deep learning techniques on the same dataset need to be compared with ML techniques. This comparison can provide more general idea on how deep learning

is influencing the accuracies on same dataset and selected features when compared with accuracies of Machine Learning models. Moreover, Dysgraphia dataset can be improved by integrating Cnn for extracting data of learners with Dysgraphia.

5 Conclusion

In this study, Dysgraphia disability has been analyzed using the Feature selection method EN. It has been analyzed in the proposed study that Motor skills affect the Dysgraphia severity and contribute majorly to developing Dysgraphia from the initial years of child development. Other features such as cognition and space knowledge have also contributed to Dysgraphia severity in children. These models which have been used previously are compared for Dysgraphia identification. Random forest yielded the highest accuracy for Dysgraphia prediction with 99.03% accuracy. The dataset can be enhanced by integrating IOT devices to get real time data. Also, classification process needs to be improved by comparing deep learning approaches in further studies.

Acknowledgement

This research work has been carried out at University of Petroleum and Energy Studies (UPES) with Project No SEED/TIDE/133/2016. The authors gratefully acknowledge the funding support received from Technology Interventions for Disabled and Elderly (TIDE) scheme under the Department of Science and Technology (DST). The authors express their gratitude towards the management of UPES for their support in research work.

Conflict of interest

Authors have no conflict of interest to declare

References

1. Hebert, M., Kearns, D. M., Hayes, J. B., Bazis, P., & Cooper, S. Why Children With
2. Dyslexia Struggle With Writing and How to Help Them. *Language, speech, and hearing services in schools*, 49(4), 843–863 (2018) https://doi.org/10.1044/2018_LSHSS-DYSLC-4.18-0024,
3. 18-0024,
4. Biotteau, M., Danna, J., Baudou, É., Puyjarinet, F., Velay, J. L., Albaret, J. M., & Chaix, Y.
5. Developmental coordination disorder and dysgraphia: signs and symptoms, diagnosis, and
6. rehabilitation. *Neuropsychiatric disease and treatment*, 15, 1873–1885 (2019)
7. <https://doi.org/10.2147/NDT.S120514>
8. Van der Gon Denier, J.J., Thuring, J. P., "The guiding of human writing movements,"
9. *Kybernetik*, vol. 2, no. 4, pp. 145-148, (1965).
10. Schneck, C. M. Visual perception. In J. Case-Smith (Ed.), *Occupational therapy for*
11. *children* (4th ed., pp. 382e412). Sydney, New South Wales, Australia: Mosby (2001)
12. Schweltnus, H., Carnahan, H. et. al., Effect of Pencil Grasp on the Speed and Legibility of
13. Handwriting in Children. *The American journal of occupational therapy: official*
14. *publication of the American Occupational Therapy Association*, 66 (2012)
15. Gerth, S. Dolk, T., et al., "Adapting to the surface: a comparison of handwriting measures
16. when writing on a tablet computer and on a paper," *Human Movement Science*, vol.48, pp.
17. 62-73 (2016)
18. Piek, J.P., Dawson, L., Smith, L.M., Gasson, N., The role of early fine and gross motor
19. (2008)
20. American Psychiatric Association . *Diagnostic and Statistical Manual of Mental*
21. *Disorders*. 4th ed. Washington DC: American Psychiatric Association; (2000)
22. Linda S Siegel, "Perspective on dyslexia". *Paediatr Child Health*, Vol 11 No 9, Nov (2006)
23. Joseph Psotka, Sharon A. Mutter, *Intelligent Tutoring Systems: Lesson Learned*. Lawrence
24. Erlbaum Associates. ISBN 978-0-8058-0192-7 (1988).
25. Döhla, D., & Heim, S. Developmental Dyslexia and Dysgraphia: What can We
26. Learn from the One About the Other?. *Frontiers in psychology*, 6, 2045 (2016)
27. <https://doi.org/10.3389/fpsyg.2015.02045>
28. Wu, Tung-Kuang & Meng, Ying-Ru & Huang, Shian-Chang. Application of
29. Artificial Neural Network to the Identification of Students with Learning Disabilities.
30. *Proceedings of the 2006 International Conference on Artificial Intelligence, ICAI'06*. 1.
31. 162-168 (2006)
32. C.Tsai, R., N.Lin, K., et al, Evaluating the users of the total score and the domain
33. scores in the cognitive abilities screening Instrument. Chinese Version (CSAI C-2.0):
34. results of confirmatory factor analysis. *International psychogeriatrics* , 19:6, 1051-1063
35. C_2007 International Psychogeriatric Association (2007) doi:10.1017/S1041610207005327
36. Sadiku, Matthew & Shadare, Adebowale & Musa, Sarhan & Akujuobi, Cajetan & Perry,
37. Roy. DATA VISUALIZATION. *International Journal of Engineering Research and*
38. *Advanced Technology (IJERAT)*. 12. 2454-6135 (2016)
39. Liu, J., Tang, T, et al, A Survey of Scholarly Data Visualization. *IEEE VOLUME 6*,
40. (2018) DOI : 10.1109/ACCESS.2018.2815030.
41. Alex Graves, Marcus Liwicki, Santiago Fernandez, ´ Roman Bertolami, Horst Bunke, and
42. Jurgen ´ Schmidhuber.. A novel connectionist system for unconstrained handwriting
43. recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. (2009)
- 44.

45. Gonzalez-Dominguez, Javier & Lopez-Moreno, I. & Sak, H. & Gonzalez-Rodriguez, J. &
46. Moreno, Pedro. Automatic language identification using Long Short-Term Memory
47. recurrent neural networks. Proceedings of the Annual Conference of the International
48. Speech Communication Association, INTERSPEECH. 2155-2159 (2014)
49. R. O. Mounica, V. Soumya, S. Krovvidi, K. S. Chandrika and R. Gayathri, "A Multi Layer
50. Ensemble Learning Framework for Learning Disability Detection in School-Aged
51. Children," *2019 10th International Conference on Computing, Communication and*
52. *Networking Technologies (ICCCNT)*, Kanpur, India, pp. 1-6, (2019) doi:
53. 10.1109/ICCCNT45670.2019.8944774.
54. R. Kariyawasam, M. Nadeeshani, T. Hamid, I. Subasinghe and P. Ratnayake, "A Gamified
55. Approach for Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia," *2019*
56. *International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka,
57. pp. 156-161 (2019), doi: 10.1109/ICAC49085.2019.9103336.
58. R. Kariyawasam, M. Nadeeshani, T. Hamid, I. Subasinghe, P. Samarasinghe and P.
59. Ratnayake, "Pubudu: Deep Learning Based Screening And Intervention of Dyslexia,
60. Dysgraphia And Dyscalculia," *2019 14th Conference on Industrial and Information*
61. *Systems (ICIIS)*, Kandy, Sri Lanka, pp. 476-481, (2019) doi:
62. 10.1109/ICIIS47346.2019.9063301.
63. Z. Dankovičová, J. Hurtuk and P. Feciľak, "Evaluation of Digitalized Handwriting for
64. Dysgraphia Detection Using Random Forest Classification Method," *2019 IEEE 17th*
65. *International Symposium on Intelligent Systems and Informatics (SISY)*, Subotica, Serbia,
66. pp. 000149-000154, (2019) doi: 10.1109/SISY47553.2019.9111567.
67. S. Rosenblum and G. Dror, "Identifying Developmental Dysgraphia Characteristics
68. Utilizing Handwriting Classification Methods," in *IEEE Transactions on Human-Machine*
69. *Systems*, vol. 47, no. 2, pp. 293-298, (2017), doi: 10.1109/THMS.2016.2628799.
70. Sarthika Dutt, Neelu Jyothi Ahuja. A Novel Approach of Handwriting Analysis for
71. Dysgraphia Type Diagnosis. *International Journal of Advanced Science and*
72. *Technology*, 29(3), 11812 (2020) Retrieved from
73. <http://serisc.org/journals/index.php/IJAST/article/view/29852>
74. A. K. Uysal, "On Two-Stage Feature Selection Methods for Text Classification," in *IEEE*
75. *Access*, vol. 6, pp. 43233-43251 (2018) doi: 10.1109/ACCESS.2018.2863547.
76. Z. Huang, C. Yang, X. Zhou and T. Huang, "A Hybrid Feature Selection Method Based on
77. Binary State Transition Algorithm and ReliefF," in *IEEE Journal of Biomedical and Health*
78. *Informatics*, vol. 23, no. 5, pp. 1888-1898 (2019), doi: 10.1109/JBHI.2018.2872811.
79. X. Zhang, L.-F. Yan, Y.-C. Hu, G. Li, Y. Yang, Y. Han, Y.-Z. Sun, Z.-C. Liu, Q. Tian, and
80. Z.-Y. Han, "Optimizing a machine learning based glioma grading system using multi-
81. parametric MRI histogram and texture features," *Oncotarget*, vol. 8, no. 29, pp. 47816-
82. 47830 (2017)
83. Wang H., Zheng H. True Positive Rate. In: Dubitzky W., Wolkenhauer O., Cho
84. KH., Yokota H. (eds) *Encyclopedia of Systems Biology*. Springer, New York, NY.
85. (2013) https://doi.org/10.1007/978-1-4419-9863-7_255
86. Bisht A., Ahuja N J. "Design and Development of Competency-based
87. Instructional Model for Instruction Delivery for Learning Disabled using Case Based
88. Reasoning", *International Journal of Recent Technology and Engineering (IJRTE)* ISSN:
89. 2277-3878, Volume-8 Issue-6 (2020)