**BRIAN WAVE CLASSIFICATION OF SYSTEM USING**

**MACHINE LEARNING APPROACHES**

A PROJECT REPORT

*Submitted by*

**KALAISELVI.S** **710617106017**

**SHAMILI.I** **710617106036**

**SRIMATHI.M** **710617106037**

*In partial fulfillment for the award*

*of*

**BACHELOR OF ENGINEERING**

IN

**ELECTRONICS AND COMMUNICATION**

**ENGINEERING**

**CSI COLLEGE OF ENGINEERING, KETTI,**

**THE NILGIRIS-643215**

**ANNA UNIVERSITY: CHENNAI 600025**

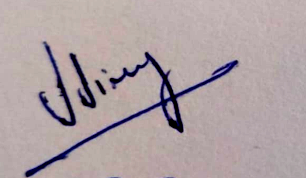
**APRIL 2021**

**ANNA UNIVERSITY: CHENNAI-600025**

**BONAFIDE CERTIFICATION**

Certified that this project report titled “**BRIAN WAVE CLASSIFICATION OF SYSTEM**

**USING MACHINE LEARNING APPROACHES”** is the bonafide work of “**KALAISELVI.S, SHAMILI.I, AND M. SRIMATHI.M”** who carried out the project work under my supervision.

| **SIGNATURE**  Prof.K.Komathy Vanitha,B.E.,M.Tech.,(Ph.D)  **HEAD OF THE DEPARTMENT**  Department of ECE  CSI College of Engineering,  Ketti, The Nilgris-643215 | **SIGNATURE**  Prof. J. Jincy, M.Tech  **SUPERVISOR**  Assistant Professor,  Department of ECE  CSI College of Engineering,  Ketti, The Nilgris-643215 |
| --- | --- |

**SUBMITTED FOR THE ANNA UNIVERSITY PROJECT VIVA-VOCE**

**EXAMINATION HELD ON -08-2021**

SDF.png

**INTERNAL EXAMINER** **EXTERNAL EXAMINER**

II

**ACKNOWLEDGEMENT**

First and foremost, we praise and thank Almighty God, whose blessings have bestowed in us the will power and confidence to carry out our project,

We have immense pleasure in expressing thanks to the Bishop of Coimbatore Diocese, Chairman and Correspondent of CSI College of Engineering **Mr. M. N. Sudhan Appadurai, M.A.,** for his contribution towards our carrier.

We have immense pleasure in expressing thanks to **Dr.J.Joshua Gnana Sekaran B.E,M.S.,Ph.D.,MISTE.,MCSI.** , Principal, CSI College of Engineering, Ooty, for hisencouragement in successful completion of this project.

We express our gratitude to **Prof. K. Komathy Vanitha BE., M.Tech., (Ph.D).,** HOD-Department of Electronics and Communication Engineering, CSI College of Engineering, Ooty, for giving us kind permission to carry out our project.

We would like to thank our Project coordinator **Prof.K. Komathy Vanitha, B.E.,M.Tech.,(Ph.D),** Department of Electronics and Communication Engineering, for their valuable support and encouragement.

We would like to place our heartfelt thanks to **Mrs. J. Jincy,M.Tech.,** Assistant professor, Department of Electronics and Communication Engineering, CSI College of Engineering, Ooty, for his guidance.

We also thank all the Staff members of the Department of Electronics and Communication Engineering, CSI College of Engineering, Ooty for being supportive

Lastly, we would like to thank our friends and family for standing with us and encouraging us throughout the project.

III

**ABSTRACT**

Learning Disability (LD) is a neurological disorder that affects the brain‘s ability to receive process, store and respond to the information. Nowadays, millions of people are suffering from learning disability. Dyslexia is one of the learning disabilities. Based on the characteristics of dyslexia, the detection of dyslexia assumes significance. The major causes for dyslexia are consumption of more medicines during pregnancy, over the counter purchase of medicines for minor ailments without the advice of physicians and uncared head injuries during childhood days. The incidence of this problem is acute in India. Attempts were made by many to detect dyslexic persons to minimize the intensity of this problem. Generally, dyslexic persons have higher percentage of forward and regressive movements and more number of fixations and regressions when compared to normal persons. In the proposed system Machine learning can be used to find meaningful patterns characterizing individual differences. In this project to overwhelm the in dyslexia\_12.4 keel datasets, an various machine learning approach used to evaluate the performance and most suitable algorithm. This model tackles the issue of uncertainty by introducing the degree of hesitation which well defines the instances with multiple class labels. The simulation results prove the performance of this SVM model which greatly assists the parents to discover the symptoms of dyslexia and recommend them to take their children to a psychologist for individual checkups competitively other algorithm KNN, decision tree and linear discriminate analyser.

**KEYWORDS:** Dyslexia,KNN,decision tree,Linear discriminant analysis .

IV

|  | **TABLE OF CONTENT** |  |  |
| --- | --- | --- | --- |
| CHAPTER NO | TITLE | PAGE NO. |  |
|  | **ABSTRACT** | IV |  |
|  | **LIST OF FIGURE** | VII |  |
|  |  |  |
|  | **LIST OF TABLES** | VIII |  |
| **1** | **INTRODUCTION** | 1 |  |
| 1.1 | INTRODUCTION | 2 |  |
| 1.2 | TYPE OF LEARNING DISABILITIES | 3 |  |
| 1.3 | OVERVIEW OF DYSLEXIA | 5 |  |
| 1.4 | IMPORTANCE OF STUDY | 7 |  |
| 1.5 | OBJECTIVE OF STUDY | 8 |  |
| 1.6 | ORGANIZATION OF THE THESIS | 8 |  |
| 1.7 | SUMMARY | 9 |  |
| **2** | **LITERATURE REVIEW** | 10 |  |
| 2.1 | INTRODUCTION | 11 |  |
| 2.2 | REVIEWS ON CONVENTIONAL | 11 |  |
|  | DYSLEXIA DETECTION METHODS |  |  |
| 2.3 | SUMMARY | 17 |  |
| **3** | **PROPOSED SYSTEM DESIGN** | 18 |  |
| 3.1 | INTRODUCTION | 19 |  |
| 3.2 | DATASET DESCRIPTION | 19 |  |
| 3.3 | PROPOSED SYSTEM | 20 |  |

V

| 3.4 | MACHINE LEARNING | 21 |
| --- | --- | --- |
|  | CLASSIFIACTION |  |
| 3.4.1 | LINEAR DISCRIMENT ANALYSIS | 21 |
| 3.4.2 | SUPPORT VECTOR MACHINE | 22 |
|  | (SVM) |  |
| 3.4.3 | DECISION TREE | 23 |
| 3.4.4 | K-NEAREST NEIGHBOUR | 24 |
| 3.5 | HARDWARE AND SOFTWARE | 25 |
|  | REQUIREMENTS |  |
| 3.5.1 | HARDWARE REQUIREMENTS | 25 |
| 3.5.2 | SOFTWARE REQUIREMENTS | 25 |
| 3.6 | SOFTWARE DESCRIPTION | 25 |
| 3.7 | SUMMARY | 29 |
| **4** | **RESULT AND DISCUSSION** | 30 |
| 4.1 | EVALUTION CRITERIA | 31 |
| 4.2 | PERFORMANCE EVALUATION | 34 |
| 4.2.1 | SENSITIVITY | 34 |
| 4.2.2 | SPECIFICITY | 34 |
| 4.2.3 | ACCURACY | 34 |
| 4.3 | SUMMARY | 38 |
| **5** | **CONCLUSION** | 39 |
| 5.1 | CONCLUSION | 40 |
| 5.2 | FUTURE WORK | 40 |
|  | **REFERENCE** | 41 |
|  | **SAMPLE CODE** | 43 |

VI

|  | **LIST OF FIGURES** |  |  |
| --- | --- | --- | --- |
| **FIGURE** | **FIGURE NAME** | **PAGE** |  |
| **NO** |  | **NO** |  |
| 3.1 | ATTRIBUTE DESCRIPTION | 20 |  |
| 3.2 | PROPOSED SYSTEM BLOCK | 21 |  |
|  | DIAGRAM |  |  |
| 3.3 | PSEUDOCODE FOR LINEAR | 22 |  |
|  | DISCRIMINATE ANALYSIS |  |  |
| 3.4 | PSEUDOCODE FOR SVM DYSLEXIA | 22,23 |  |
|  | DATASET PREDICTION MODEL |  |  |
| 3.5 | PSEUDOCODE FOR DECISION TREE | 23 |  |
|  | DYSLEXIA DATASET PREDICTION |  |  |
|  | MODEL |  |  |
| 3.6 | PSEUDOCODE FOR K-NEAREST NEIGHBOR | 24 |  |
|  |  |
|  | DYSLEXIA DATASET PREDICTION MODEL24 |  |  |
| 4.1 | KNN CONFUSION MATRIXES | 32 |  |
| 4.2 | DECISION TREE CONFUSION MATRIXES | 33 |  |
| 4.3 | LDA AND SVM CONFUSION MATRIXES | 33 |  |
| 4.4 | SENSITIVITY COMPRESSION RESULTS | 35 |  |
| 4.5 | SPECIFICITY COMPRESSION RESULTS | 36 |  |
| 4.6 | ACCURACY COMPRESSION RESULTS | 37 |  |

VII

|  | **LIST OF TABLES** |  |
| --- | --- | --- |
| **TABLE** | **TABLE NAME** | **PAGE** |
| **NO** |  | **NO** |
| **4.1** | CONFUSION MATRIX | 31 |
| **4.2** | ALGORITHM MODELS COMPRESSION | 35 |
|  | RESULTS |  |
| **4.3** | ALGORITHM MODELS COMPRESSION | 36 |
|  | RESULTS |  |
| **4.4** | ALGORITHM MODELS ACCURACY | 36 |
|  | COMPRESSION RESULTS |  |

VIII

**CHAPTER 1**

**INTRODUCTION**

1

**1.1 INTRODUCTION**

Brain waves, the electrical signals generated by the brain, hold the key to unlocking the mysteries of human cognition, behavior, and neurological health. By analyzing and classifying these intricate patterns, researchers and clinicians can gain valuable insights into the inner workings of the mind, paving the way for innovative advancements in fields such as neural engineering, neuropsychology, and brain-computer interfaces. This introduction will explore the fundamental principles of brain wave classification, delving into the various machine learning approaches that have been employed to harness the power of these neural signals and uncover their hidden potential.

Learning disabilities (LD) refer to a group of neurological disorders that affect the brain's ability to process and comprehend information effectively. These disorders can impact various cognitive processes, such as reading, writing, mathematics, reasoning, and attention.

Nature of Learning Disabilities: Learning disabilities are lifelong conditions that affect individuals differently. They are not indicative of intelligence but rather represent differences in how the brain processes information. These differences can manifest as difficulties in acquiring, organizing, retaining, or expressing information.

Types of Learning Disabilities: There are several types of learning disabilities, including dyslexia (reading disorder), dyscalculia (mathematics disorder), dysgraphia (writing disorder), auditory processing disorder (difficulty processing auditory information), and attention-deficit/hyperactivity disorder (ADHD), among others. Each type of LD presents unique challenges and may require different approaches to intervention and support.

Impact on Education and Daily Life: Learning disabilities can significantly impact academic performance, social interactions, and overall well-being. Students with LD may struggle in school, experience low self-esteem, and face challenges in transitioning to higher education or the workforce. However, with appropriate support and accommodations, individuals with LD can succeed academically and thrive in various aspects of life.

2

Identification and Assessment: Diagnosing learning disabilities involves comprehensive assessment by qualified professionals, including educators, psychologists, and medical professionals. Assessment typically includes reviewing medical history, conducting cognitive and academic evaluations, and observing the individual's learning behaviors and challenges.

Intervention and Support: Effective intervention for learning disabilities involves a multi-disciplinary approach that may include specialized instruction, assistive technology, accommodations (e.g., extended time on tests), and targeted interventions to address specific skill deficits. Early intervention is key to helping individuals with LD overcome challenges and reach their full potential.

Legal and Ethical Considerations: Laws such as the Individuals with Disabilities Education Act (IDEA) in the United States ensure that individuals with learning disabilities receive appropriate educational services and accommodations. It is essential to uphold the rights and dignity of individuals with LD and provide equal access to education and opportunities.

**1.2 Types of Learning Disabilities**

The majority of learning disabilities are language-based and difficulties using oral and written language remain the single most significant deterrent to educational growth Wallach. The following are the types of learning disability.

**a. Dysgraphia:**

Dysgraphia is characterized by difficulties in learning to write, distinct from general cognitive delays or neurological disorders. Challenges in dysgraphia may include poor handwriting, difficulty with spelling, and slow writing speed. Factors such as automaticity in letter retrieval, orthographic processing, and fine motor skills influence handwriting skills.

**Dyslexia**

Dyslexia is a specific language-based disorder characterized by difficulties in decoding words, particularly in relation to phonological processing. Individuals with dyslexia may have trouble with reading, writing, and spelling, despite having normal intelligence and adequate educational opportunities. Phonological dyslexia specifically refers to difficulty in making sense of speech sounds and their connections to letters. Orthographic dyslexia involves difficulty in remembering how letters and words look, leading to reading and spelling challenges.

**c. Dyscalculia:**

Dyscalculia involves difficulties with mathematical computations and problem-solving, often associated with neurological dysfunction. Challenges in dyscalculia may include concept formation, procedural learning, and visual-motor integration, impacting problem-solving abilities.

**d. Dyspraxia:**

Dyspraxia is characterized by motor difficulties that can affect body movements, resulting in clumsiness, handwriting problems, or speech difficulties. Unlike dyslexia and dyscalculia, dyspraxia does not directly impact reading, writing, spelling, or mathematics but may affect balance and coordination.

Each type of learning disability presents unique challenges and may require different approaches to intervention and support. Understanding the specific nature of these disabilities is crucial for developing effective strategies to help individuals overcome their difficulties and achieve academic success. In this research work, the focus is primarily on dyslexia, with a detailed exploration of its concept and implications.

4

**1.3. OVERVIEW OF DYSLEXIA**

Dyslexia is a specific learning disability that primarily affects reading and related language-based skills. It is characterized by difficulties in decoding words, spelling, and fluent reading, despite adequate intelligence, educational opportunities, and motivation. Individuals with dyslexia may struggle with recognizing and processing letters, sounds, and words, leading to challenges in reading comprehension and written expression.

Key features of dyslexia include:

**i. Phonological Processing Difficulties**:

Individuals with dyslexia often have trouble with phonological processing, which involves recognizing and manipulating the sounds of language. This difficulty can affect their ability to map letters to their corresponding sounds and decode words accurately. Orthographic Processing Challenges: Dyslexia may also involve difficulties in orthographic processing, which refers to recognizing and remembering the visual aspects of words, such as letter forms and spelling patterns. Individuals with dyslexia may struggle with sight word recognition and spelling consistency.

**ii. Reading Fluency and Comprehension Issues:**

Dyslexia can result in slow and laborious reading, as well as difficulties with reading fluency and comprehension. Individuals may have trouble understanding the meaning of text, making inferences, and recalling information.

**iii. Impact on Writing Skills:** Dyslexia can also affect writing skills, including spelling, grammar, and organization. Individuals may have difficulty expressing their thoughts coherently on paper and may avoid writing tasks due to frustration or anxiety.

**iv. Variable Presentation:**

Dyslexia can vary widely in its presentation and severity among individuals. Some may exhibit mild difficulties that are easily managed with appropriate support, while others may struggle significantly despite intervention efforts.

**iv. Associated Challenges:** Dyslexia may be associated with other difficulties, such as attention-deficit/hyperactivity disorder (ADHD), language disorders, and executive function deficits. These additional challenges can further impact academic performance and social-emotional well-being. Overall, dyslexia is a complex and multifaceted learning disorder that requires comprehensive assessment and targeted intervention to address the specific needs of individuals affected by it. Early identification and appropriate support are essential for helping individuals with dyslexia overcome challenges and reach their full potential in academic and personal pursuits.

**1.4. IMPORTANCE OF THE STUDY**

The importance of this study lies in addressing the critical need for early detection and intervention in dyslexia, a prevalent learning disability. Below are the refined points that underscore the significance of this research:

**Enhancing Learning Abilities:** Detecting dyslexia early is vital for improving learning abilities. This study aims to provide a comprehensive approach to identifying dyslexia, which is crucial for enhancing learning outcomes.

**Role of Physicians, Parents, and Society:** Physicians, parents, and society play pivotal roles in dyslexia detection. This research equips them with valuable insights and tools to identify dyslexia early on, facilitating timely intervention and support.

**Empowering Parents:** Early recognition of dyslexia empowers parents to understand and address their children's needs effectively. By recognizing early signs and accessing appropriate services, parents can better support their children's educational journey.

**Societal Impact:** Dyslexia affects a growing number of individuals, and its early identification and treatment can positively impact society as a whole. By addressing dyslexia early, individuals can overcome challenges and contribute more effectively to society.

**Psychological Well-being:** Counseling and support provided to dyslexic individuals can significantly improve their psychological well-being and confidence levels. This research aims to contribute to more effective counseling strategies, thereby enhancing the overall competence and self-esteem of dyslexic individuals.

**Preventing Academic Struggles:** Undiagnosed dyslexia can lead to prolonged academic struggles and frustration. Early detection through the methodologies proposed in this study can prevent such struggles and pave the way for academic success.

**Hospital-Based Early Detection:** Converting the research findings into an embedded system for use in hospitals can revolutionize dyslexia detection. This application can enable early identification of dyslexia, leading to timely intervention and improved outcomes for affected individuals.

By addressing these issues, this research not only contributes to the scientific understanding of dyslexia but also has practical implications for individuals, families, and society at large. The implementation of the proposed methodology in jupyter notebook ensures its accessibility and usability for researchers, educators, and healthcare professionals.

**1.5. OBJECTIVES OF THE STUDY**

The objectives of the study are as follows:

**Increasing Awareness:** The primary objective is to raise awareness about dyslexia, particularly in India where awareness levels may be inadequate. By highlighting the significance of early detection, the study aims to encourage individuals to seek diagnosis and support for dyslexia.

**Early Detection:** The study seeks to facilitate early detection of dyslexia, enabling individuals to receive timely intervention and support. By identifying dyslexia at an early stage, the study aims to mitigate its impact on learning and academic performance.

**Utilizing Dyslexia Keel Dataset:** The study focuses on analyzing the Dyslexia Keel dataset to explore the feasibility of dyslexia detection solely from this dataset. This objective aims to assess the effectiveness of using existing datasets for dyslexia research and detection purposes.

**Developing Machine Learning Approaches:** The study aims to develop and implement various machine learning approaches for dyslexia detection. By employing different algorithms and techniques, the study seeks to enhance the accuracy and reliability of dyslexia detection methods.

**Performance Evaluation:** The study intends to evaluate the performance of the proposed dyslexia detection methods using established metrics such as sensitivity, specificity, accuracy, precision, F-measure, and P-value. This objective aims to assess the efficacy of the developed approaches in accurately identifying dyslexia.

Overall, the objectives of the study encompass raising awareness, facilitating early detection, utilizing existing datasets, developing machine learning approaches, and evaluating performance metrics to contribute to the field of dyslexia research and detection.

**1.6. ORGANIZATION OF THE THESIS**

The thesis is organized into several chapters, each focusing on different aspects of dyslexia detection using machine learning approaches. Here's a brief overview of each chapter:

**Introduction and Background:** This chapter provides an introduction to image processing techniques, including feature extraction, selection, and various classification methods.

It also introduces the concept of learning disabilities, with a particular focus on dyslexia, and outlines the need for the research work along with its objectives.

**Literature Review:** Chapter 2 conducts a survey of conventional dyslexia detection methods, including feature extraction, selection, and classification techniques.

It covers existing research on dyslexia detection, focusing on methods utilizing eye movements as a biometric indicator.

**Dataset Description and Preprocessing:** This chapter describes the eye movement biometric database version 1 used in the research. It details the preprocessing steps applied to the dataset and discusses feature extraction and selection techniques utilized in the study. Classification techniques such as Decision Tree, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) also Covolutioal neural network (Cnn) covered.

**Implementation and Results:** Chapter 4 discusses the implementation of machine learning approaches for dyslexia detection. It compares the results of classification methods on different datasets and presents the average performance metrics of all classifiers.

**Conclusion and Future Work:** The final chapter concludes by summarizing the findings, highlighting the efficacy of the FANFIS-SVM classification method for dyslexia detection. It also outlines the scope for future research and potential areas for further investigation in the field of dyslexia detection using machine learning.

Overall, the thesis provides a comprehensive overview of dyslexia detection methods, implementation of machine learning approaches, and insights into future research directions.

**1.7. SUMMARY**

This chapter carries information concerning learning disability, its types, and different types of Dyslexia. This chapter has a focus on the need for this kind of research, objectives of research undertaken and the way in which the research the research report is organized as a thesis.

**CHAPTER 2**

**LITERATURE REVIEW**

10

**2.1 INTRODUCTION**

Researchers, neurologists, investigators, parents and teachers have developed various architecture models, simulated systems, prototypes for reasonably accurate prediction/detection of dyslexia. This chapter is an attempt to review conventional dyslexia detection methods, eye movement based dyslexia detection, feature extraction, feature selection methods, artificial neural network, adaptive neuro fuzzy inference system, particle swam optimization technique, support vector machine and fuzzy membership function with adaptive neuro fuzzy inference system. Relevant literature related to these works are condensed and given below.

**2.2 REVIEWS ON CONVENTIONAL DYSLEXIA DETECTION METHODS**

Stanberry et al. (2006) reported differences in functional Magnetic Resonance Imaging (fMRI) connectivity in adults with and without dyslexia during phoneme mapping. Analysis of correlations in the low-frequency range showed that regions known to activate during an "on-off" phoneme-mapping task exhibited synchronous signal changes when the task was administered continuously (without any "off" periods). This research showed that three functional networks, which were defined on the basis of documented structural deficits in dyslexics and included regions associated with phonological processing, it differs significantly in spatial extent between good readers and dyslexics.

Aylward et al. (2008) explained different neuro imaging technologies for revealing the biological basis of reading and dyslexia. fMRI is most suited to localisation of function, and hence to investigating the neural networks that underpin efficient (or inefficient) reading. Electro EncephaloGraphy (EEG) is sensitive to millisecond differences in timing, hence it is suited for studying the time course of processing Neuro imaging studies of developmental dyslexia are then reviewed, focusing on (a) the neural networks recruited for reading, (b) the time course of neural activation and (c) the neural effects of remediation. Representative studies using the different methodologies are selected. This study shows that dyslexic brain is characterised by under-activation of the key neural networks for reading.

11

Karstenspecht et al. (2008) studied fMRI, for adults, considered at risk for dyslexia were compared with an age-/gender-matched control group for differences in brain activation when presented with visual stimuli differing in demands for literacy processing. Stimuli were nameable pictures, brand logos familiar to individuals, and written words – these were either regularly spelled using early-acquired rules (alphabetic) or more complex (orthographic).

Brain responses distinguished between the presentation conditions, as afunction of group, within many cortical areas. SooYeonJi & Kayvan Najarian (2008) designed a hierarchical method that applies an optimization algorithm based on Modified Maximum Correlation Model (MMCM) that can detect small variations across the two study groups. Based on this method, they also hypothesize that dyslexiamight represent functional Magnetic Resonance Imaging (fMRI) brain signal activities in specific regions of the brain that are distinguishable from healthy brain fMRIs. In this work, they have presented a model to detect the brain activation using a hierarchical optimization algorithm. In particular, an optimization algorithm is applied to the specific regions of the brain identified to find patterns that are significantly different across two groups of subjects.

The method was specialized for optimal brain fMRI filtering and signal detection in noisy images. The limitation of this study was the small size of the dataset used to identify the fMRI brain activity. Even though the dataset was small sized, the method successfully processes the active raw time series in both dyslexic and healthy subjects in order to distinguish regions differentially activated in the two groups.

Schulz et al. (2008) recorded brain activity of 47 persons (16 with dyslexia, 31 controls) with fMRI and Event-Related Potentials (ERP) in two separate counter balanced sessions. The person silently read and occasionally judged simple sentences with semantically congruous or incongruous endings. fMRI and ERP activation during sentence reading and semantic processing was analyzed across all the individuals and also by comparing an individual with dyslexic and healthy subjects. For sentence reading, the analysis is made based on response to all words in a sentence. Sentence reading was characterized by activation in a left-lateralized language network. Semantic processing was characterized by activation in left-hemispheric regions of the inferior frontal and superior temporal cortex and by an electrophysiological N400 effect after 240 ms with consistent left anterior source localization. A person with dyslexia showed decreased activation for sentence reading in inferior parietal and frontal regions, and for semantic processing in inferior parietal

12

regions. Together, this suggests that semantic impairment during sentence reading reduces dyslexicperson‘s response in left anterior brain regions.

Soo-YeonJi et al. (2009) evaluated the capability of a hierarchical method that performed an optimization algorithm based on MMCM. The optimization algorithm was designed by adopting modified maximum correlation model to detect active regions that contain significant responses. The optimization algorithm was examined based on two groups of datasets, dyslexia and healthy subjects, to verify the ability of the algorithm that enhances the quality of signal activities in the interested regions of the brain. After verifying the algorithm, Discrete Wavelet Transformation (DWT) is applied to identify the difference between healthy and dyslexia subjects. In this research work, they have concluded that the wavelet transform is an appropriate non-stationary signal analysis method, which may be suitable for differentiating two different conditions.

Eden et al. (1994) found that eye movement tracking on non-reading tasks could reliably differentiate between good and poor readers. Poor readers tended to have jerky and erratic eye movements when attempting to visually track a moving target. The author hypothesized that deficits in eye movement tracking among poor readers appear to be poor eye movement control. Biscaldi et al. (1994) explained saccades in five noncognitive tasks. Parameters of the eye-movement data were collected for each person. On the basis of their reading, writing, and other cognitive performances, twelve people were considered dyslexic and were divided into two groups (Dl and D2). Group statistical comparisons revealed significant differences between control and dyslexic subjects. The standard tasks of the dyslexic subjects have poorer fixation quality, failed more often to hit the target at once, had smaller primary saccades and shorter reaction times to the left as compared with the control group. The control group and group Dl dyslexics showed an asymmetrical distribution of reaction times, but in opposite directions. Group D2 dyslexics made more anticipatory and express saccades, they undershot the target more often in comparison with the control group, and almost never overshot it. In the sequential tasks group Dl subjects made fewer and larger saccades in a shorter time and group D2 subjects had shorter fixation durations than the subjects of the control group.

Maria De Luca et al. (1999) explained eye movements in non-linguistic and linguistic tasks. Stability of fixation on a stationary stimulus was examined. Performance of dyslexics was no different from that of an age matched control group. Similarly, no difference was observed between

13

the two groups when they were requested to saccade to a rightward or leftward target. On the other hand, while reading short passages, dyslexics showed an altered pattern of eye movements with more frequent and smaller rightward saccades as well as longer fixation times. The reading pattern was analysed by eye tracking. Numerous fixations were used to read a single word in a fragmented way. Longer words showed a higher number of fixations. It was concluded that surface dyslexia is not associated with oculo-motor dysfunction and the study of eye movements in reading reveals the processing through orthography-technology conversion characteristic of surface dyslexia. The importance is stressed of examining selected groups of subjects in the psycho physiological study of dyslexia.

Hutzler & Wimmer (2004) Participants were German dyslexic readers who—compared to English dyslexic readers—suffer mainly from slow laborious reading and less from reading errors. The eye movements of eleven dyslexic boys and age-matched controls were recorded during reading of text passages and pseudoword lists. For both text and pseudoword reading, the dyslexic readers exhibited more and much longer fixations, but relatively few regressions. Increased length of words and pseudowords led to a greater increase in number of fixations for dyslexic than normal readers. Comparisons across studies suggest that the present German dyslexic eye movement findings differ from English-based findings by a lower frequency of regressions (presumably due to the higher regularity of German) and from Italian findings by longer fixation duration (presumably due to the greater syllabic complexity of German).

Hawelka & Wimmer (2005) theorized that shorter saccades are the source of greater fixations among children with Reading Disability as compared to normal readers. Shorter saccades are common in letter-by-letter reading and contribute to a slow and laborious reading style. Hawelka & Wimmer (2005) examined types of eye movements (e.g. fixations and saccades), speed of reading, and errors in reading in dyslexic and control children. Results of the study indicated that dyslexic children made fewer errors than the control group; however, their reading speed was significantly slower than the Control subjects. Differences in reading rate were associated with the number of eye movements made during reading. That is, participants with more eye movements had slower reading speeds. These findings support the letter-by-letter reading pattern thought to be characteristic of reading disabled children. Stefan Hawelka et al. (2005) presented eye-movements, which represent a great interest in studying the specificity of the reading difficulties that individuals

14

with developmental dyslexia have. In the studied dyslexic children were pair-matched with control children in a sentence reading task. The children read sentences in Bulgarian – a Cyrillic alphabet language with regular orthography. Target nouns with controlled frequency and length were embedded in the sentences. Eye movements revealed highly significant group differences in the gaze time and the total fixation times, word frequency and word length effects as well as interaction for both frequency and length with the group factor.

The non-reading task used by Hutzler et al. (2006) was considered to be still closer to the perceptual and oculomotor demands of reading. In thisstudy, the dyslexic and control participants had to read aloud series of pseudowords (e.g., GUFT). Their eye movement patterns on this task were compared to those recorded in a string processing task where they had to search through lists of consonant strings (e.g., GDRK, LBQD) for items with two adjacent identical letters (e.g., VPLL). The consonant strings were built from the pseudo-words by replacing vowels by consonants. The spatial arrangement of pseudo-words and consonant strings was the same. The eye movement patterns of dyslexic and control readers did not differ when performing the visual search task, whereas they strongly differed in the reading task. Such findings were interpreted as evidence against the hypothesis of visual perceptual or oculomotor problems in developmental dyslexia.

Evgenia Hristova et al. (2006) analyzed the relationship between dyslexia and eye movements. 22 patients were included in the study, 11patients have a diagnosis of dyslexia and 11 subjects were used as a control group (normal readers). All patients underwent careful orthoptic and ophthalmological visit; eye movements were quantified by Ober-2 system. Ocular motility was divided into three phases: stability analysis, analysis of fixation pauses, analysis of tracking saccades (left and right horizontal axis); speed reading, saccades and regressions through the reading of a text. The stability analysis on fixating a still target showed a significant (p<0.001) difference between dyslexic and control group, outlining an increased amount of loss of fixation in dyslexic subjects.

Chloe Pradoa et al. (2007) explained eye movements of 14 French dyslexic persons having a VA span reduction and 14 normal readers were compared in two tasks of visual search and text reading. The dyslexic participants made a higher number of rightward fixations in reading only. They simultaneously processed the same low number of letters in both tasks whereas normal readers processed far more letters in reading. Importantly, the person ‘s VA span abilities related to the

15

number of letters simultaneously processed in reading. The typical eye movements of some dyslexic readers in reading thus appear to reflect difficulties to increase their VA span according to the task request.

Kerstin et al. (2010) used eye movements approach to advance the understanding of impaired information processing in acquired central dyslexia of stroke patients with aphasia. Till now there has been no research attempting to analyze both word based viewing time measures and local fixation patterns in dyslexic readers. The goal of the study was to find out whether specific eye movement parameters reflect pathologically preferred segmental reading in contrast to lexical reading. They compared oral reading of single words of normal controls (n = 11) with six aphasic participants (two cases of deep, surface and residual dyslexia each). Their mean fixation duration was already prolonged during first pass reading reflecting their attempts of immediate access to lexical information. After first pass reading, re-reading time was significantly increased in all participants with acquired central dyslexia due to their exceedingly higher monitoring demands for oral reading.

Servet Bayram et al. (2012) examined and compared dyslexic and normal readers, reading habits and eye movements during reading passages and pseudowords. Participants were 15 Turkish dyslexic students, suffering mainly from reading disorder. In addition, there were 15 Turkish students who were regular readers and who did not have problems with reading. During reading passages and pseudowords, the eye movements of participants were recorded. For both text and pseudoword reading, the dyslexic readers exhibited more and much longer fixations, but relatively few regressions. An increased length of words and pseudowords led to a greater increase in the number of fixations for dyslexics rather than normal readers.

Peyrin et al. (2012) explained Visual Attention (VA) span disorders have been reported in dyslexic persons. This study investigates whether this cognitively-based dissociation has a neurobiological counterpart through the investigation of two cases of developmental dyslexia. Learning level showed a phonological disorder but preserved VA span whereas Frontal Gyrus (FG)exhibited the reverse pattern. During a phonological rhyme judgement task, LL showed decreased activation of the left inferior Frontal Gyrus whereas this region was activated at the level of the controls in FG. Conversely, during a visual categorization task, FG demonstrated decreased activation of the parietal lobules whereas these regions were activated in LL as in the controls.

16

These contrasted patterns of brain activation thus mirror the cognitive disorders‘dissociation. Novak et al. (2004) presented a Self-Organizing Map (SOM) and Genetic Algorithm (GA)

for extraction of set of features for detection of eye movement signal in both horizontal and vertical manner. They concluded that the reading speed enlarged with the likelihood of the patient being healthy. In this method, an inductive modeling technique was applied to data set resulting in extraction of six features which were used as the input to Self-Organizing Map (SOM). Three clusters were finally formed by the SOM proving that the proposed methodology is suitable for automatic dyslexia analysis.

Lei yu et al. (2004) applied feature selection for many applications where data has hundreds and thousands of features. The authors proposed a new framework of efficient feature selection through relevance and redundancy analysis. This work has been implemented in supervised learning where data contains many irrelevant and/or redundant features. Wu et al. (2008) proposed feature selection for learning disability diagnosis problem. In this work, a GA-based feature selection algorithm is proposed as the pre-processing step. And this wrapper-based GA feature selection procedure can improve the learning disability identification accuracy.

**2.3 SUMMARY**

This chapter discusses about conventional dyslexia detection methods, eye movements based dyslexia detection, feature extraction methods, feature selection methods, classification based dyslexia detection methods in a detailed manner.

17

**CHAPTER 3**

**PROPOSED SYSTEM DESIGN**

18

**3.1 INTRODUCTION**

Dyslexics can largely learn to read, write and study effectively when they use methods designed to their unique learning style. Based on recent progress in theory and dyslexia measurement techniques, the analysis of movements has become one of the major methodological tools in experimental reading research. This chapter discusses the classification, feature extraction and selection techniques and eye movement biometric database version. The proposed research methods are analyzed based on this KEEL Dyslexia dataset.

**3.2. DATASET DESCRIPTION**

We collected EEG signal data from 10 college students while they watched MOOC video clips. We extracted online education videos that are assumed not to be confusing for college students, such as videos of the introduction of basic algebra or geometry. We also prepare videos that are expected to confuse a typical college student if a student is not familiar with the video topics like Quantum Mechanics, and Stem Cell Research. We prepared 20 videos, 10 in each category. Each video was about 2 minutes long. We chopped the two-minute clip in the middle of a topic to make the videos more confusing.

The students wore a single-channel wireless MindSet that measured activity over the frontal lobe. The MindSet measures the voltage between an electrode resting on the forehead and two electrodes (one ground and one reference) each in contact with an ear.

After each session, the student rated his/her confusion level on a scale of 1-7, where one corresponded to the least confusing and seven corresponded to the most confusing. These labels if further normalized into labels of whether the students are confused or not. This label is offered as self-labelled confusion in addition to our predefined label of confusion.

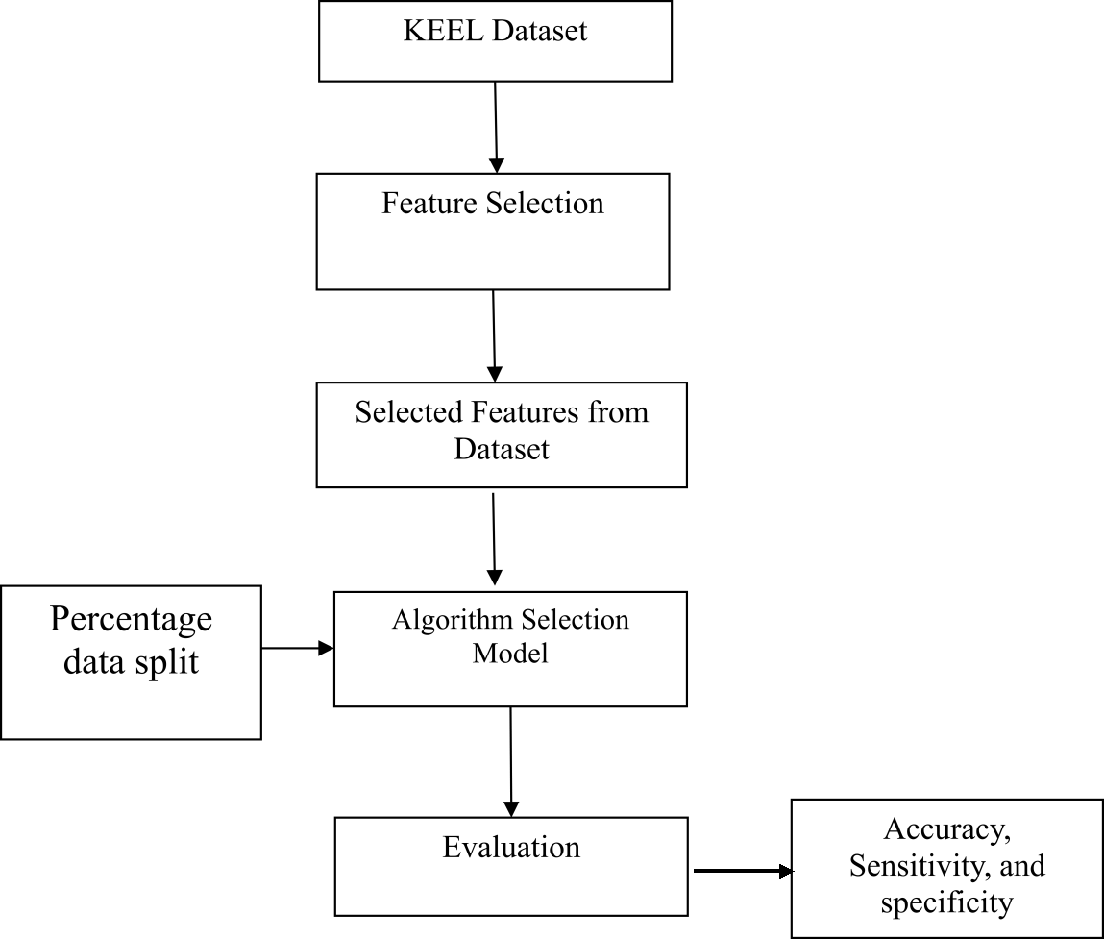
The present dataset contains vague data in both input and output variables. The output variable is a subset of the 4 labels that follow:

1. No dyslexic.
2. Control and revision.
3. Dyslexic.
4. Inattention, hyperactivity or other problems.

The vagueness data is available both in input and output and it consists of missing value also. In this work, the missing values are handled by applying different machine learning approach based imputation method to produce the complete dataset. The missing values of a particular instance are selected and their remaining attributes are compared with other complete set instances.

After determining k-nearest neighbors, the missing value is filled by their obtained mean value of those neighbors’s corresponding attribute value. The input dataset consist of interval value is transformed into midpoints and then they are converted into intuitionistic fuzzy domain representation.

**3.3 PROPOSED SYSTEM**

Implementation for developing the Dyslexia prediction uses jupyter noteook. The algorithm will provide an optimal solution in Dyslexia detection process in the KEEL datasets and identify the in Dyslexia or not in supervised manner. Many researchers have developed different machine learning techniques to diagnose the Dyslexia. Different models have been developed using different algorithms for prediction of the Dyslexia. But no precise model has been developed in diagnosing.the disease accurately. Hence, in this proposed research, an integrated framework model developed. **Figure 3.2** **Proposed System Block Diagram**

**3.4.** **MACHINE LEARNING CLASSIFICATION**

**3.4.1. Linear Discriminant Analysis:**

Linear Discriminant Analysis were created by (Paul D. Allison, 1999) (William McKelvey Richard et. al, 1994). Linear Discriminant Analysis, also called as logit model, when the target variable is a categorical variable with two categories - for example, active or inactive, healthy or unhealthy, win or loss, purchase product or does not purchase product etc. This association is usually formulated as an equation in which the independent variables have parametric coefficients that enable future values of the dependent variable to be predicted.



**Input:** Dyslexia Dataset

**OUTPUT:** Classification result

**Method:**

x=random(x) #picks a random starting point

Valuethreshold=0.000001

While true;

Gradient = compute\_gradient(x)

next\_x=step(x, gradient, alpha=-0,001) #gradient negative step

if distance (next\_x, x)<threshold ; # when converging attained process stop break;

x=next\_x #continue if we are not

return x



**Figure 3.3 pseudocode for Linear Discriminate Analysis**

**3.4.2. Support Vector Machine (SVM)**

SVMs were created by Cortes and Vapnik (1995) for parallel characterization of any complex data. Their approach might be generally portrayed as different cases like class partition, Overlapping, classes and Nonlinearity. Normally, the class partition is meant finding the isolating hyperplane between the two classes by boosting the edge between the classes' nearest focuses as shown in the figure 2. The focused channel will be lying on the limits are called bolster vectors and the center of the edge is our ideal isolating hyperplane.



**Input:** Dyslexia Dataset

**OUTPUT:** Classification result

**Method:** #performance estimation by using 10-fold cv

**Step1:** Begin

**Step2:** For j=1:k

**Step3:** Training set= k-1subsets;



22

**Step4:** Testing set=remaining subsets; #Training set parameters obtained

**Step5:** Parameter\_ optimization (k);

**Step6:** Again test on testing set;

For end;

**Step7:** Return #accuracyavg entire dataset

end



**Figure3.4 pseudocode for SVM Dyslexia Dataset Prediction Model 3.4.3. DECISION TREE**

In statistics, Decision Tree classifiers are a family of simple "probabilistic classifiers" based on applying Decision Tree theorem with strong independence assumptions between the features. They are among the simplest Bayesian network models but coupled with kernel density estimation, they can achieve higher accuracy levels.

Decision Tree classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.



**Input:** Dyslexia Dataset

**Output:** Group of attributes

**Method:**

**Step1:** Start keeping first attributes and the class attribute.

**Step2:** Compare the attribute name from the key1 list and key2 list, where key1 is the list to store attributes names based on the ascending order of the entropy value, and key2 is the list to store attributes names in original order.

**Step3:** both are same then remove the attributes from the dataset and also remove the attribute from the key2 list and evaluate. **Step4:** do step until last attributes in the dataset.

**Figure3.5 pseudocode for Decision Tree Dyslexia Dataset Prediction Model**

23

**3.4.4. K-nearest neighbor**

K-nearest neighbor classifier is one of the introductory supervised classifier, which every data science learner should be aware of. Fix & Hodges proposed K-nearest neighbor classifier algorithm in the year of 1951 for performing pattern classification task. For simplicity, this classifier is called as Knn Classifier. To be surprised k-nearest neighbor classifier mostly represented as Knn, even in many research papers too. Knn address the pattern recognition problems and also the best choices for addressing some of the classification related tasks.

The simple version of the K-nearest neighbor classifier algorithms is to predict the target label by finding the nearest neighbor class. The closest class will be identified using the distance measures like Euclidean distance.



**Input:** Dyslexia Dataset

**Output:** Group of attributes

**Method:**

1. Calculate “d(x, xi)” i =1, 2, ….., n; where d denotes the Euclidean distance between the points.
2. Arrange the calculated n Euclidean distances in non-decreasing order.
3. Let k be a +ve integer, take the first k distances from this sorted list.
4. Find those k-points corresponding to these k-distances.
5. Let ki denotes the number of points belonging to the ith class among k points i.e. k ≥ 0
6. If ki >kj ∀ i ≠ j then put x in class i.

**Figure3.6 Pseudocode for K-Nearest Neighbor Dyslexia Dataset Prediction Model**

24

**3.5. HARDWARE AND SOFTWARE REQUIREMENTS**

**3.5.1. HARDWARE REQUIREMENTS**

1. Processor: Intel core 2.30GHZ
2. Memory: 4GB RAM
3. Operating System: Windows 10
4. Hard disk: 500 gb

**3.5.2. SOFTWARE REQUIREMENTS**

1. **Software:** Jupyter noteook

**3.6. SOFTWARE DESCRIPTION**

Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, including Python, R, Julia, and Scala, making it a versatile tool for data analysis, scientific computing, machine learning, and more.

Here's a brief description of Jupyter Notebook's features and functionalities:

**Interactive Computing:** Jupyter Notebook provides an interactive computing environment where you can write and execute code in individual cells. This allows for iterative development and immediate feedback, making it ideal for exploratory data analysis and prototyping.

**Support for Multiple Languages:** While Jupyter Notebook is commonly associated with Python, it supports a wide range of programming languages through its kernel system. Each notebook can be associated with a specific kernel, enabling you to work with languages such as R, Julia, and others.

**Rich Output:** In addition to code execution, Jupyter Notebook allows you to generate rich output, including plots, tables, images, and interactive widgets. This makes it easy to visualize data and communicate insights within the notebook itself.

**Markdown Support:** Jupyter Notebook supports Markdown, a lightweight markup language, allowing you to create formatted text, headings, lists, links, and more. Markdown cells can be interspersed with code cells to provide context, explanations, and documentation within the notebook.

**Integration with Libraries and Tools:** Jupyter Notebook seamlessly integrates with popular libraries and tools for data analysis and visualization, such as NumPy, Pandas, Matplotlib, Seaborn, and scikit-learn. This allows you to leverage the full power of these libraries within your notebooks.

**Collaboration and Sharing:** Jupyter Notebook files are saved in a JSON-based format (.ipynb), which can be easily shared and version-controlled using platforms like GitHub. Additionally, Jupyter Notebook supports the export of notebooks to various formats, including HTML, PDF, and slideshows, making it convenient for collaboration and presentation.

**Extensibility:** Jupyter Notebook is highly extensible, with a vibrant ecosystem of extensions and plugins. You can customize the notebook interface, add new functionalities, and integrate with external tools to suit your specific workflow and requirements.

Overall, Jupyter Notebook provides a flexible and powerful environment for interactive computing, data analysis, and collaborative research. Its intuitive interface and rich features make it a popular choice among data scientists, researchers, educators, and professionals in various fields.

**CHAPTER 4**

**RESULT AND DISCUSSION**

30

**4.1 Evaluation Criteria**

To analyze the performance of the proposed model, accuracy, precision and recall were used for evaluating classification results and sensitivity and specificity were used for evaluating regression results. SVM Algorithm was chosen to solve this problem. Table 4.1 shows the Confusion matrix which is typically used to evaluate performance of data mining algorithms.

| **Subjects** | **Predicted positive** | **Predicted Negative** |
| --- | --- | --- |
|  |  |  |
| Actual positive | True Positive (TP) | False Negatives (FN) |
|  |  |  |
| Actual negative | False positives (FB) | True Negatives (TN) |
|  |  |  |



**True positive**

Cancerous patients correctly diagnosed as malignant

**False positive**

Healthy people wrongly diagnosed as malignant

**True negative**

Healthy people correctly diagnosed as benign

**False negative**

Cancerous patients wrongly diagnosed as benign



**Table 4.1 Confusion Matrix**

Both sensitivity and FROC uses the terms in the confusion matrix. Two independent result sets were requested for the same test data using each of these evaluation measures.

To evaluate the performance of the, the following parameters are used:

= True positive is the result when the examination data are positive for a subject with dyslexia. = False positive is the result when the examination data are negative for a subject with dyslexia. =True negative is the result when the examination data are negative for a subject without

31

dyslexia.

=False negative is the result when the examination data are positive for a subject without

dyslexia.

Accuracy: 0.827984966753397

Classification Report:

precision recall f1-score support

0 0.83 0.82 0.82 1698

1 0.82 0.84 0.83 1761

micro avg 0.83 0.83 0.83 3459

macro avg 0.83 0.83 0.83 3459

weighted avg 0.83 0.83 0.83 3459

samples avg 0.83 0.83 0.83 3459

**Figure 4.1 KNN Confusion Matrixes**

Accuracy: 0.9508528476438277

Classification Report:

precision recall f1-score support

0 0.95 0.95 0.95 1698

1 0.95 0.95 0.95 1761

micro avg 0.95 0.95 0.95 3459

macro avg 0.95 0.95 0.95 3459

weighted avg 0.95 0.95 0.95 3459

samples avg 0.95 0.95 0.95 3459

**Figure 4.2 Decision Tree Confusion**

**4.2 Performance evaluation**

**4.2.1 Sensitivity**

Sensitivity is also called the True Positive Rate (TPR). It evaluates the percentage of actual positive which are dyslexic subjects class. It is the representation of acceptably classified percentage of dyslexic subjects. The sensitivity is defined as below:



**4.2.2 Specificity**

Specificity evaluates the percentage of actual negatives which are related to non-dyslexic subjects class. Specificity is also called as True Negative Rate (TNR). It is the accurateness in identifying non-dyslexic subjects and defines the percentage of correctly classified non-dyslexic subjects.



**4.2.3 Accuracy**

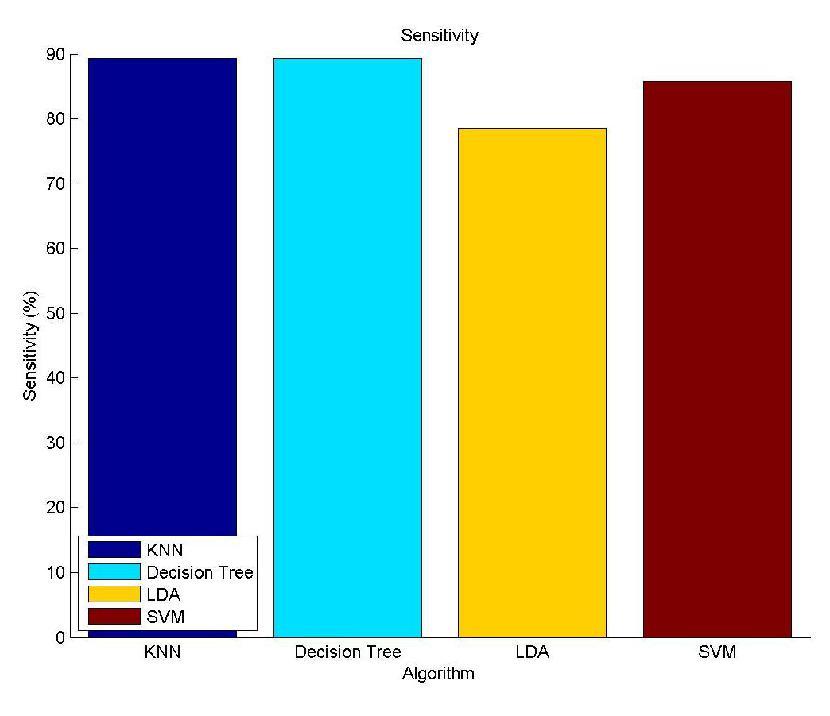
Accuracy is defined as the overall correctness of the model and is calculated as the sum of actual classification parameters (Tp + Tn) separated by the total number of classification parameters (Tp + Tn + Fp + Fn)



To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.1 and performance parameters final values. Total Sensitivity of SVM value is 89.28 %, the conclusion was that the model has performed quite well on this independent dataset. Figure 4.4 also display the graph results for proposed system comparatively K Nearest Neighbour, Decision tree, and Linear Discriminant Analysis.

34

**Table 4.2 Algorithm models compression results**

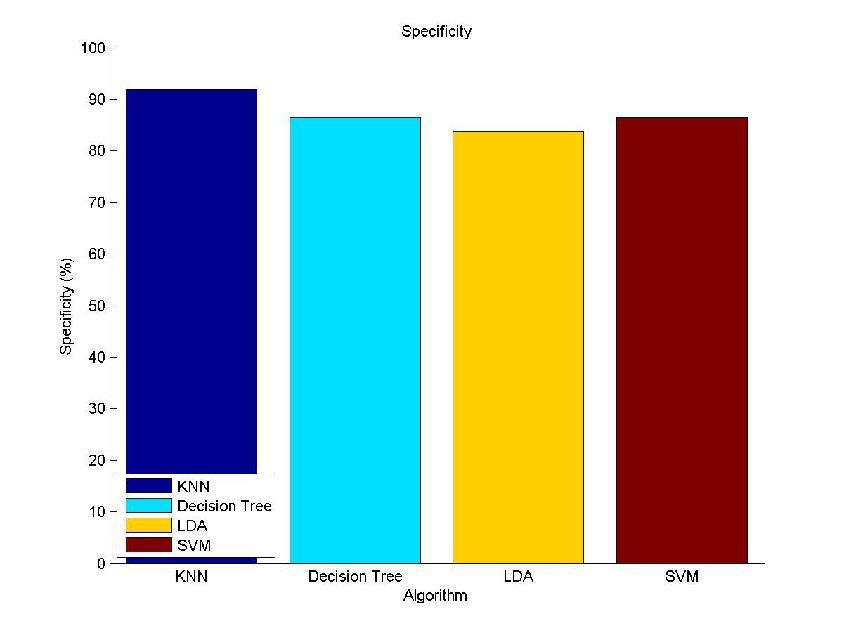


**Figure 4.4 sensitivity compression results**

To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.2 and performance parameters final values. Total Specificity of Decision tree value is 95.04 %, the conclusion was svm that the model has performed quite well on this independent dataset comparatively K Nearest Neighbour, SVM, and Cnn. Figure 4.5 also display the graph results for proposed system proved the better for dyslexia 12-4 imbalanced Dataset.

35

**Table 4.3 Algorithm models compression results**



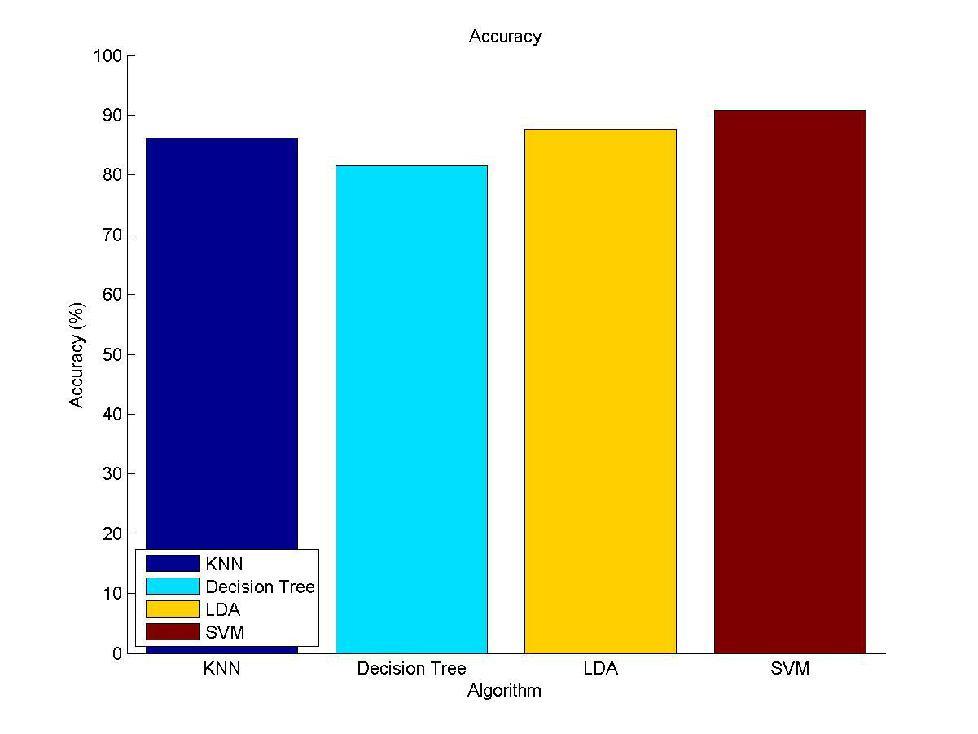
**Figure 4.5 specificity compression results**

|  | **Algorithm** | **Accuracy /Efficiency** |
| --- | --- | --- |
|  | |  |
| K Nearest Neighbour | | 82.97 |
|  | |  |
| Decision tree | | **95.04** |
|  |  |  |
| Cnn |  | 94.73 |
|  |  |  |
|  | |  |
| Support Vector Machine | | 90.7692 |
|  |  |  |

**Table 4.4 Algorithm models Accuracy compression results**

36

To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.3 and performance parameters final values. Total accuracy of SVM value is 90.76%, the conclusion was svm that the model has performed quite well on this independent dataset comparatively K Nearest Neighbour, Decision tree, and Linear Discriminant Analysis. Figure 4.6 also display the graph results for proposed system.



**Figure 4.6 Accuracy compression results**

The performance of SVM method was compared with that described in previous chapter existing algorithms. Prediction results of the data mining and statistical learning algorithms compressions are shown in Figure 4.4 4.5 and 4.6. In existing method compression KNN algorithm is showed slight improvement when applied to the independent dataset. Comparatively KNN algorithm slightly improves and best result for other two algorithms (Decision tree and Cnn) in various parameters (accuracy, specificity, sensitivity). Based on the result algorithms are ranked as Decision tree, Cnn, Support Vector Machine, KNN

37

**4.3 SUMMARY**

This section shows the results obtained from various classification methods applied for dyslexia 12-4 KEEL dataset. The experimentation results of various classification methods have been carried out with the help of six parameters such as sensitivity, specificity, accuracy, and precision,. All these evaluation results conclude that the proposed Decision tree method provides higher detection accuracy than KNN, SVM and Cnn.

38

**CHAPTER 5**

**CONCLUSION**

39

**5.1 CONCLUSION**

Identification of dyslexia affected persons is a problem today in the medical field as diagnosis is a difficult task. Through this work the researcher is attempting to make the diagnosis easy for the problem mostly prevalent in developing countries like India. Then the dyslexia 12-4 KEEL dataset features are passed through a various machine learning classifier like support vector machine, k-nearest neighbour Decision tree and Linear Discriminant Analysis classifiers. The results emanating out of the classifier are evaluated. The evaluation is based on parameters like sensitivity, specificity, and accuracy. This method is not only simple to manage large data simultaneously but also yields good results than previous works. This study is beneficial to physicians for easy and effective diagnosis of learning disability. Academics especially computer scientists can replicate the suggested models to other fields too.

Prediction results of the data mining and statistical learning algorithms compressions are discussed in chapter 4. In machine learning methods compression KNN algorithm is showed slight improvement when applied to the independent dataset. Comparatively KNN algorithm slightly improves and best result for other two algorithms (Decision tree and Linear Discriminant Analysis) in various parameters (accuracy, specificity, sensitivity). Based on the result algorithms are ranked as Support Vector Machine, KNN, Decision tree and Linear Discriminant Analysis.

**5.2 FUTURE WORK**

With the experience gained through this work, the researcher is suggesting to fellow researchers to focus on early diagnosis of learning disability by applying soft computing techniques. Thereby the impact of dyslexia on per-school children can very well be diagnosed and that too at the right time to contain the stress due to unrecognized disorder, such research endeavours will be beneficial to those children suffering due to dyslexia, teachers engaged in special education, and physicians. To be precise, the focus of future research is to identify early diagnosis of dyslexia and proper classification of learning disability to support the special education community in their quest

to be with mainstream

40

**REFERENCES**

1. Aylward, EH, Richards, TL, Berninger, VW, Nagy, WE, Field, KM, Grimme; AC, Richards, AL, Thomson, JB & Cramer, SC 2008, ‗Instructional treatment associated with changes in brain activation in children with dyslexia ‘, Educational Research, vol. 50, no. 2, pp. 135–148.
2. Biscaldi, M, Fischer, B & Aiple, F 1994, ‗Saccadic eye movements of dyslexic and normal reading children Perception ‘, vol. 23, no. 1, pp. 45 – 64.
3. ChloéPradoa, Matthieu Dubois & SylvianeValdois, ‗The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span ‘, Vision Research, vol. 47, issue. 19, pp. 2521–2530.
4. Eden, GF, Stein, JF, Wood, MH & Wood, FB 1994, ‗Differences in Eye Movements and Reading Problems in Dyslexic and Normal Children, Vision Research‘, vol. 34, no. 10, pp. 1345-1358.
5. Evgenia Hristova, Alexander Gerganov, Ekaterina Todorova & Severina Georgieva 2006, ‗Eye-Movements of Dyslexic Children Reading in Regular Orthography: Exploring Word Frequency and

Length Effects‘, Exploring Word Frequency and Length Effects, pp. 56-70.

1. Hawelka, S & Wimmer, H 2005, ‗Impaired Visual Processing of Multi-Element Arrays is Associated with Increased Number of Eye ovements in Dyslexic Reading‘, Vision Research, vol. 45, no. 1, pp. 855-863.
2. Hultquist, AM 2006, ‗An Introduction to Dyslexia for Parents and Professionals‘. London: Jessica

Kingsley Publishers.

1. Hutzler, F & Wimmer, H 2004, ‗Eye movements of dyslexic children when reading in a regular orthography‘, Brain and Language, vol. 89, no. 1, pp. 235–242.
2. Hutzler, F, Kronbichler, M, Jacobs, AM & Wimmer, H 2006. ‗Perhaps correlational but not causal:

No effect of dyslexic readers ‘magnocellular system on their eye movements during reading. Neuropsychologia, vol. 44, no. 4, pp. 637–648.

1. Karstenspecht, Kenneth hugdahl, Sonja ofte, Merethenygård, Atlebjørnerud, Elena plante & Turidhelland 2008, ‗Brain activation on pre-reading tasks reveals at-risk status for dyslexia in 6-year-old children ‘, Scandinavian Journal of Psychology.

41

1. Kerstin I Schattka, Ralph Radach & Walter Huber 2010, ‗Eye movement correlates of acquired central dyslexia Eye movement correlates of acquired central dyslexia ‘, Neuropsychologia, vol. 48, pp. 2959–2973.
2. Lei yu 2004, ‗Efficient Feature Selection via Analysis of Relevance and Redundancy ‘, Journal of Machine Learning Research, 1205–1224 Submitted 1/04; Revised 5/04; Published 10/04.
3. Maria De Luca, Enrico Di Pace, Anna Judica, Donatella Spinelli, Pierluigi Zoccolotti 1999, ‗Eye movement patterns in linguistic and non-linguistic tasks indevelopmental surface dyslexia ‘, Neuropsychologia, vol. 37, pp. 1407-1420
4. Novak, D, Kordk, P, Macas, M, Vyhnalek, M, Brzezny, R, Lhotska, L 2004, ‗School Children

Dyslexia Analysis using Self Organizing Maps ‘, IEEE 26th Annual International Conference, San Francisco, CA, pp. 576-580.

1. Peyrin, C, Lallier, M, Démonet, JF, Pernet, CM, Baciu, JF, Le Bas, S Valdois 2012, ‗Neural dissociation of phonological and visual attention span disorders in developmental dyslexia ‘, FMRI evidence from two case reports, Brain & Language, pp. 182-192.
2. ServetBayram, MücahitCamnalbur & EsadEsgin 2012, ‗Analysis of dyslexic students ‘, reading disorder with eye movement tracking ,Cypriot Journal of Educational Sciences, vol. 7, issue. 2, pp. 129-148
3. SooYeonJi & Kayvan Najarian 2008, ‗A Modified Maximum Correlation Modeling Method for fMRI Brain Mapping‘, Application for Detecting Dyslexia - 978-1-4244-2890-8/08/$25.00 ©2008 IEEE.
4. Soo-YeonJi, Kevin Ward & KayvanNajarian 2009, ‗Brain mapping and detection of functional patterns in fMRI using wavelet transform‘, application in detection of dyslexia - BMC Medical Informatics and Decision Making 2009, 9(Suppl 1):S6 doi:10.1186/1472-6947-9-S1- S6.
5. Stanberry, LI1, Richards, TL, Berninger, VW, Nandy, RR, Aylward, EH, Maravilla, KR, Stock, PS & Cordes, D 2006, Low-frequency signal changes reflect differences in functional connectivity between good readers and dyslexics during continuous phoneme mapping, MagnReson Imaging, vol. 24, no. 3, pp. 217-29.
6. Wu, Tung-Kuang, Shian-Chang Huang & Ying-RuMeng 2008, ‗Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities‘. Expert Systems with Applications, vol. 34, no. 3, pp. 1846-1856

42

**A. SAMPLE CODE**

clc;

clear all;

close all;

fpintf('\*\* DYLEXIA CLASSIFIACTION \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\n'); fprintf('\*\* 12 features with 65 students dataset \*\*\*\*\*\*\*\*\*\*\*\*\*\n'); fprintf('\*\* 4 classification \*\*\*\*\*\*\*\*\*\*\*\*\*\*\n');

fprntf('\*\* Dyslexic, Control and revision, Dyslexic,Inattention hyperactivity or other problems

\*\*\*\*\*\*\*\*\*\n');

fprintf('\n');

fpritf('\n');

fprintf('\n');

fprintf(' \*\*\*\* Read Dyslexic DATASET \*\*\*\*\n');

fprintf('\n');

load dylexia.mat

dylexia=cell2mat(dylexia3);

Varame1 = dylexia(:, 1);

VarName2 = dylexia(:, 2);

VarName3 = dylexia(:, 3);

VarName4 = dylexia(:, 4);

VarName5 = dylexia(:, 5);

VarName6 = dylexia(:, 6);

VarName7 = dylexia(:, 7);

VarName8 = dylexia(:, 8);

VarName9 = dylexia(:, 9);

VarName10 = dylexia(:, 10);

43

VarName11 = dylexia(:, 11);

VarName12 = dylexia(:, 12);

VarName13 = dylexia(:, 13);

fprintf(' \*\*\*\* DATASET SEPARATION \*\*\*\*\n'); fprintf('\n');

%% Clear temporary variables

clearvars filename delimiter formatSpec fileID ans dylexia3; %% Create the nx11 matrix

dataSet = [VarName1, VarName2, VarName3, VarName4, VarName5, VarName6, VarName7, VarName8, VarName9, VarName10, VarName11,VarName12, VarName13]; dataSet = double(dataSet);

labels = dataSet(:, 13);

* Normalise values in [0,1] normalisedData = [];

for k=1:size(dataSet, 2)-1

normalisedData = [normalisedData, (dataSet(:, k) - min(dataSet(:, k))) / (max(dataSet(:, k)) - min(dataSet(:, k)))];

end

% for i = 1:10

% [train,test] = crossvalind('Kfold',labels,10);

% mdl = fitcknn(normalisedData(train,:),labels(train),'NumNeighbors',3); % predictions = predict(mdl,normalisedData(test,:));

% classperf(cp,predictions,test); % end

indices = crossalind('Kfold',labels,10);

%%K Nearest Neighbour(KNN) model training cpKnn = clasperf(labels);

for i = 1:10

test = (indices == i); train = ~test;

44

mdl = ClassificatioKN.fit(normalisedData(train,:),labels(train),'NumNeighbors',5); predictions = predict(mdl,normalisedData(test,:)); classprf(cpKnn,predictions,test);

end

disp(' \*\*\*\*\* KNN CLASSIFICATION \*\*\*\*\*\*'); fprintf('\n');

fprintf('Accuracy: %f\n', 1 - cpKnn.ErrorRate);

KNNAccuracy=(1 - cpKnn.ErrorRate)\*100;

%msgbox(sprintf('KNN accuracy = %f',KNNAccuracy));

fprintf('Sensitivity: %f\n', cpKn.Sensitivity);

KNNSensitivity=cpKn.Sensitivity\*100;

%msgbox(sprintf('KNN Sensitivity = %f',KNNSensitivity));

fprintf('Specificity: %f\n', cpKnn.Specificity);

KNNSpecificity=cpKnn.Specificity\*100;

%msgbox(sprintf('KNN Specificity = %f',KNNSpecificity));

cpKnn.CoutingMatrix

fprintf('\n');

* Decision tree model training cpDTree = clasperf(labels);

for i = 1:10

test = (indices == i); train = ~test;

mdl = ClassificationTree.fit(normalisedData(train,:),labels(train)); predictions = predict(mdl,normalisedData(test,:)); classperf(cpDTree,predictions,test);

end

disp('\*\*\*\*\*\* Decision Tree CLASSIFICATION \*\*\*\*\*\*'); fprintf('Accuracy: %f\n', 1 - cpDTree.ErrorRate); DTAccuracy=(1 - cpDTree.ErrorRate)\*100; %msgbox(sprintf('Decision Tree accuracy = %f',DTAccuracy));

45

fprintf('Sensitivity: %f\n', cpDTree.Sensitivity); DTSensitivity=cpDTree.Sensitivity\*100; %msgbox(sprintf('Decision Tree Sensitivity = %f',DTSensitivity)); fprintf('Specificity: %f\n', cpDTree.Specificity); DTSpecificity=cpDTree.Specificity\*100; %msgbox(sprintf('Decision Tree Specificity = %f',DTSpecificity)); cpDTree.CountigMatrix

fprintf('\n');

* Linear Discriminant Analysis classifier cpRgn = classperf(labels);

for i = 1:10

test = (indices == i); train = ~test; y=labels(train);

mdl = ClassificationDiscriminant.fit(normalisedData(train,:),labels(train)); predictions = predct(mdl,normalisedData(test,:)); classperf(cpRgn,predictions,test);

end

disp('\*\*\*\*\*\* LDA CLASSIFICATION \*\*\*\*\*\*\*\*'); fprintf('Accuracy: %f\n', 1 - cpRgn.ErrorRate); DAAccuracy=(1 - cpRgn.ErrorRate)\*100; %msgbox(sprintf('LDA Accuracy = %f',DAAccuracy)); fprintf('Sensitivity: %f\n', cpRgn.Sensitivity); DASensitivity=cpRgn.Sensitivity\*100; %msgbox(sprintf('LDA Sensitivity = %f',DASensitivity)); fprintf('Specificity: %f\n', cpRgn.Specificity); DASpecificity=cpRgn.Speciicity\*100; %msgbox(sprintf('LDA Specificity = %f',DASpecificity));

46

cpRgn.CountingMatrix

%cm = confusionchart(mdl,labels(train));

fprintf('\n');

* support vector machine cpSvm = classperf(labels); for i = 1:10

test = (indices == i); train = ~test;

SVMstruct = svmtrain(normalisedData(train,:),labels(train));

classes = svmcassify(SVMstruct,normalisedData(test,:)); classperf(cpSvm,classes,test);

end

disp('\*\*\*\*\*\* SVM CLASSIFICATION \*\*\*\*\*\*\*\*');

fprintf('Accuracy: %f\n', 1 - cpSvm.ErrorRate);

SVMAccuracy= (1 - cpSvm.ErrorRate)\*100;

%msgbox(sprintf('SVM Specificity = %f',SVMAccuracy));

fprintf('Sensitivity: %f\n', cpSvm.Sensitivity);

SVMSensitivity=cpSvm.Sensitivity\*100;

%msgbox(sprintf('SVM Sensitivity = %f',SVMSensitivity));

fprintf('Specificity: %f\n', cpSvm.Specificity);

SVMSpecificity=cpSvm.Specificity\*100;

%msgbox(sprintf('SVM Specificity = %f',SVMSpecificity));

cpSvm.CoutingMatrix

%cp=confusionchart(cpSvm.CountingMatrix);

fprintf('\n');

47

%%

* % % Naive Bayes model training
* cpNBayes = classperf(labels);
* for i = 1:11
* test = (indices == i);
* train = ~test;
* mdl = NaiveBayes.fit(normalisedData(train,:),labels(train));
* predictions = predict(mdl,normalisedData(test,:));
* classperf(cpNBayes,predictions,test);
* end
* disp('Naive Bayes:');
* fprintf('Accuracy: %f\n', 1 - cpNBayes.ErrorRate);
* fprintf('Sensitivity: %f\n', cpNBayes.Sensitivity);
* fprintf('Specificity: %f\n', cpNBayes.Specificity);
* cpNBayes.CountingMatrix
* fprintf('\n');

load VarName1

load VarName2

load VarName3

load VarName4

load VarName5

load VarName6

load VarName7

load VarName8

load VarName9

load VarName10

load VarName11

load VarName12

load label

48

* accuracy graph figure1 = figure;

x = [KNNAccuracy; DTAccuracy; DAAccuracy; SVMAccuracy];

bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest')

xlabel('Algorithm');

ylabel('Accuracy (%)')

title('Accuracy')

box off

%fullfile('C:\Users\Desktop\project\A data\SegmentedCharWithBlanks',['figure' num2str(k) '.jpg']))

saveas(figure1,fullfile('performance','Accuracy.jpg'))

* sensitivity graph figure2=figure;

x = [SVMSensitivity; DASensitivity; DTSensitivity; KNNSensitivity]; bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest') xlabel('Algorithm');

ylabel('Sensitivity (%)') title('Sensitivity')

box off

saveas(figure2,fullfile('performance','Sensitivity.jpg'))

figure3=figure;

x = [SVMSpecificity; DASpecificity; DTSpecificity; KNNSpecificity]; bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest') xlabel('Algorithm');

ylabel('Specificity (%)') title('Specificity')

49

box off

saveas(figure3,fullfile('performance','Specificity.jpg'))

**4.2 Performance evaluation**

**4.2.1 Sensitivity**

Sensitivity is also called the True Positive Rate (TPR). It evaluates the percentage of actual positive which are dyslexic subjects class. It is the representation of acceptably classified percentage of dyslexic subjects. The sensitivity is defined as below:



**4.2.2 Specificity**

Specificity evaluates the percentage of actual negatives which are related to non-dyslexic subjects class. Specificity is also called as True Negative Rate (TNR). It is the accurateness in identifying non-dyslexic subjects and defines the percentage of correctly classified non-dyslexic subjects.



**4.2.3 Accuracy**

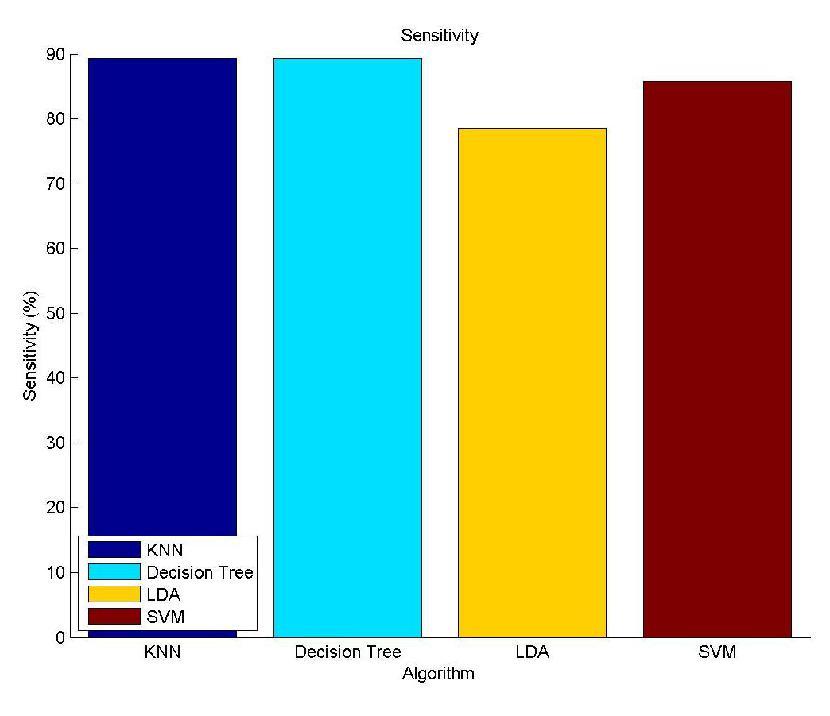
Accuracy is defined as the overall correctness of the model and is calculated as the sum of actual classification parameters (Tp + Tn) separated by the total number of classification parameters (Tp + Tn + Fp + Fn)



To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.1 and performance parameters final values. Total Sensitivity of SVM value is 89.28 %, the conclusion was that the model has performed quite well on this independent dataset. Figure 4.4 also display the graph results for proposed system comparatively K Nearest Neighbour, Decision tree, and Linear Discriminant Analysis.

34

**Table 4.2 Algorithm models compression results**

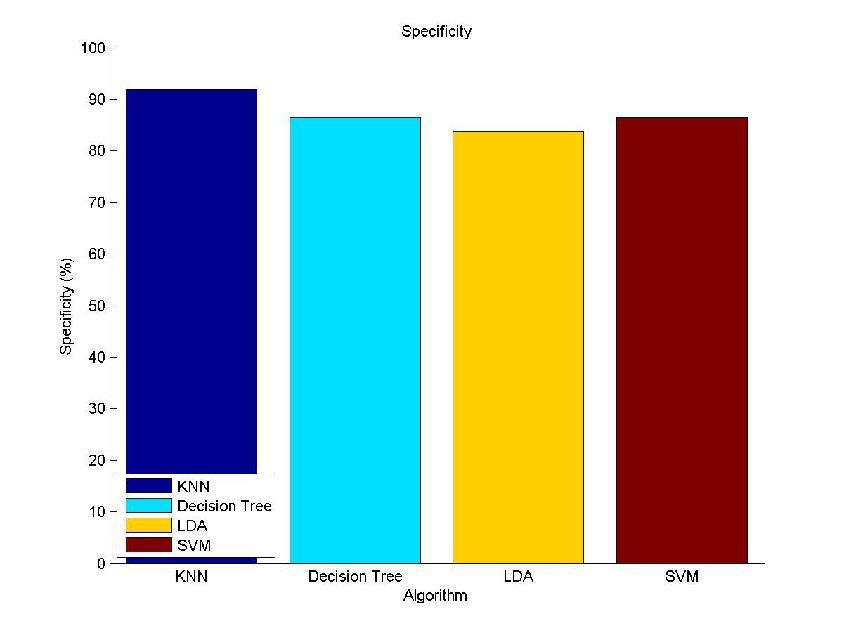


**Figure 4.4 sensitivity compression results**

To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.2 and performance parameters final values. Total Specificity of Decision tree value is 95.04 %, the conclusion was svm that the model has performed quite well on this independent dataset comparatively K Nearest Neighbour, SVM, and Cnn. Figure 4.5 also display the graph results for proposed system proved the better for dyslexia 12-4 imbalanced Dataset.

35

**Table 4.3 Algorithm models compression results**



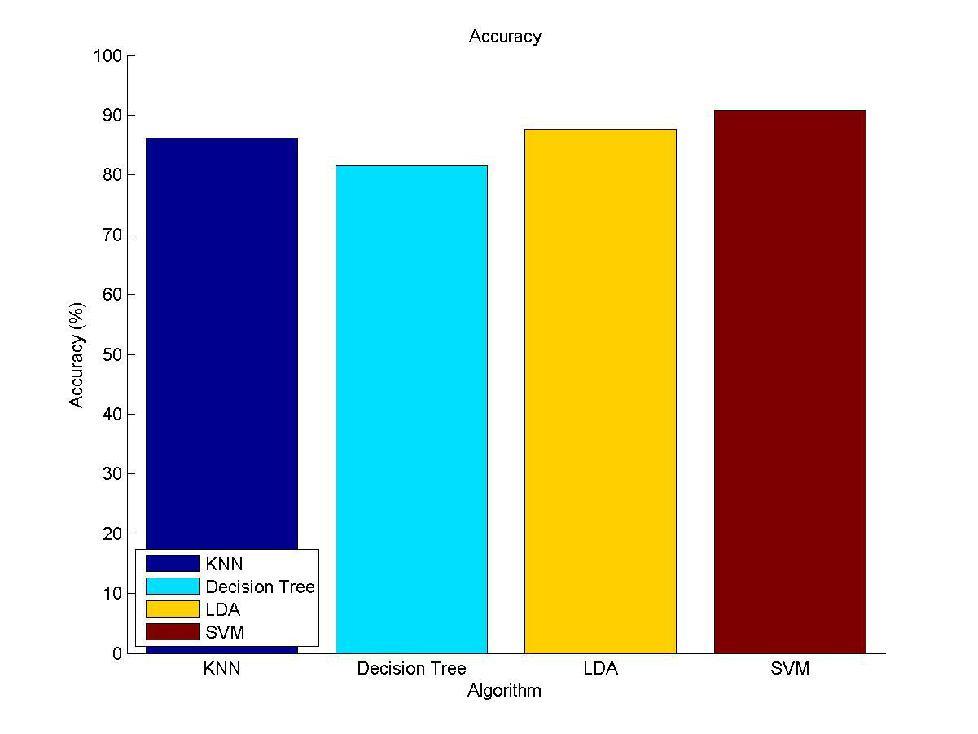
**Figure 4.5 specificity compression results**

|  | **Algorithm** | **Accuracy /Efficiency** |
| --- | --- | --- |
|  | |  |
| K Nearest Neighbour | | 82.97 |
|  | |  |
| Decision tree | | **95.04** |
|  |  |  |
| Cnn |  | 94.73 |
|  |  |  |
|  | |  |
| Support Vector Machine | | 90.7692 |
|  |  |  |

**Table 4.4 Algorithm models Accuracy compression results**

36

To obtain the discrimination power and the feature set of dyslexia 12-4 Dataset users are noticed with respect to the generated values as shown in the table 4.3 and performance parameters final values. Total accuracy of SVM value is 90.76%, the conclusion was svm that the model has performed quite well on this independent dataset comparatively K Nearest Neighbour, Decision tree, and Linear Discriminant Analysis. Figure 4.6 also display the graph results for proposed system.



**Figure 4.6 Accuracy compression results**

The performance of SVM method was compared with that described in previous chapter existing algorithms. Prediction results of the data mining and statistical learning algorithms compressions are shown in Figure 4.4 4.5 and 4.6. In existing method compression KNN algorithm is showed slight improvement when applied to the independent dataset. Comparatively KNN algorithm slightly improves and best result for other two algorithms (Decision tree and Cnn) in various parameters (accuracy, specificity, sensitivity). Based on the result algorithms are ranked as Decision tree, Cnn, Support Vector Machine, KNN

37

**4.3 SUMMARY**

This section shows the results obtained from various classification methods applied for dyslexia 12-4 KEEL dataset. The experimentation results of various classification methods have been carried out with the help of six parameters such as sensitivity, specificity, accuracy, and precision,. All these evaluation results conclude that the proposed Decision tree method provides higher detection accuracy than KNN, SVM and Cnn.

38

**CHAPTER 5**

**CONCLUSION**

39

**5.1 CONCLUSION**

Identification of dyslexia affected persons is a problem today in the medical field as diagnosis is a difficult task. Through this work the researcher is attempting to make the diagnosis easy for the problem mostly prevalent in developing countries like India. Then the dyslexia 12-4 KEEL dataset features are passed through a various machine learning classifier like support vector machine, k-nearest neighbour Decision tree and Linear Discriminant Analysis classifiers. The results emanating out of the classifier are evaluated. The evaluation is based on parameters like sensitivity, specificity, and accuracy. This method is not only simple to manage large data simultaneously but also yields good results than previous works. This study is beneficial to physicians for easy and effective diagnosis of learning disability. Academics especially computer scientists can replicate the suggested models to other fields too.

Prediction results of the data mining and statistical learning algorithms compressions are discussed in chapter 4. In machine learning methods compression KNN algorithm is showed slight improvement when applied to the independent dataset. Comparatively KNN algorithm slightly improves and best result for other two algorithms (Decision tree and Linear Discriminant Analysis) in various parameters (accuracy, specificity, sensitivity). Based on the result algorithms are ranked as Support Vector Machine, KNN, Decision tree and Linear Discriminant Analysis.

**5.2 FUTURE WORK**

With the experience gained through this work, the researcher is suggesting to fellow researchers to focus on early diagnosis of learning disability by applying soft computing techniques. Thereby the impact of dyslexia on per-school children can very well be diagnosed and that too at the right time to contain the stress due to unrecognized disorder, such research endeavours will be beneficial to those children suffering due to dyslexia, teachers engaged in special education, and physicians. To be precise, the focus of future research is to identify early diagnosis of dyslexia and proper classification of learning disability to support the special education community in their quest

to be with mainstream

40

**REFERENCES**

1. Aylward, EH, Richards, TL, Berninger, VW, Nagy, WE, Field, KM, Grimme; AC, Richards, AL, Thomson, JB & Cramer, SC 2008, ‗Instructional treatment associated with changes in brain activation in children with dyslexia ‘, Educational Research, vol. 50, no. 2, pp. 135–148.
2. Biscaldi, M, Fischer, B & Aiple, F 1994, ‗Saccadic eye movements of dyslexic and normal reading children Perception ‘, vol. 23, no. 1, pp. 45 – 64.
3. ChloéPradoa, Matthieu Dubois & SylvianeValdois, ‗The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span ‘, Vision Research, vol. 47, issue. 19, pp. 2521–2530.
4. Eden, GF, Stein, JF, Wood, MH & Wood, FB 1994, ‗Differences in Eye Movements and Reading Problems in Dyslexic and Normal Children, Vision Research‘, vol. 34, no. 10, pp. 1345-1358.
5. Evgenia Hristova, Alexander Gerganov, Ekaterina Todorova & Severina Georgieva 2006, ‗Eye-Movements of Dyslexic Children Reading in Regular Orthography: Exploring Word Frequency and

Length Effects‘, Exploring Word Frequency and Length Effects, pp. 56-70.

1. Hawelka, S & Wimmer, H 2005, ‗Impaired Visual Processing of Multi-Element Arrays is Associated with Increased Number of Eye ovements in Dyslexic Reading‘, Vision Research, vol. 45, no. 1, pp. 855-863.
2. Hultquist, AM 2006, ‗An Introduction to Dyslexia for Parents and Professionals‘. London: Jessica

Kingsley Publishers.

1. Hutzler, F & Wimmer, H 2004, ‗Eye movements of dyslexic children when reading in a regular orthography‘, Brain and Language, vol. 89, no. 1, pp. 235–242.
2. Hutzler, F, Kronbichler, M, Jacobs, AM & Wimmer, H 2006. ‗Perhaps correlational but not causal:

No effect of dyslexic readers ‘magnocellular system on their eye movements during reading. Neuropsychologia, vol. 44, no. 4, pp. 637–648.

1. Karstenspecht, Kenneth hugdahl, Sonja ofte, Merethenygård, Atlebjørnerud, Elena plante & Turidhelland 2008, ‗Brain activation on pre-reading tasks reveals at-risk status for dyslexia in 6-year-old children ‘, Scandinavian Journal of Psychology.

41

1. Kerstin I Schattka, Ralph Radach & Walter Huber 2010, ‗Eye movement correlates of acquired central dyslexia Eye movement correlates of acquired central dyslexia ‘, Neuropsychologia, vol. 48, pp. 2959–2973.
2. Lei yu 2004, ‗Efficient Feature Selection via Analysis of Relevance and Redundancy ‘, Journal of Machine Learning Research, 1205–1224 Submitted 1/04; Revised 5/04; Published 10/04.
3. Maria De Luca, Enrico Di Pace, Anna Judica, Donatella Spinelli, Pierluigi Zoccolotti 1999, ‗Eye movement patterns in linguistic and non-linguistic tasks indevelopmental surface dyslexia ‘, Neuropsychologia, vol. 37, pp. 1407-1420
4. Novak, D, Kordk, P, Macas, M, Vyhnalek, M, Brzezny, R, Lhotska, L 2004, ‗School Children

Dyslexia Analysis using Self Organizing Maps ‘, IEEE 26th Annual International Conference, San Francisco, CA, pp. 576-580.

1. Peyrin, C, Lallier, M, Démonet, JF, Pernet, CM, Baciu, JF, Le Bas, S Valdois 2012, ‗Neural dissociation of phonological and visual attention span disorders in developmental dyslexia ‘, FMRI evidence from two case reports, Brain & Language, pp. 182-192.
2. ServetBayram, MücahitCamnalbur & EsadEsgin 2012, ‗Analysis of dyslexic students ‘, reading disorder with eye movement tracking ,Cypriot Journal of Educational Sciences, vol. 7, issue. 2, pp. 129-148
3. SooYeonJi & Kayvan Najarian 2008, ‗A Modified Maximum Correlation Modeling Method for fMRI Brain Mapping‘, Application for Detecting Dyslexia - 978-1-4244-2890-8/08/$25.00 ©2008 IEEE.
4. Soo-YeonJi, Kevin Ward & KayvanNajarian 2009, ‗Brain mapping and detection of functional patterns in fMRI using wavelet transform‘, application in detection of dyslexia - BMC Medical Informatics and Decision Making 2009, 9(Suppl 1):S6 doi:10.1186/1472-6947-9-S1- S6.
5. Stanberry, LI1, Richards, TL, Berninger, VW, Nandy, RR, Aylward, EH, Maravilla, KR, Stock, PS & Cordes, D 2006, Low-frequency signal changes reflect differences in functional connectivity between good readers and dyslexics during continuous phoneme mapping, MagnReson Imaging, vol. 24, no. 3, pp. 217-29.
6. Wu, Tung-Kuang, Shian-Chang Huang & Ying-RuMeng 2008, ‗Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities‘. Expert Systems with Applications, vol. 34, no. 3, pp. 1846-1856

42

**A. SAMPLE CODE**

clc;

clear all;

close all;

fpintf('\*\* DYLEXIA CLASSIFIACTION \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\n'); fprintf('\*\* 12 features with 65 students dataset \*\*\*\*\*\*\*\*\*\*\*\*\*\n'); fprintf('\*\* 4 classification \*\*\*\*\*\*\*\*\*\*\*\*\*\*\n');

fprntf('\*\* Dyslexic, Control and revision, Dyslexic,Inattention hyperactivity or other problems

\*\*\*\*\*\*\*\*\*\n');

fprintf('\n');

fpritf('\n');

fprintf('\n');

fprintf(' \*\*\*\* Read Dyslexic DATASET \*\*\*\*\n');

fprintf('\n');

load dylexia.mat

dylexia=cell2mat(dylexia3);

Varame1 = dylexia(:, 1);

VarName2 = dylexia(:, 2);

VarName3 = dylexia(:, 3);

VarName4 = dylexia(:, 4);

VarName5 = dylexia(:, 5);

VarName6 = dylexia(:, 6);

VarName7 = dylexia(:, 7);

VarName8 = dylexia(:, 8);

VarName9 = dylexia(:, 9);

VarName10 = dylexia(:, 10);

43

VarName11 = dylexia(:, 11);

VarName12 = dylexia(:, 12);

VarName13 = dylexia(:, 13);

fprintf(' \*\*\*\* DATASET SEPARATION \*\*\*\*\n'); fprintf('\n');

%% Clear temporary variables

clearvars filename delimiter formatSpec fileID ans dylexia3; %% Create the nx11 matrix

dataSet = [VarName1, VarName2, VarName3, VarName4, VarName5, VarName6, VarName7, VarName8, VarName9, VarName10, VarName11,VarName12, VarName13]; dataSet = double(dataSet);

labels = dataSet(:, 13);

* Normalise values in [0,1] normalisedData = [];

for k=1:size(dataSet, 2)-1

normalisedData = [normalisedData, (dataSet(:, k) - min(dataSet(:, k))) / (max(dataSet(:, k)) - min(dataSet(:, k)))];

end

% for i = 1:10

% [train,test] = crossvalind('Kfold',labels,10);

% mdl = fitcknn(normalisedData(train,:),labels(train),'NumNeighbors',3); % predictions = predict(mdl,normalisedData(test,:));

% classperf(cp,predictions,test); % end

indices = crossalind('Kfold',labels,10);

%%K Nearest Neighbour(KNN) model training cpKnn = clasperf(labels);

for i = 1:10

test = (indices == i); train = ~test;

44

mdl = ClassificatioKN.fit(normalisedData(train,:),labels(train),'NumNeighbors',5); predictions = predict(mdl,normalisedData(test,:)); classprf(cpKnn,predictions,test);

end

disp(' \*\*\*\*\* KNN CLASSIFICATION \*\*\*\*\*\*'); fprintf('\n');

fprintf('Accuracy: %f\n', 1 - cpKnn.ErrorRate);

KNNAccuracy=(1 - cpKnn.ErrorRate)\*100;

%msgbox(sprintf('KNN accuracy = %f',KNNAccuracy));

fprintf('Sensitivity: %f\n', cpKn.Sensitivity);

KNNSensitivity=cpKn.Sensitivity\*100;

%msgbox(sprintf('KNN Sensitivity = %f',KNNSensitivity));

fprintf('Specificity: %f\n', cpKnn.Specificity);

KNNSpecificity=cpKnn.Specificity\*100;

%msgbox(sprintf('KNN Specificity = %f',KNNSpecificity));

cpKnn.CoutingMatrix

fprintf('\n');

* Decision tree model training cpDTree = clasperf(labels);

for i = 1:10

test = (indices == i); train = ~test;

mdl = ClassificationTree.fit(normalisedData(train,:),labels(train)); predictions = predict(mdl,normalisedData(test,:)); classperf(cpDTree,predictions,test);

end

disp('\*\*\*\*\*\* Decision Tree CLASSIFICATION \*\*\*\*\*\*'); fprintf('Accuracy: %f\n', 1 - cpDTree.ErrorRate); DTAccuracy=(1 - cpDTree.ErrorRate)\*100; %msgbox(sprintf('Decision Tree accuracy = %f',DTAccuracy));

45

fprintf('Sensitivity: %f\n', cpDTree.Sensitivity); DTSensitivity=cpDTree.Sensitivity\*100; %msgbox(sprintf('Decision Tree Sensitivity = %f',DTSensitivity)); fprintf('Specificity: %f\n', cpDTree.Specificity); DTSpecificity=cpDTree.Specificity\*100; %msgbox(sprintf('Decision Tree Specificity = %f',DTSpecificity)); cpDTree.CountigMatrix

fprintf('\n');

* Linear Discriminant Analysis classifier cpRgn = classperf(labels);

for i = 1:10

test = (indices == i); train = ~test; y=labels(train);

mdl = ClassificationDiscriminant.fit(normalisedData(train,:),labels(train)); predictions = predct(mdl,normalisedData(test,:)); classperf(cpRgn,predictions,test);

end

disp('\*\*\*\*\*\* LDA CLASSIFICATION \*\*\*\*\*\*\*\*'); fprintf('Accuracy: %f\n', 1 - cpRgn.ErrorRate); DAAccuracy=(1 - cpRgn.ErrorRate)\*100; %msgbox(sprintf('LDA Accuracy = %f',DAAccuracy)); fprintf('Sensitivity: %f\n', cpRgn.Sensitivity); DASensitivity=cpRgn.Sensitivity\*100; %msgbox(sprintf('LDA Sensitivity = %f',DASensitivity)); fprintf('Specificity: %f\n', cpRgn.Specificity); DASpecificity=cpRgn.Speciicity\*100; %msgbox(sprintf('LDA Specificity = %f',DASpecificity));

46

cpRgn.CountingMatrix

%cm = confusionchart(mdl,labels(train));

fprintf('\n');

* support vector machine cpSvm = classperf(labels); for i = 1:10

test = (indices == i); train = ~test;

SVMstruct = svmtrain(normalisedData(train,:),labels(train));

classes = svmcassify(SVMstruct,normalisedData(test,:)); classperf(cpSvm,classes,test);

end

disp('\*\*\*\*\*\* SVM CLASSIFICATION \*\*\*\*\*\*\*\*');

fprintf('Accuracy: %f\n', 1 - cpSvm.ErrorRate);

SVMAccuracy= (1 - cpSvm.ErrorRate)\*100;

%msgbox(sprintf('SVM Specificity = %f',SVMAccuracy));

fprintf('Sensitivity: %f\n', cpSvm.Sensitivity);

SVMSensitivity=cpSvm.Sensitivity\*100;

%msgbox(sprintf('SVM Sensitivity = %f',SVMSensitivity));

fprintf('Specificity: %f\n', cpSvm.Specificity);

SVMSpecificity=cpSvm.Specificity\*100;

%msgbox(sprintf('SVM Specificity = %f',SVMSpecificity));

cpSvm.CoutingMatrix

%cp=confusionchart(cpSvm.CountingMatrix);

fprintf('\n');

47

%%

* % % Naive Bayes model training
* cpNBayes = classperf(labels);
* for i = 1:11
* test = (indices == i);
* train = ~test;
* mdl = NaiveBayes.fit(normalisedData(train,:),labels(train));
* predictions = predict(mdl,normalisedData(test,:));
* classperf(cpNBayes,predictions,test);
* end
* disp('Naive Bayes:');
* fprintf('Accuracy: %f\n', 1 - cpNBayes.ErrorRate);
* fprintf('Sensitivity: %f\n', cpNBayes.Sensitivity);
* fprintf('Specificity: %f\n', cpNBayes.Specificity);
* cpNBayes.CountingMatrix
* fprintf('\n');

load VarName1

load VarName2

load VarName3

load VarName4

load VarName5

load VarName6

load VarName7

load VarName8

load VarName9

load VarName10

load VarName11

load VarName12

load label

48

* accuracy graph figure1 = figure;

x = [KNNAccuracy; DTAccuracy; DAAccuracy; SVMAccuracy];

bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest')

xlabel('Algorithm');

ylabel('Accuracy (%)')

title('Accuracy')

box off

%fullfile('C:\Users\Desktop\project\A data\SegmentedCharWithBlanks',['figure' num2str(k) '.jpg']))

saveas(figure1,fullfile('performance','Accuracy.jpg'))

* sensitivity graph figure2=figure;

x = [SVMSensitivity; DASensitivity; DTSensitivity; KNNSensitivity]; bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest') xlabel('Algorithm');

ylabel('Sensitivity (%)') title('Sensitivity')

box off

saveas(figure2,fullfile('performance','Sensitivity.jpg'))

figure3=figure;

x = [SVMSpecificity; DASpecificity; DTSpecificity; KNNSpecificity]; bar(diag(x),'stacked')

set(gca,'xticklabel',{'KNN','Decision Tree','LDA','SVM'});

legend({'KNN','Decision Tree','LDA','SVM'}, 'Location','southwest') xlabel('Algorithm');

ylabel('Specificity (%)') title('Specificity')

49

box off

saveas(figure3,fullfile('performance','Specificity.jpg'))

50

51