Learning traits and capture mode of learning disability with classification in e-learning for detecting learning disability using machine learning

Masooda Modak
Research Scholar –SP.I.T., Mumbai, Assistant professor–SIES Graduate School of Technology
Corresponding author email: masoodamodak29@gmail.com

Prachi Gharpure
Director SVKM’s NMIMS University, Indore Campus

M Sasikumar
Director, CDAC, Santacruz

Abstract---Learning Disabilities (LD) are a type of disability that affects people of normal or above-average intelligence. The ability to learn is harmed, and this could last a lifetime. Some children may have a single learning problem, while others may have multiple learning disorders that overlap. Learning disability may include disabilities in various areas related to reading, language and mathematics. Learning disabled children are a broad collection of kids who may face challenges in a variety of areas. For example, one child with a learning disability may have major reading challenges, whereas another may have no reading difficulty at all but struggles with written communication. Learning difficulties are developmental abnormalities that commonly appear during the course of a child’s schooling. These limitations cause a considerable gap between an individual’s genuine potential and day-to-day performance. The purpose of this study is to provide a taxonomy of the many learning qualities of LD, as well as the types of characteristics that cater to which learning disabilities, and to identify the modalities in which a particular learning disability can be captured. Based on these characteristics , we design an e-learning system to detect the presence of learning disability using machine learning.

Keywords---learning disability, detection learning disability, machine learning, learning analytics.
Introduction

Learning Disabilities (LD) are a type of disability that affects people of normal or above-average intelligence. The ability to learn is affected and this may be true for one's lifetime. Some children may have a single learning problem, while others may have multiple learning disorders that overlap. Learning disability may include disabilities in various areas related to reading, language and mathematics. Learning Disability (LD) according to the WHO's International Classification of Diseases the LD are known as "Specific developmental disorders of scholastic skills". They could be classified as Dyslexia, Dysgraphia, Dysclaculia, Dyspraxia. For parents and school officials, the challenges linked with LD have always been a source of worry and concern. Children with LD can and can learn successfully and become winners in society with the correct aid at the right time, proper assessment, and remediation. Identifying, diagnosing, and assisting children with learning disabilities is a difficult task. Based on these findings, it is indicated that early detection of learning disabilities in children is critical in identifying and recommending remedial treatments to parents.

With proper recognition and support, the learning disabled can successfully complete the assigned task. Learning problems do not go away with age or training. Despite the manifested learning difficulties, the LD students could develop compensatory strategies or preferences in learning to provide assistance while they are studying. However, even after putting more efforts, when these children are compared to their peers, they show significant difficulties in reading tasks. Surprisingly, not all students with learning disabilities are diagnosed and/or given assistance before beginning their education. As a result, many kids with learning challenges have not been officially identified with LD through a psychometric evaluation technique. As a result, a large number of students may be using an e-learning platform without possessing the necessary learning skills. There are numerous e-learning resources available for learning challenged individuals. However, they do not take into account or adapt to the unique learning qualities of each student. Every learning handicapped person may have a distinct type of learning disability. One dyslexic student, for example, may struggle with letter and number reversals, whereas another dyslexic student struggles to comprehend written language.

Children with LD can and can learn successfully and become winners in society with the correct aid at the right time, proper assessment, and remediation. Identifying, diagnosing, and assisting children with learning disabilities is a difficult task. Based on these findings, it is indicated that early detection of learning disabilities in children is critical in identifying and recommending remedial treatments to parents. Many times, the student is completely unaware that a specific LD symptom applies to him. Students are sometimes hesitant to recognize that they have one or more LD symptoms. The discovery of these learning disabilities has been aided by a number of researchers. This paper presents an overview of the numerous research projects that have been undertaken in this area. Once the disability has been identified, the student will be able to learn according to his or her own learning needs and preferences, which may result in positive outcomes.
**Related Work**

For the detection of LD, there are a variety of formal and informal tests available. The formal detection approach entails the child undergoing a battery of testing, including intellect, language, and achievement exams. The results of such tests are accepted by schools and institutions in order to provide special education to LD children. However, this necessitates numerous sessions with a qualified LD examiner. Parents are hesitant to undergo such testing because it is a time-consuming process. The informal exams, on the other hand, are designed especially for the early detection of LD so that LD children can receive immediate remediation. The informal detection of LD employs various computational techniques for detection of different types of LD (Informal Testing). The author [20] have discussed the various informal testing methods based on Local Characteristic checklist (LCC), Eye movement, Game, Questionnaire, EEG, Test performance.

**Static Questionnaire/Local Characteristics Checklist**

The works in [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] is based on Static Questionnaire/Local Characteristics Checklist. The results of the proposed study [22, 23, 26] suffered from the same flaw: instructors and parents were hesitant to accept he forecast result due to the black box prediction logic and the lack of clarity in the rules or logic of prediction. However, by using NN, [26] was able to recognise 15 out of 55 people as having LD with an accuracy of 80%. [27] also achieved a decent result by using PCA to reduce the attributes from 16 to 7 with 99 percent accuracy. In addition, the work of [28] reduced the number of attributes from 16 to 4 with a classification accuracy of 98.6 percent.

**Test Performance**

The contributions of [32, 33, 34, 35, 36, 37] is towards using the test performance to gather the learning characteristics of the children. The primary drawback in this approach is that the test results are solely utilised as a basis for judging a child, which is extremely biased. However, [34], analysed 240 samples and discovered 80 LD samples. They were also able to obtain a 95.8% accuracy rate by applying LVQ. [37] used the LVQ and SLP algorithms to achieve accuracy of 84 percent and 91.8%, respectively.

**Electroencephalogram (EEG)**

We have the detection of learning disability done by [38, 39, 40] using Electroencephalogram (EEG). The electroencephalogram, or EEG for short, is a tool for monitoring and detecting brain activities. EEG is a recording of brain electric potential oscillations taken from electrodes on the human scalp. The electrodes on the scalp can detect the electrical activity of the brain in response to diverse stimuli. The authors in [39], are using SVM classifiers using EEG inputs. The results were gathered from 50 volunteers between the ages of 24 and 35, 20 of whom were Dyslexic and 30 of whom were habitual readers. The TP for Dyslexics was 78.7%, whereas it was 82.84 percent for ordinary readers. The
author in [40] achieved 90% accuracy with MLP, however the number of samples was relatively small.

**Eye movements**

The work is also done in detecting learning disability using Eye movements[41,42,43,44,45] The authors in [42] has recorded the eye movements of 76 people, primarily fixations and saccades, and has achieved a high level of sensitivity. For stimulus A, the probability of remaining fixated is 80%, while for stimulus C, the optimal saccades have a sensitivity of 70% and specificity of 58%. Another attempt [43] used an eye movement method to capture the eye movements of 76 school pupils. The classification accuracy was 94.6 percent using videooculographic (VOG) and a Fuzzy membership function using Adaptive Neuro- Fuzzy Inference System-PSO. Another good result was reached by[18], in which the samples were given a test to read and eye tracking was performed along with the user preference for font utilising surveys while reading. There were 97 participants in total, 48 of them were diagnosed with dyslexia, with an accuracy rate of 80.18 percent.

**Game Development**

Finally, we have game development also contributing to the informal detection of LD [46,47,48,49,50] The game DytectiveR48 was first tested on 243 people and achieved 83 percent accuracy using Support Vector Machines in a held-out test set with 100 participants. The authors increased the performance by expanding the game’s scope to include HCI measurements, which were covered by 226 performance measures. The model was trained using the SVM classifier, and it has an accuracy of 84.62 percent [50].

The research so far has been focused on the child’s static characteristics. Another work involves a questionnaire that the students must complete. The detection of Learning Disability is done using a static questionnaire or with the involvement of parents in the screening process, according to the research review. There is a reliance on the responses provided to such questionnaires, which may or may not be honest. The reason for this is because the child may be completely unaware that he is having difficulties. He can also be hesitant to embrace this fact. Furthermore, the EEG procedures used to detect LD may not be very pleasant for youngsters to go through. It’s another exam that a child goes through to see if they have LD. Although eye tracking measurements provide good accuracy, relying only on them may not be very useful because not all eyes can be tracked. Also, the eye movement data is merely tracked, which must then be matched to the context, and the results are not straightforward or easy to interpret.

As a result, we require a system that is not reliant on questionnaire responses or parental involvement. We need a system that can recognise a student’s learning behaviour in the context of the setting in which it occurs in order to diagnose a learning deficit. Once the disability has been identified, the student will be able to learn according to his or her specific learning needs and preferences, which may result in improved performance for the LD learner. Thus, we want to create a system where a learner can be given a user interface via an LMS. The content
on the learner’s specified topic will be supplied. The learner’s performance is tracked as he is consuming this content, and the learner’s performance data is saved. The learning analytics engine now analyses the learner’s learning behaviour or pattern, which is based on the performance data. The learner model is revised by the LA engine based on the findings of the analysis. When this learner requests content for a future topic, the improved model is used to provide the content to the learner (more appropriate representation of the learner). If the revised learner model has characteristics similar to a learning disabled (already stored profile for a learning disabled), the system can diagnose the learning disability. The next section is the framework of the proposed model for detecting a learning disability.

**Framework for Detection of Learning Disability**

The following diagram depicts the system design for detecting a learning disability.

![System Framework to detect Learning Disability](image)

Figure 1. System Framework to detect Learning Disability

As shown in figure 1, the learner accesses the system through a GUI, provided through a LMS. The user may request the system for a learning content of a particular topic. The content selector according to the characteristics of the learner will choose or select the contents and will present it to the learner. The learner could then start learning from the content. As the learner will be learning, the performance monitor will capture the performance of the learner. The performance data could be in the form of time to study, click streams, assignment scores, his eye gaze, his facial expressions, etc. The performance data could then be used by the learning analytics engine to analyze and revise the learner model. The learner model is continuously updated for the changing learning requirements of the learner and according to the characteristics present in the learner model diagnoses the presence of a learning disability. The description and function of the various modules are explained below.

**Learning Management System**

The LMS component will allow conception of easily reached learning environment for children that can increase their learning capabilities. The LMS aims to create
learning environments which is centered on students as learners. Using LMS the students could learn more from what they actually do instead of following the instructions of what they should do. LMS is needed to systematize and administer the content for the learners. The LMS will provide a user interface to the LD learner through which the user can interact, request and retrieve the learning content from the system.

**Adaptive Engine**

Adaptivity can be provided when the information about the LD learners is made available from the learner model. The engine should consider different aspects when providing students with adapted courses. Every LD learner have his/her own way of learning making it necessary to have adaptive navigation support and adaptive presentation. The aim is representation of the content in a way that the LD learner will understand whatever content is presented to him. Another approach is an adaptive sequencing of learning courses/activities for each particular learner based on a learner model updated after each interaction with the learner.

**Learner model**

A learner model is responsible for storing the LD student information. Basically, this model may represent the LD learners’ knowledge, interests, preferences, goals, background, and individual traits during their learning process, allowing for personalized learning and adaptation towards their learning requirements. The learner model will be continuously revised by the analytics engine to map the dynamic learning requirements of the LD learner. Based on their learner profile details and the topic selected by the learners, the system will provide the learning content to the user. If there are any changes in learners’ behavior, the learner profile will be updated immediately.

**Performance tracker**

While learning, the performance monitor will record the various browsing actions such as time spent on learning materials, type of learning content, pass over of learning contents, if any, based on the navigation logs, learning content preferences, etc. If there are any changes in learners’ behavior, the system updates the learner profile immediately. The learners may be assessed based on their test performance. These test performances could be used to identify the knowledge level, learner state, and content complexity. Performance tracker detects LD students’ online behavior via LMS and the LD student profile database is updated continuously. With the help of this performance tracker, we can identify the mindset of LD student at particular time period.

**Learning analytics engine**

The performance data is input to the analytics engine to analyze and learn about their reading difficulties, learning style, and cognitive deficits in order to facilitate suggestion and revision of the learner model. Learning analytics present teachers, parents and students with quick feedback about learning processes of
the students. This feature of LA allows shaping the model of education[26]. Learning analytics can detect learning difficulties in individual learners. Language problems (dysorthographia and dyslexia), arithmetic problems (dyscalculia) and even motor skills problems (dyspraxia) can distort learning achievements. Rapid detection and remediation can make a notable difference; especially when the intervention through learning analytics takes into account all the factors resulting from the assessment[51,52].

**Characteristics of Learning Disability and the Various Modes in Which They can be Captured**

To detect a learning disability, it is necessary to identify the various characteristics that cater to a particular type of learning disability. The first step towards it being to identify the characteristics of learning disability according to age/grade. Table 1 summarizes the characteristics with respect to the age/grade[10,11,12,13].

<table>
<thead>
<tr>
<th>Table 1. Characteristics of Learning disability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-school(age 3 to 5 years)</td>
</tr>
<tr>
<td>1. Pronouncing words is difficult</td>
</tr>
<tr>
<td>2. Rhyming words is difficult</td>
</tr>
<tr>
<td>3. Multiple syllables mispronunciation.</td>
</tr>
<tr>
<td>4. Difficulty in recalling the correct word.</td>
</tr>
<tr>
<td>5. Difficulty in learning the alphabet and numbers.</td>
</tr>
<tr>
<td>6. Recalling the days of the week is not easy.</td>
</tr>
<tr>
<td>7. Difficult identify the colors and shapes.</td>
</tr>
<tr>
<td>8. Writing his or her name is difficult</td>
</tr>
</tbody>
</table>
Once, the various characteristics are studied, the next step is to find the various characteristics of Learning disability that cater to a particular type of LD, that is, reading disability, writing disability, maths difficulty, etc. Table 2 presents the various learner characteristics/features and the corresponding learning difficulty they refer to.

### Table 2: Learner characteristics/features and the Type of Learning Disorder/Disability

<table>
<thead>
<tr>
<th>Reference No.</th>
<th>Learner characteristics/features</th>
<th>Type of Learning Disorder/Disability</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>Difficulty in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Reading decoding</td>
<td>Reading disorders</td>
</tr>
<tr>
<td></td>
<td>2. Reading fluency</td>
<td>Written expression disorders</td>
</tr>
<tr>
<td></td>
<td>3. Reading comprehension</td>
<td>Math disorders</td>
</tr>
<tr>
<td></td>
<td>4. Spelling accuracy</td>
<td>Language disorders</td>
</tr>
<tr>
<td></td>
<td>5. Grammar and punctuation accuracy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Clarity or organization of written Expression</td>
<td>Attention disorders</td>
</tr>
<tr>
<td></td>
<td>7. Number sense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8. Memorization of arithmetic facts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9. Accurate math reasoning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10. Reduced vocabulary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11. Limited sentence structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12. Social pragmatic communication.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13. Combined presentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14. Mainly inattentive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Predominantly hyperactive/impulsive.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>learning to read (word reading that is either inaccurate or slow and laborious)</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>constant need to reread written material to understand its meaning.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Multiple errors in meaning</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Inaccurate or slow counting, persistent difficulties in retrieving.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>number facts, multiple errors in arithmetic calculation, poor mathematical problem solving.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dyslexia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Dyscalculia</td>
</tr>
<tr>
<td>2.</td>
<td>Dyscalculia</td>
</tr>
<tr>
<td>3.</td>
<td>Dyscalculia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dyscalculia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>May make mistakes while reading or writing, reading or recalling numbers.</td>
</tr>
<tr>
<td>2.</td>
<td>Mistakes while adding numbers because of number substitutions or omissions or reversals.</td>
</tr>
<tr>
<td>3.</td>
<td>Unable to recall past or future event plans and sequences.</td>
</tr>
<tr>
<td>4.</td>
<td>Incorrect answers while doing maths operations like add, subtract, multiply or divide.</td>
</tr>
<tr>
<td>5.</td>
<td>Difficulty remembering maths formulas, rules, sequences and procedures.</td>
</tr>
<tr>
<td>6.</td>
<td>Difficulty with left and right orientation.</td>
</tr>
<tr>
<td>7.</td>
<td>Difficulty following steps and directions in solving maths.</td>
</tr>
<tr>
<td>8.</td>
<td>Slow in understanding math concepts in word problems.</td>
</tr>
<tr>
<td>9.</td>
<td>Confuse operations signs or perform them in the wrong order.</td>
</tr>
<tr>
<td>10.</td>
<td>Confuse part to whole relationships.</td>
</tr>
<tr>
<td>11.</td>
<td>Difficulty keeping score during games</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Visual-Spatial Difficulties</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Faces difficulty in reading and organizing words from left to right</td>
</tr>
<tr>
<td>2.</td>
<td>Writes letters spreading in all directions, and letters and words that may jumble together on the page</td>
</tr>
<tr>
<td>3.</td>
<td>Difficult to write on a line and inside margins</td>
</tr>
<tr>
<td>4.</td>
<td>Difficulty in understanding map and drawing.</td>
</tr>
<tr>
<td>5.</td>
<td>Copies text slowly.</td>
</tr>
<tr>
<td>6.</td>
<td>Faces problem in holding a pencil or cutting food, tying shoes.</td>
</tr>
<tr>
<td>7.</td>
<td>Is unable to use scissors well or to color inside the lines.</td>
</tr>
<tr>
<td>8.</td>
<td>Holds his wrist, arm, body or paper in an awkward position when writing.</td>
</tr>
<tr>
<td>9.</td>
<td>Has trouble getting ideas down on paper quickly.</td>
</tr>
<tr>
<td>10.</td>
<td>Has trouble understanding the rules of games.</td>
</tr>
<tr>
<td>11.</td>
<td>Has a hard time following directions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Fine Motor Difficulties</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>Has difficulty in understanding spelling rules</td>
</tr>
<tr>
<td>13.</td>
<td>Has trouble telling if a word is misspelled</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Language Processing Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>Spelling Issues/Handwriting Issues</td>
</tr>
<tr>
<td>15.</td>
<td>Spells words incorrectly and in many different ways</td>
</tr>
<tr>
<td>17.</td>
<td>Blends printing and cursive.</td>
</tr>
<tr>
<td>18.</td>
<td>Has trouble reading his own writing</td>
</tr>
<tr>
<td>19.</td>
<td>Avoids writing.</td>
</tr>
<tr>
<td>20.</td>
<td>Gets a tired or cramped handed when he writes.</td>
</tr>
<tr>
<td>21.</td>
<td>Erases a lot.</td>
</tr>
</tbody>
</table>

**Grammar and Usage Problems**

| 22. | Doesn’t know how to use punctuation. |
| 23. | Overuses commas and mixes up verb tenses. |
| 24. | Doesn’t start sentences with a capital letter. |
| 25. | Doesn’t write in complete sentences but writes in a list format. |

| 26. | Has trouble telling a story and may start in the middle |
| 27. | Leaves out important facts and details, or provides too much information. |
| 28. | Assumes others know what he’s talking about. |
| 29. | Uses vague descriptions. |
| 30. | Writes jumbled sentences. |
| 31. | Never gets to the point, or makes the same point over and over. |
| 32. | Is better at conveying ideas when speaking. |

**Organization of Written Language**

| [5] | Difficulty in both single word decoding and phonology. |
| 1. | Can’t read. |
| 2. | Impairment of both retrieval and storage of speech sounds. |
| 3. | Letters and words switch. |
| 4. | Blurred and wavy words. |
| 5. | Words appear together. |
| 6. | Whole line appears together. |
| 7. | Words jump, vibrate, not clear, move back and forth |
| 8. | Shuffled words. |

**Dyslexia**
1. Poor Vocabulary and Sentence structure.
2. Difficulty listening and following directions, especially multi-step directions in noisy environments
3. Greater difficulty with verbal than nonverbal tasks
4. Misinterpretation of questions
5. Tend to be distractible, especially when background noise is present
6. Have poor social communication skills or difficulty making and/or keeping friends
7. Inability to sing in tune and poor musical ability
8. Difficulty understanding riddles and jokes
9. Poor reading comprehension
10. Trouble sounding out new words and poor fluency when reading aloud
11. Poor spelling skills
12. Confusion or reversal of letters
13. Difficulty remembering people’s names
14. Display poor memory for words and numbers
15. Difficulty with complex language such as word problems
16. Difficulty finding the right words to use when talking
17. Slow or delayed responses to verbal instructions

**Auditory processing disorder**

1. Difficulty with speech sounds (phonemes)
2. Deficiencies in remembering phonemes and manipulating them.
3. Difficult to remember auditory information or follow multiple instructions.

Further, we would need to capture Learning Disability in an e-learning mode. The next step is to check how these learning difficulties can be captured in an e-learning mode. Thus, we have Table 3 discussing the various Learning difficulties, their classification and the different types of assessment questions that could be asked to capture this learning difficulty. This activity is supported by the mentioned citations and also the observations done as part of internship undertaken under a learning centre giving remedial education to students facing learning difficulties.
Table 3: Capture Method of Learning Difficulties in e-learning Mode

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Learning Difficulties</th>
<th>Learning Disability Classification</th>
<th>Reference No.</th>
<th>Example of assessment questions/Method</th>
<th>How the difficulty is captured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Difficult to comprehend</td>
<td>Reading difficulty, Language difficulty, Auditory processing disorder</td>
<td>[1,15,16,17]</td>
<td>Give a short story to read. Mrs. Park’s class wanted to help people in need. Mrs. Park said, “We have to make money. How can we do this?” “I know what we can do,” Ron said. “We can bake cakes and other foods people like. Then, we can have a bake sale.” That is just what the class did. All the children helped make the food. A lot of hungry people came to eat it. Soon, the class had money for people in need. Who was it that the class wanted to help? <a href="Direct">People in need</a> What did Ron say the class could bake? <a href="Direct">Cakes and other foods people like.</a> The story says, “A lot of hungry people came to eat it.” What does hungry mean? [Wanting or needing food]. (Indirect/Inference) Who gave the class the money for the people in need? [The people who paid for food at the bake sale]. (Indirect/Inference) What is the main idea of the passage? [Main Idea]</td>
<td>Captured through Generators of the quiz/MCQ/response /grade/exact string match. Is the child able to answer Direct and Indirect answers from the given passage? Questions to be annotated as to whether they are direct, inference, Vocabulary/Grammar, title, etc.</td>
</tr>
<tr>
<td>2.</td>
<td>Problems with spelling and/or written composition.</td>
<td>Writing Difficulty, Language difficulty</td>
<td>[1,19,15]</td>
<td>Spelling Sight Words: (said, your, was, could, sure, done, who, what, laugh, again) Spelling pseudo words{ trung, sud, splown, quab, bliff, wim, glay, bon, repent, yonk, libbot, cleep, rutch, feme, habble, riade, gratio}</td>
<td>Captured through spellings typed through keyboard and then compared against the actual spellings.</td>
</tr>
<tr>
<td>3.</td>
<td>Difficulties in Sequencing-- Poor</td>
<td>Dyslexia</td>
<td>[18,9]</td>
<td>Q) What comes after Wednesday?? The child may not be able to answer immediately, A generator of sequences is maintained which randomly throws 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>A child might see the math problem as $16 - 3 = ?$ on the blackboard, but write it as $61 - 3 = ?$.</td>
<td>but may need to go back over the whole list, &quot;Sunday, Monday, Tuesday, Wednesday...&quot; before she or he can answer.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Reduced Vocabulary Language Disorder, Auditory Processing Disorder</td>
<td>Requires reading the words and provide synonyms, antonyms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Short term memory(vis/auditory) Dyslexia, Auditory processing disorder</td>
<td>Captured through responses given through the MCQ. Also the time could be recorded for each response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Difficulties in understanding pictorial/verbal information and/or in expressing oneself. Writing disorder</td>
<td>The child reads a phone number and then click on the numbers in order. Click on the days of the week in order.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Difficulties in making connections between words/sentences and the ideas they represent</td>
<td>The child hears an audio (one sentence) and then is asked the sequence of words that he listened in the audio.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Difficulty in interpreting, understanding and recalling auditory/visual information</td>
<td>The child uses his knowledge and ability to relate the meaning of words and their use or context. Captured through responses given through the MCQ. Also the time could be recorded for each response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Difficulty in understanding pictorial/verbal information and/or in expressing oneself.</td>
<td>Match the following: What is your date of birth To find hobbies To check location To calculate age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Difficulty in interpreting, understanding and recalling auditory/visual information</td>
<td>The child hears an audio of the sequence “cow 9 up 3 5”) and is expected to respond with the words first and then the numbers in the order they were dictated. MCQs or written response can be scanned and converted to doc and compare the response to be correct or incorrect.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 9 | Sentence Construction difficulty/ Writing fluency | Language disorder, Writing disorder | [1,7] | Showing a picture of a cake along with the jumbled words, eg,  
  • good cake is  
  • fat pig is  
  Child needs to make a meaningful sentence,  
  As  
  Cake is good Pig is fat.  
  Captured through responses given through the MCQ. Also the time could be recorded for each response |
| 10 | Difficulty with phonemes (phonologic al processing Segmenting and remembering phonemes to identify sounds within words. It is difficult for the child to translate a syllable they have heard into written form. | Reading disorder, Auditory Processing Disorder | [7,18] | The child hears a letter and needs to map it to its visual representation.  
  Orthographic and phonetically similar letters, e.g. b, q, p within a time frame  
  Recognize letters based on their sounds rather than their names.  
  For example, the child hears the phoneme /n/ and then a table is shown containing the target <n> as well as distractors.  
  The child listens to the sound (phoneme) of a letter and identifies it from among the distractors  
  A child needs to change the order of letters to spell a real word  
  to change the order of syllables in order to spell an actual word |
| 11 | Miscue | Writing disorder | [19,18] | These exercises targets child’s phonological awareness at a word level such as  
  Addition of letters, *arrround (around), omission of letters, *empty (empty); substitution of letters, *scholl (school);  
  Omissions. Omitting a word (Tom saw [a] cat.)  
  • Insertions. Inserting additional words (Tom saw the cat.)  
  • Substitutions. Changing order of letters in a word (no for on, was for saw) |
<table>
<thead>
<tr>
<th></th>
<th>Difficulty with left and right orientation, like putting numbers correctly in U,T,H places for the respective operations</th>
<th>Maths difficulty</th>
<th>[3]</th>
<th>Provide a small grid as an interface with places or position to enter the numbers, so as to keep a check if the student has entered the numbers correctly in the required positions, and then can check if the result of operation is correct/incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>Difficulty in calculating squares of small numbers and algebraic formulas</td>
<td>Maths difficulty</td>
<td>[1,2]</td>
<td>Provide a small grid as an interface with places or position to enter the numbers, so as to keep a check if the student has entered the numbers correctly in the required positions, and then can check if the result of operation is correct/incorrect</td>
</tr>
</tbody>
</table>

**Design and Methodology**

The figure shows the design and methodology for the detection system.
As needed, the first development was to

1. Create a e-learning system on Moodle LMS
2. Create courses for the age groups 11-13 years i.e. for grades 6,7 and 8 through the Admin user considering the difficulties faced by Dyslexic students. Each course (of a particular grade) has an English part with quizzes and activities assessing the reading and comprehension capabilities. There is also a Maths section in each of these courses assessing on the basis of mental or logical capabilities along with issues like being able to perform carry operations.
3. Setting up plugins for auto enrollment of users on the basis of the matching preferences, grade from the profile, plugin for providing audio input to the users.
4. Create dedicated user accounts for the students to login and attempt the course.
5. Once, the users attempt the course, their data is generated in the form of their responses, grades and logs.
6. The data generated is extracted by the admin and is used for analysis purposes.
7. The data is then analyzed using python language to convert the raw data into meaningful information.
8. The aforementioned analysis is performed using a set of verified parameters and corresponding threshold values. The verified parameters are as follows:-
   a) Are there any Miscues (Addition, Substitution and Reversal of words) problem.
b) Ability to answer Direct and Indirect Questions in English Comprehension.
c) Ability to describe a picture and a topic in a few sentences.
d) Able to identify Word Relations.
e) Able to solve simple and complex Math problems.
f) Able to perform carry operations.
g) Whether there is presence of short term memory loss.

Algorithm Implementation

The speeches of the users recorded in Moodle can be downloaded only in MP3 format. So, these recordings have to be converted to WAV format before passing them to the speech recognition part. To convert the recording from one audio format to another, we have used the pydub package. Audio file in WAV format is passed as a parameter to the method recognize google() which gives the text output.

Specific ways of comparing two strings is provided in the difflib package. The difflib.ndiff() method gives the difference between two words in terms of the characters needed to be added or subtracted to convert one string to another. To preprocess the text, we have used the nltk() package. There are multiple inbuilt methods available for preprocessing out of which we have used word tokenize() from nltk.tokenize which tokenizes (list of words) a string passed as parameter to this method. Another method of this package used is stopwords from nltk.corpus. In this method, when the parameter ‘english’ is passed to the function stopwords.words(), a list of stopwords in english language is obtained.

Miscue Analysis

Miscue analysis is an analytical procedure for assessing students’ reading comprehension based on samples of oral reading. The miscue analysis is performed on the speech input of a user, obtained through an assessment in Moodle where they have to read aloud a given passage. The recognize google() method gives the output of text converted from the speech in string. To find whether the user has a miscue problem, the following steps are followed:

1. Terms used in Miscue Analysis:
   a. original - list of the words (excluding punctuation) of the actual passage
   b. obtained - list of words of converted text from the speech input of that passage.
   c. count - set for each difficulty to obtain the corresponding no. of mistakes.
   d. current - pointing to the current word in a list.
   e. next - pointing to the next word in a list.

2. Compare both the lists parallelly word by word for identifying mis- matches.
3. If current of original = current of obtained, then
   a. For word substitutions: if next of original = next of obtained, then count +1, else if, next +1 of original = next +1 of obtained, then count +2.
   b. For word additions: if current of original = next of obtained, then count +1, else if, current of original = next +1 of obtained, then count +2.
c. For word omissions: if next of original = current of obtained, then count +1, else if, next +1 of original = current of obtained, then count +2.

b) (d) For letter reversals in word: if 'b','m','n','p, in current of original = 'd','w','u','q' in current of obtained or vice-versa, then count +1.

4. For mispronunciation of a word: Create words - a list of unique words in original
   \[ \forall \text{current in words, No. of occurrences of current in obtained} \]
   \[ \text{No. of occurrences of current in original} > 0.3, \text{then count +1} \]

5. For prefix /suffix addition in a word:
   \[ \text{if current of original} \]
   \[ \in \text{current of obtained, then if current of obtained} \]
   \[ \neg \text{current of original} \in \text{set of few prefixes/suffixes, then count +1.} \]

Here difflib.ndiff() has been used.

Total Mistakes = words substituted + words added + words omitted +
words with letter reversals + mispronounced words + words with pre – fix/
suffix added

i.e add all the count in steps 3 to

\[ \text{Percentage of Mistakes} = \frac{\text{Total Mistakes}}{\text{Total no. of words in the passage}} \]
\[ \text{If Percentage of Mistakes} > \]
30\%, User has miscue problem., else, no miscue problem

Picture Description

The picture description is used to analyze the visual and the audio processing skills of the student which is done by providing a picture of a garden and asking them to describe the image in their own words by recording their speech on the system which is later obtained from Moodle. To understand whether the user has the ability to describe the picture well or not we capture 5 recordings each capturing a particular sentence. Thus have created a corpus or a text file containing 10 possible sentences explaining the picture as accurately and closely as possible and then finding out the similarity between the spoken sentence and the corpus.

We have calculated the similarity by using cosine similarity [22][23] technique, it is a good method for determining semantic similarity between sentences. To determine we perform the following steps they are as follows:

1. Each of the MP3 recordings of the user is given as input as mentioned in the tkinter() section and then converted to text as mentioned in the SpeechRecognition() section.
2. The string obtained is converted to lowercase using lower() and then tokenized and the set of keywords called X Set is created by removing the stop words using nltk()
3. Now, we iterate over the corpus sentences which are stored in a text file. In each iteration, a single sentence is considered, it is converted to lowercase
using lower() and then, stopwords are removed using nltk() and a keywords set is obtained and stored in Y Set.
4. A union of X Set and Y Set called rvector is created and two lists l1 and l2 are initialized.
5. For each keyword in rvector : if keyword in X Set, then append the value 1 in l1 else, append 0 in l1. Similarly we append the values 1 or 0 in l2 if present in Y Set.
6. Thus, lists l1 and l2 are vectors (0,1) of the same size.
7. Counter c is initialized to 0.
8. Now, the cosine formula is used to calculate the similarity between the two vectors i.e. the dot products of the vectors l1 and l2 are stored in a variable c.
9. Then c is divided by the product of square roots of l1 and l2 stored in variable cosine.
10. This process is repeated such that similarity between the recorded sentence and each sentence in the corpus is obtained from variable cosine and is appended in a list called list1.
11. Now, we simply check if the maximum percentage value in list1 is greater than the threshold of 40% then, the recorded sentence is classified as similar to the corpus else, it is classified as not similar.
12. This whole process is repeated for each MP3 recording of the user.
13. If more than 60% (3 out of 5) of the sentences have the cosine similarity of less than 40% with the corpus, then it means that the user has difficulty in describing the picture, else he does not have this difficulty.

**Topic Description**

The topic description mainly helps us understand whether the student is able to frame correct sentences and gauge the understanding of the topic. To find out whether the student has this difficulty or not, we have created an assessment in Moodle where a topic is provided and the students are made to speak 5 sentences on the topic given.

The TXT file or the corpus containing approximately 70 sentences that best describe the topic. A package called wordcloud is used to generate the word cloud. We have made use of the method WordCloud from the package where stopwords removal is performed using STOPWORDS method. The file is passed as a parameter to a function in this method called WordCloud().generate(). The matplotlib package is used to visualize the word cloud as. It also highlights the significant textual data points. Word clouds are widely used for analyzing data from social network websites.

Then we plot the bar graph of words vs frequencies where the top 50 frequent words are plotted as shown in Figure 3.
The MP3 recording from Moodle is converted to WAV format and then provided as an input to the program. It is converted to text, lowercased, tokenized and then a set of keywords is obtained. This is compared with a set of keywords obtained from the corpus and both are converted to vectors of equal length and then cosine similarity is calculated to find out the similarity between the 2 vectors. Cosine Similarity is the dot product of the 2 vectors divided by the magnitude of the vectors. If the Cosine Similarity is higher (closer to 1) then similar else, not similar. This whole process is repeated for each WAV recording of the user. If 3 out of 5 i.e. 60% of the sentences have the cosine similarity closer to 0, then it means that the user has difficulty in describing the topic, else he does not have this difficulty.

**Short Term Memory**

The short term memory loss is assessed by recording the speech input of the user obtained via the courses attempted on Moodle. To determine whether the user has short term memory loss, we provide an audio recording containing a small passage including sequences of specific objects to be memorized. We have created a text file containing the sequence of the objects to be spoken by the user in the order in which it appears in the recording.
The users recording is downloaded from the admins profile and then externally analyzed with our corpus. The corpus is a text file which is initially opened and read line by line. The text file is opened using file pointer in read mode and a FOR loop is run over the text file to get each line. The converted speech to text is then compared exactly with every line in the text file. Then the spoken text is then compared within FOR loop with each and every line obtained and then string matching is performed i.e. exact string matching is performed between the spoken text and text file (corpus). If the percentage of short term memory loss < 100% then we can say that the user cannot retain the sequence i.e. the user has short term memory loss otherwise the user does not have short term memory loss.

**Results and Analysis**

The students of the grades 6, 7 and 8 attempted the lessons through moodle. The students were belonging to the two classes LD and non-LD. The dataset consists of 25 to 30 samples belonging to both classes in each of the three grades 6, 7 and 8 with a total of 80 + students. After dataset creation, the first step was to clean the data. The training and testing tuples are divided into 80:20 ratio. ML algorithms are applied on the dataset which is stored in a dataframe. The dataframe is preprocessed and finally, it is used for building and training the model using algorithms. Here, the dataframe is preprocessed by checking for null values and formatting the data i.e. train-test split in 80:20 ratio and segregating output label. Two separate models are built using Logistic Regression and SVM. Logistic regression is a statistical model that uses a logistic function to model a binary class or multi-class label. It estimates the parameters of a logistic model in binary regression. A binary logistic model (used here) has a dependent variable with two possible classes which is represented by values labeled "0" and "1". SVM algorithm is used for classification, regression and outliers detection. It is a supervised learning algorithm. This algorithm is effective in high dimensional spaces. It is memory efficient because it uses a subset of training points called support vectors in the decision function. Some disadvantages are that it does not directly provide probability estimates, which have to be computed using a five-fold cross-validation.

Building Logistic Regression Model, fitting it on the train-test split and generating predictions.

```python
from sklearn.linear_model import LogisticRegression
1 lr = LogisticRegression()
2 lr.fit(x_train, y_train)
3 predictions = lr.predict(x_test)
```

![Figure 4: Logistic Regression Model](image)

Building SVM Model, fitting it on the train-test split and generating predictions.

```python
from sklearn.svm import SVC
1 svcclassifier = SVC(kernel='linear')
2 svcclassifier.fit(x_train, y_train)
3 y_pred = svcclassifier.predict(x_test)
```

![Figure 5: SVM Model](image)
Testing

Figure 6: Test Tuples

```python
import sklearn.metrics

# Testing tuples

t1 = [[7,5,4,23,5,12,6,4,3,0,1,1,1,0,0,0,1,1,1]]
print("Class Prediction of unknown tuple 't1': ", end=" ")
print(lr.predict(t1))

t2 = [[1,2,1.38,2.34,3.2,2.4,1.5,0,1,0,0,0,0,0,0,1,0,0]]
print("Class Prediction of unknown tuple 't2': ", end=" ")
print(lr.predict(t2))
```

Class Prediction of unknown tuple 't1': [1]
Class Prediction of unknown tuple 't2': [0]

Figure 6 shows the testing of the model using 2 unknown samples t1 and t2. The testing shows the prediction of unknown tuples to be classified as LD(class '1') or non LD(class '0') by the trained model.

Figure 7: Logistic Regression Model Evaluation

```python
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix, accuracy_score

print("Classification Report:
")
print(classification_report(y_test, predictions))
print("Confusion Matrix:")
print(confusion_matrix(y_test, predictions))
print("Accuracy:", end=" ")
print(accuracy_score(y_test, predictions))
```

Classification Report:

<table>
<thead>
<tr>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.83</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>0.80</td>
<td>0.89</td>
</tr>
</tbody>
</table>

avg / total | 0.92   | 0.90    | 0.90    | 10      |

Confusion Matrix:

```python
[[5 0]
 [1 4]]
```

Accuracy: 0.9

Figure 7 shows the performance of Logistic Regression model. It can be seen that the accuracy of the model is 90%.
Figure 8: SVM Model Evaluation

Figure 8 shows the performance of Logistic Regression model. It can be seen that the accuracy of the model is 100%.

The Evaluation Metrics used in both the models are:

- Precision (Specificity): TP / (TP + FP)
- Recall (Sensitivity): TP / (TP + FN)
- F1 Score: 2 * (Recall * Precision) / (Recall + Precision)
- Support: No. of tuples belonging to the corresponding Class
- Accuracy: (TP + TN) / (TP + FP + FN + TN)

On comparing both the models, it is evident that the SVM Model has superior values for these metrics as compared to the Logistic Regression Model but, the values achieved by the SVM model are theoretically ideal values i.e. all the values are 1.00. This means that model has an accuracy of 100% which is not practically achievable and acceptable. This implies the overfitting behavior of the SVM model. Therefore, we would prefer the Logistic Regression model.

The Accuracy of this model is 0.9. This tells that our model gives 90% correct predictions which is acceptable and pretty good for an ML classifier. High Precision indicates FP rate and the model has average precision = 0.92. Recall value over 0.5 is preferred the model produces the recall value of 0.9.

The average of Precision and Recall calculate F1 score and can be considered more helpful than accuracy when there is and an uneven class distribution. The F1 Score of this model is 0.9. The Confusion matrix shows that five tuples are correctly classified into class ‘1’ i.e. LD. There are no FPs which means that no tuples that are incorrectly classified into class ‘1’. There is only one FN which means that one tuple is incorrectly predicted as belonging to class ‘0’. There are four tuples correctly predicted as belonging to the class ‘0’.
Drawbacks of the Conducted Research

The system developed through Moodle has a few assumptions and drawbacks. They are listed as below:

1. The system implemented is conducted using the lessons including picture and topic description. As mentioned, our system has employed a Google translator recognize google() to convert speech to text conversion. So, the correctness of the miscue analysis depends on the accuracy of the recognizer, the speech and volume of the speaker. So, it is very possible that the translated text is incorrect leading to incorrect results.

2. For topic description, even thought the system has including many sentences in the corpus, still it is possible for many correct sentences to identify as not similar. A non LD student may have a good description quality and may frame the sentences using different vocabulary and this may not match to the corpus, giving a not similar statement.

3. For Detecting short term memory, it is possible that the child listens to 5 words and correctly replies with only one word in wrong order; this does not necessarily mean short term memory. It should cover the sequence recalling capability in some practical way.

4. The system tries to understand the learning behavior of the student through the types of lessons the students are attempting, but the lessons and questions are not annotated so as to understand the type of difficulty the child may be facing in attempting the lessons. This may be beneficial to provide assistance to the students.

Conclusion

The existence of Learning Disability may affect the individuals throughout their life. LD usually exists for a lifetime. An LD individual can overcome his difficulties only when they are observed and detected. The major task, thus being the detection of Learning Disability. As a first step towards the detection through any approach, its utmost important to study the characteristics of a learning disabled and then understanding the characteristics that cater to a particular type of learning disability. The study in this paper provides a better understanding of the various characteristics of the LD. We further develop a system that understand the learning behavior of a student in the environment in which its happening so as to detect the learning disability. Once the disability is diagnosed the student will be able to learn according to his/her learning requirements/preferences that may lead to positive performance of the LD learner. The drawbacks of the system are discussed in the previous section regarding the various types of lessons and the way the learning difficulty is captured. Thus we may further need a system to capture the learning difficulties by giving the age appropriate lessons, annotating the various types of questions asked, so that the informal detection of learning disability will allow the student to be able to learn according to his or her own learning needs and preferences, which may result in positive outcomes.
References

16. Auditory proc- disorder).Available at: https://www.auditorycenter.com/symptoms-of-auditory-processing-
disorder-apd/symptoms-of-apd-in-elementary-school-children/[Auditory
proc disorder][Accessed 23 January 2022].


31. R Akhil ; Athul Soori ; Dilip Shankar ; M Mrinal Krishnan ; B R Poorna, Detecting Specific Learning Disabilities (decision tree), 2017 International


