A STATISTICAL AND MACHINE LEARNING APPROACH TO EARLY DETECTION AND AFFECTED AREA PREDICTION IN NEUROMUSCULAR DISORDERS

PROJECT REPORT

Submitted by

DEEPANEESH R V

(23MST005)

Under the Guidance of

Dr.V.SANGEETHA
Assistant Professor

In partial fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE IN STATISTICS

Bharathiar University, Coimbatore.

DEPARTMENT OF STATISTICS
PSG COLLEGE OF ARTS & SCIENCE
An autonomous college-Affiliated to Bharathiar University
Re-accredited with A++ Grade by NAAC (4th cycle)
College with Potential for Excellence
(Status awarded by the UGC)
Star College Status Awarded by DBT-MST
An ISO 9001:2015 Certified Institution
Coimbatore-641 014

April 2025

CERTIFICATE

CERTIFICATE

This is to certify that the project entitled **"A STATISTICAL AND MACHINE LEARNING APPROACH TO EARLY DETECTION AND AFFECTED AREA PREDICTION IN NEUROMUSCULAR DISORDERS"** submitted to the **PSG COLLEGE OF ARTS & SCIENCE, Coimbatore-14**, a partial fulfilment of the requirements for the award of degree in **MASTER OF SCIENCE IN STATISTICS**, a record of original research work done by **DEEPANEESH R V (23MST005)** during the period of his study in the **DEPARTMENT OF STATISTICS, PSG COLLEGE OF ARTS & SCIENCE**, Coimbatore-14 of the Bharathiar University under my supervision and guidance.

PSG COLLEGE OF ARTS & SCIENCE ★ DEPARTMENT OF STATISTICS ★

ATTESTED

CERTIFIED

15/04/2025

Head of the Department

HEAD
DEPARTMENT OF STATISTICS
PSG COLLEGE OF ARTS & SCIENCE
COIMBATORE - 641 014

Faculty guide

Dr.V. SANGEETHA M.Sc., M.Phil., Ph.D.,
Assistant Professor
Department of Statistics
PSG College of Arts & Science
Coimbatore - 641 014.

17/04/25

Internal Examiner

17/04/25

External Examiner

DECLARATION

DECLARATION

I, DEEPANEESH R V hereby declare that the project entitled "A STATISTICAL AND MACHINE LEARNING APPROACH TO EARLY DETECTION AND AFFECTED AREA PREDICTION IN NEUROMUSCULAR DISORDERS" submitted to the PSG COLLEGE OF ARTS & SCIENCE in partial fulfilment of the requirement for the award of degree of MASTER OF SCIENCE IN STATISTICS is a record of original research work done by me under the supervision and guidance of Dr.V.SANGEETHA, ASSISTANT PROFESSOR, DEPARTMENT OF STATISTICS, PSG COLLEGE OF ARTS & SCIENCE, Coimbatore-14.

Date: 17.04.2025

Place: Coimbatore

Signature of the Student

(DEEPANEESH R V)

ACKNOWLEDGEMENT

ACKNOWLEDGEMENT

An acknowledgement is not a formality, but it is the greatest possible way through which I express my gratefulness whole heartedly to the following persons who are responsible for the successful completion of this project work.

I avail this opportunity to thank my respected guide **Dr.V.SANGEETHA, Assistant Professor,** Department of Statistics, PSG College of Arts & Science, Coimbatore-14 for her encouragement, kindness, valuable suggestions, and moral support for completion of the project.

I am grateful to **Dr.G.UMA**, **Associate Professor & Head of the Department of Statistics,** PSG College of Arts & Science, Coimbatore-14, for her efficiency, diligence and discipline that helped me to carry out all my endeavours. I would like to express my deep sense of gratitude for her constant support and encouragement during the entire span of my project.

I would like to place on my thanks to **Dr. M. SENGUTTUVAN**, Principal, PSG College of Arts and Science, Coimbatore-14 for permitting to do this study.

I am greatly indebted to **Dr. T. KANNAIAN**, Secretary, of PSG College of Arts & Science, Coimbatore-14 for permitting me to do this study.

I express my sincere thanks to the entire Faculty members, Lab technicians and those who directly and indirectly helped me in the successful completion of this project.

I wish to express my gratitude to my family members and friends who are great support to my success.

DEEPANEESH R V

CONTENTS

CONTENTS

INTRODUCTION

CHAPTER I

INTRODUCTION

1.1 Neuromuscular Disorders

Neuromuscular disorders comprise a diverse array of conditions that adversely affect the nerves, muscles, and the intricate communication pathways between them. These disorders can lead to a spectrum of challenges, including progressive muscle weakness, chronic pain, and significant difficulties in movement and coordination. Unlike temporary muscle fatigue, which often resolves with rest, neuromuscular diseases are typically chronic in nature, with potential for gradual worsening over time if not adequately managed. Common symptoms encompass muscle weakness, numbness, loss of balance and coordination, difficulty swallowing or breathing, chronic pain, and muscle atrophy.

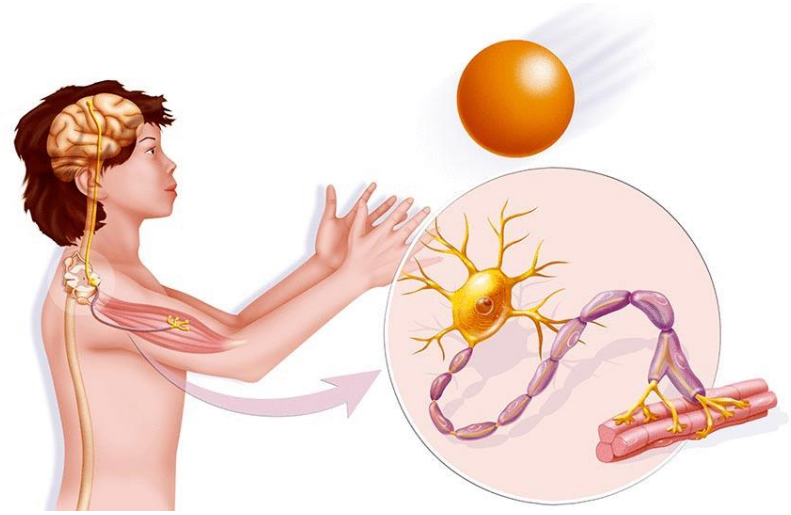

Figure 1.1

The etiologies of neuromuscular disorders are multifactorial, arising from a complex interplay of genetic, autoimmune, metabolic, and degenerative factors. For example, genetic conditions like muscular dystrophy are inherited and result in progressive muscle degeneration due to mutations affecting muscle proteins. In contrast, autoimmune disorders such as myasthenia gravis arise when the body's immune system mistakenly targets the neuromuscular junction, disrupting the communication between nerves and muscles. Diseases like amyotrophic lateral sclerosis (ALS) involve the progressive degeneration of motor neurons, leading to profound disability and impacting not only physical functions but also respiratory capabilities. Environmental factors, including exposure to certain toxins, drugs, or viruses, can also play a critical role in the onset and progression of these disorders.

The impact of neuromuscular disorders is substantial, affecting millions of individuals globally and carrying far-reaching implications for both patients and society as a whole. Many affected individuals experience severe disabilities that impede their ability to engage in daily activities, thereby diminishing their overall quality of life and independence. If left untreated, these conditions can lead to further physical decline, reducing the ability to work, participate in social interactions, and maintain a fulfilling lifestyle. Comprehensive management often requires long-term medical care, physical therapy, the use of assistive devices, and rehabilitative services to alleviate symptoms and maintain function

These disorders can manifest at any age, impacting both pediatric and adult populations. Certain conditions, however, show distinct age-related prevalence; for instance, Duchenne muscular dystrophy primarily affects young boys, while conditions such as ALS and peripheral neuropathy are more frequently diagnosed in the aging population. Among those affected, complications involving muscle and nerve function are not uncommon, leading to joint issues and secondary problems stemming from prolonged muscle weakness and immobility.

The muscle groups most commonly impacted by neuromuscular diseases include those in the hips, thighs, shoulders, and upper arms—regions critical for mobility and stability. In conditions like Guillain-Barré syndrome and Charcot-Marie-Tooth disease, peripheral nerves become compromised, resulting in symptoms such as numbness, tingling, and weakness, which can further complicate everyday tasks. Additionally, the neuromuscular junctions, where muscle fibers and nerves connect, are often disrupted, leading to further complications in muscular control. While joint problems are typically not the primary concern in neuromuscular disorders, they can develop due to protracted periods of muscle weakness, causing stiffness, pain, and eventual deformities over time.

Early diagnosis and timely intervention are crucial in managing neuromuscular disorders effectively. Healthcare professionals employ a battery of diagnostic tools, including genetic testing, electromyography (EMG), nerve conduction studies, and muscle biopsies, to identify the specific condition and guide treatment. While many neuromuscular disorders currently lack a definitive cure, a variety of treatment options—such as pharmacological interventions, physical therapy, occupational therapy, and adaptive equipment—can significantly improve patients' quality of life and help slow the progression of the disease. Emerging research in gene therapy and regenerative medicine holds promise for future breakthroughs in treatment

neuromuscular disorders present a considerable health challenge, contributing to progressive muscle weakness, nerve damage, and mobility challenges for millions. Despite these hurdles, advancements in medical research and patient-centered care continue to enhance the outlook for those affected. With early intervention, tailored management strategies, and ongoing scientific innovation, there exists hope for improved treatments and a better quality of life for individuals grappling with neuromuscular disorders.

1.2 History of Neuromuscular Disorders

Neuromuscular disorders are a diverse group of medical conditions that affect the muscles and the nerves that control them, significantly impacting motor function and overall quality of life. The exploration of these complex disorders began in the early 19th century when pioneering neurologists and physiologists laid the groundwork for our understanding of muscle physiology and the nervous system's role in motor control.

One of the first significant contributions to this field was made by **Sir Charles Bell**, a British anatomist and surgeon, who, in the early 1800s, identified the critical link between nerve injuries and resulting muscle weakness. His work marked a turning point in the study of neuromuscular relationships, asserting that damage to peripheral nerves can lead to observable impairments in muscle function. Over the remainder of the 19th century, notable figures such **as Jean-Martin Charcot—often referred to as the father of modern neurology—and William Gowers** enhanced our understanding by classifying various neuromuscular diseases and elucidating the characteristics of motor neuron diseases like amyotrophic lateral sclerosis (ALS).

As medical science advanced into the 20th century, the field began to benefit from technological developments that improved diagnostic capabilities. The evolution of electrophysiology, particularly the adoption of electromyography (EMG), proved invaluable in assessing the electrical activity of muscles and diagnosing neuromuscular disorders with greater precision. This era also witnessed the identification of genetic mutations linked to hereditary neuromuscular diseases, including various forms of muscular dystrophy and myopathy. These discoveries allowed researchers to delve deeper into the molecular pathophysiology of these conditions, fostering a clearer understanding of the underlying mechanisms that drive muscle degeneration and dysfunction. The late 20th century and early 21st century ushered in a revolutionary era in medicine with advancements in molecular biology and genetic testing. This evolution dramatically transformed approaches to diagnosing

neuromuscular disorders, enabling the identification of specific genetic markers and mutations that are responsible for many hereditary conditions. Additionally, these developments have paved the way for innovations in treatment options, including gene therapies targeting the root causes of disorders and personalized medicine approaches tailored to the individual genetic profile of patients. Today, researchers are at the forefront of an exciting phase of investigation into the genetic, biochemical, and immunological factors contributing to neuromuscular disorders. With ongoing studies aimed at discovering targeted therapies, the hope is to not only improve patient outcomes but also enhance the quality of life for individuals affected by these challenging conditions. The collaborative efforts across various fields, including genetics, neurology, and rehabilitation science, continue to hold promise for breakthroughs that could change the landscape of treatment and management for neuromuscular disorders.

1.3 Major Functional Challenges in Neuromuscular Disorders

Neuromuscular disorders impact various components of the musculoskeletal and nervous systems, resulting in a range of complications. The major issues can be categorized into problems related to muscles, nerves, joints, and the combined dysfunction of muscles and nerves.

1.3.1 Muscle-Related Issues

Muscle dysfunction is one of the primary concerns in neuromuscular disorders, leading to progressive weakness and reduced mobility. Patients often experience muscle atrophy (wasting), cramps, and stiffness, which make simple movements difficult. Over time, the loss of muscle strength affects posture, balance, and coordination, increasing the risk of falls and injuries. In severe cases, respiratory muscles become weak, leading to breathing difficulties and the need for ventilatory support. Conditions such as muscular dystrophy, polymyositis, and myopathies are primarily muscle-related and require long-term physical therapy, medications, and assistive devices to maintain mobility and function.

1.3.2 Nerve-Related Issues

Nerve damage in neuromuscular disorders disrupts the transmission of signals between the brain, spinal cord, and muscles. This leads to symptoms such as numbness, tingling, burning sensations, and loss of reflexes. Motor nerves that control movement may degenerate, causing paralysis or severe muscle weakness, while sensory nerve damage may result in decreased sensation, making patients prone to unnoticed injuries. Autonomic nerve dysfunction can

further lead to problems with heart rate regulation, digestion, and temperature control. Disorders like peripheral neuropathy, amyotrophic lateral sclerosis (ALS), and Charcot-Marie-Tooth disease primarily affect the nervous system and require nerve-targeting treatments, pain management, and rehabilitation therapy.

1.3.3 Joint-Related Issues

Joint complications arise due to muscle weakness, imbalances, and contractures, where muscles or tendons become permanently tight. This leads to joint stiffness, pain, deformities, and restricted range of motion. Over time, limited mobility increases the risk of joint degeneration, arthritis, and postural abnormalities, such as scoliosis or kyphosis. These issues make daily activities like standing, walking, or even sitting uncomfortable and challenging. Patients with conditions like spinal muscular atrophy, muscular dystrophy, and arthrogryposis often require orthopedic interventions, including braces, corrective surgery, and regular physiotherapy to maintain joint flexibility and function.

1.3.4 Combined Muscle and Nerve Dysfunction

Many neuromuscular disorders involve both muscle and nerve damage, leading to a combination of motor impairment, sensory loss, and autonomic dysfunction. When nerve signals fail to reach the muscles properly, it results in progressive weakness, fatigue, and impaired coordination. Conditions such as myasthenia gravis, Guillain-Barré syndrome, and ALS cause difficulties in movement, speech, swallowing, and even breathing. These disorders often require a multidisciplinary approach involving medications, immunotherapy, physical therapy, and assistive devices to slow disease progression and improve quality of life.

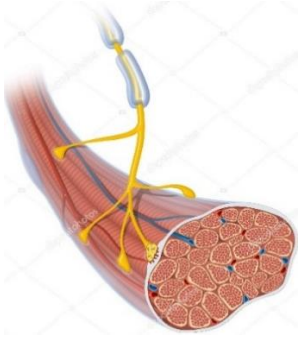

Figure 1.2

1.4 Prevalence and Influencing Factors of Neuromuscular Disorders

Neuromuscular disorders encompass a wide variety of conditions that can affect individuals of all ages, genders, and backgrounds, significantly impacting muscle function and overall health. These disorders can arise in any population; however, certain types exhibit a higher prevalence and distinct characteristics based on demographic factors such as age, genetics, and hormonal influences. For instance, Duchenne muscular dystrophy (DMD)—a severe and progressive form of muscular dystrophy—typically manifests in early childhood, predominantly affecting males due to its X-linked genetic inheritance. DMD is characterized by gradual muscle weakness and degeneration, which can lead to profound mobility challenges and a decreased quality of life as the condition progresses. As affected children grow, they may experience difficulties with activities such as walking and eventually require wheelchairs, greatly impacting their physical health and independence. In contrast, disorders like amyotrophic lateral sclerosis (ALS), a progressive neurodegenerative disease that affects motor neurons in the brain and spinal cord, usually manifest in adults, with onset typically ranging from middle age to later ALS affects both men and women, epidemiological studies have shown that men are diagnosed more frequently than women, highlighting potential gender-specific risk factors.

Notably, gender differences significantly influence the prevalence and clinical presentation of various neuromuscular diseases. Myasthenia gravis, an autoimmune disorder characterized by weakness and rapid fatigue of voluntary muscles, is more frequently diagnosed in females, particularly among younger women. Despite its higher incidence in this demographic, the disorder can also affect older individuals and males, indicating a complex interplay of factors that determines its distribution across different populations. Additionally, genetic disorders such as myotonic dystrophy, which affects both males and females, can present with different patterns of severity and age of onset based on gender.

Research suggests that males often exhibit symptoms earlier and with greater intensity than their female counterparts, underscoring the importance of considering both genetic predispositions and gender in the evaluation of neuromuscular conditions. Genetic susceptibility plays a crucial role in the development of many neuromuscular disorders, especially hereditary forms like various muscular dystrophies. Individuals with a family history of these conditions are often at a significantly heightened risk, and genetic counseling may provide essential insights for those affected or at risk. Beyond genetic factors, environmental influences are also pivotal in the onset of acquired neuromuscular diseases. Exposure to toxic

substances, such as heavy metals or certain chemicals, physical trauma from accidents, and viral infections can serve as triggers for these conditions, complicating the landscape of diagnosis and prognosis. Ultimately, the interplay of age, gender, genetic predisposition, and environmental factors creates a multifaceted framework that healthcare professionals must navigate when diagnosing and treating neuromuscular disorders. Gaining a deeper understanding of these nuances is essential for providing targeted interventions that not only address the specific needs of patients but also improve overall outcomes and quality of life for those living with neuromuscular conditions.

1.5 Factors Influencing the Prevalence of Neuromuscular Disorders

Neuromuscular disorders can affect people of all ages, backgrounds, and genders, leading to a variety of symptoms such as muscle weakness and loss of coordination. While these conditions can arise in anyone, certain disorders may be more prevalent in specific demographic groups due to genetic factors or age. For example, spinal muscular atrophy (SMA) and muscular dystrophy often occur more frequently in populations with particular genetic predispositions, while age-related conditions like amyotrophic lateral sclerosis (ALS) are typically seen in middle-aged adults. Recognizing these patterns is crucial for early detection and tailored care.

- **Genetic Factors:** Numerous neuromuscular disorders have a hereditary component, indicating that they can be transmitted through familial lines. Specific genetic mutations are responsible for a variety of conditions, including the well-known muscular dystrophies. For instance, Duchenne muscular dystrophy predominantly impacts males and is marked by progressive muscle degeneration, while myotonic dystrophy presents a range of symptoms, including muscle weakness and stiffness. Families with a history of these disorders often experience a ripple effect, witnessing the impact of these conditions across several generations, as the genetic predisposition increases the likelihood of multiple family members being affected.

- **Age :** Neuromuscular disorders can affect individuals across the entire lifespan, demonstrating a remarkable variability in age of onset. For instance, Duchenne muscular dystrophy, a genetic condition characterized by rapid muscle degeneration, typically manifests during early childhood, often before the age of five. In contrast, amyotrophic lateral sclerosis (ALS), a progressive neurodegenerative disease that impacts motor neurons, generally presents itself in middle adulthood, usually between

the ages of 40 and 70. Furthermore, conditions like myasthenia gravis, which leads to fluctuating muscle weakness due to an autoimmune response, are noteworthy for their ability to affect individuals of all ages, although they are most commonly diagnosed in adults. This spectrum of age-related onset highlights the complex nature of neuromuscular disorders and the diverse challenges they pose to patients and their families.

- **Gender :** plays a significant role in the prevalence of various neuromuscular disorders, with notable differences observed between males and females. For example, Duchenne muscular dystrophy predominantly affects boys, a phenomenon attributed to its X-linked inheritance pattern, which means the defective gene responsible for the condition is located on the X chromosome. As a result, males, who have only one X chromosome, are more severely impacted. In contrast, disorders such as myasthenia gravis are more frequently diagnosed in females, highlighting the gender disparities in the presentation and diagnosis of these conditions. This variation underscores the importance of considering gender as a critical factor in understanding and addressing neuromuscular disorders.

- **Environmental and Lifestyle Factors:** While genetics undeniably contribute significantly to the development of neuromuscular disorders, various environmental elements can also play a crucial role in their onset and progression. For instance, prolonged exposure to environmental toxins, such as heavy metals or industrial chemicals, can result in toxic myopathies, where muscle fibers are damaged and lead to muscle weakness. Additionally, viral infections may trigger autoimmune responses that can culminate in conditions like polymyositis, where the body's immune system mistakenly attacks its own muscle tissue. Certain medications, particularly those that are cytotoxic or have neurotoxic properties, can also contribute to acquired neuromuscular disorders, thereby underscoring the importance of understanding both environmental and lifestyle influences in the context of neuromuscular health.

- **Autoimmune Conditions:** A number of neuromuscular diseases arise from autoimmune reactions, where the body's immune system erroneously targets and attacks its own tissues. Notable conditions within this spectrum include myasthenia gravis, characterized by weakness and rapid fatigue of voluntary muscles, and Guillain-

Barré syndrome, which often begins with weakness and tingling in the legs and can rapidly progress to more severe motor impairment. These disorders can manifest in individuals of any age, but certain autoimmune neuromuscular conditions are observed to occur more frequently among women or older adults. The unpredictable nature of these diseases often poses unique challenges, impacting patients' daily lives and requiring comprehensive medical management.

1.6 Impact of Neuromuscular Disorders on Individuals

Neuromuscular disorders significantly affect individuals by disrupting the intricate interplay between muscles and the nervous system, leading to difficulties in muscle contraction and movement. These conditions can arise from a variety of factors, including dysfunctions within the muscle fibers themselves, problems with the peripheral nerves that send signals to these muscles, or issues at the neuromuscular junctions—the critical sites where nerve impulses are transmitted to initiate muscle action. The consequences of these disorders can vary widely, impacting mobility, strength, and overall quality of life in profound ways. Individuals may experience muscle weakness, cramps, spasms, and fatigue, making simple tasks challenging and often leading to a decline in independence. Moreover, the emotional and psychological toll of coping with such conditions can further complicate the experiences of those affected.

- **Muscle Weakness and Atrophy:** Numerous neuromuscular disorders result in progressive muscle weakness, leading to a gradual decline in strength and function. This weakening can significantly impact daily activities such as walking, lifting, or even breathing. For example, in conditions like muscular dystrophy, muscle fibers are gradually damaged over time, resulting in both weakness and atrophy (muscle wasting).

- **Movement Difficulties:** Neuromuscular disorders often impair the coordination and control of voluntary movements. For example, individuals with amyotrophic lateral sclerosis (ALS) may find it challenging to perform basic motor functions like speaking, swallowing, and walking due to the progressive degeneration of motor neurons that transmit signals from the brain to the muscles.

- **Fatigue:** People with neuromuscular diseases often experience significant fatigue, as their muscles struggle to sustain the energy needed for normal movement and function.

This can make everyday tasks feel exhausting, even if they are not physically demanding.

- **Respiratory Problems:** In severe cases of neuromuscular disorders, such as ALS or muscular dystrophies, individuals may face respiratory complications. Weakness in the muscles responsible for breathing can lead to difficulties in maintaining adequate lung function, which may necessitate mechanical support such as ventilators to assist with breathing.

- **Pain and Discomfort:** Certain neuromuscular disorders, like myopathies or neuropathies, can cause muscle pain, cramping, or soreness. This pain can be constant or may flare up during specific activities, contributing to overall discomfort.

- **Sensory Impairments:** Conditions such as peripheral neuropathy can damage the nerves that transmit sensory information from the skin, muscles, and joints to the brain, leading to sensations of tingling, numbness, or loss of feeling in certain body parts.

- **Gait and Posture Problems:** Weakness in leg muscles or loss of coordination can impact walking, balance, and posture. Individuals with neuromuscular disorders may experience difficulty walking, an altered gait, or may need mobility aids such as canes or wheelchairs for assistance.

- **Cognitive and Speech Impairments:** Some neuromuscular disorders, including ALS, can also affect cognitive and speech functions. Individuals may struggle with forming words or controlling the muscles used in speech, making communication challenging. Cognitive changes can occur in some cases of ALS, although they are less common.

- **Loss of Independence:** Over time, as the disorder progresses and symptoms worsen, individuals may experience a significant loss of independence. They may require assistance with daily activities such as dressing, bathing, and eating.

1.7 Types of Neuromuscular Disorders

Neuromuscular disorders (NMDs) encompass a wide range of conditions that affect the functioning of muscles and the nerves controlling them. These disorders can be classified based

on their origin—either genetic or acquired—and their impact on muscle function. They vary in severity, from mild to severe, and may result in progressive disability. Below is a comprehensive overview of different types of neuromuscular disorders, including their curability, potential for early-stage treatment, specific treatment approaches, and methods for identification based on symptoms.

Muscular dystrophies are a group of genetic disorders characterized by progressive weakness and degeneration of skeletal muscles. These disorders predominantly affect males, as most forms are X-linked. Examples include Duchenne Muscular Dystrophy (DMD), Becker Muscular Dystrophy (BMD), Myotonic Dystrophy, and Facioscapulohumeral Muscular Dystrophy (FSHD). Although these conditions are incurable, management options are available, such as physical therapy, corticosteroids, and experimental gene therapies. Common symptoms include muscle weakness, difficulty walking, delays in motor skill development, and respiratory complications in later stages.

- **Amyotrophic Lateral Sclerosis (ALS)** is a progressive neurodegenerative disorder that affects motor neurons, resulting in a loss of voluntary muscle control. It primarily affects adults, with a higher incidence in males over the age of 40. Although there is currently no known cure, medications such as Riluzole and Edaravone may help slow the progression of the disease. Supportive therapies, including breathing assistance and speech therapy, can aid in managing symptoms, which may include muscle stiffness, weakness, difficulty speaking and swallowing, and ultimately, respiratory failure.

- **Spinal Muscular Atrophy (SMA)** is a genetic disorder that leads to the loss of motor neurons, resulting in progressive muscle weakness and wasting. While it predominantly affects infants and children, it can also occur in adults. SMA is classified into different types based on severity, with Type 1 being the most severe and Type 4 typically presenting in adulthood. Although there is no cure for SMA, early treatment with gene therapies such as Spinraza and Zolgensma can significantly improve outcomes. Common symptoms of SMA include muscle weakness, difficulty swallowing, respiratory infections, and delays in motor development.

- **Myasthenia Gravis (MG)** is an autoimmune disorder in which the immune system mistakenly attacks the neuromuscular junction, resulting in muscle weakness. This condition can affect individuals of all ages, but it is more commonly seen in young

women and older men. While there is currently no cure for MG, it can be effectively managed with medications such as acetylcholinesterase inhibitors and immunosuppressants. Common symptoms include muscle fatigue, drooping eyelids, and difficulty breathing and swallowing, which may worsen with physical exertion**.**

- **Charcot-Marie-Tooth Disease (CMT)** is a hereditary disorder that affects the peripheral nerves, leading to muscle weakness and sensory loss. It can manifest in either childhood or adulthood and affects both genders equally. Although there is currently no cure for CMT, symptoms can be managed through physical therapy, the use of orthotic devices, and pain management strategies. Common symptoms include foot deformities, muscle atrophy, difficulty walking, and loss of sensation in the extremities.

- **Guillain-Barré Syndrome (GBS)** is an autoimmune disorder in which the immune system attacks the peripheral nerves, resulting in paralysis. It can affect individuals of all ages but is more commonly seen in adults, particularly males. Unlike many other neuromuscular disorders, GBS often allows for complete recovery, especially when treated early with options such as intravenous immunoglobulin (IVIG) and plasma exchange. Symptoms typically include sudden weakness, tingling sensations, paralysis, and difficulty breathing, which can develop rapidly over the course of days or weeks.

- **Inclusion Body Myositis (IBM**) is an inflammatory muscle disease characterized by progressive muscle weakness. It mainly affects older adults, typically those over 50 years of age, and is more prevalent in men. While there is no cure for this condition, physical therapy may help slow its progression. Patients often experience muscle atrophy, difficulty gripping objects, and issues with swallowing, which can make daily activities increasingly challenging.

- **Lambert-Eaton Myasthenic Syndrome (LEMS)** is an autoimmune disorder that disrupts communication between nerves and muscles. It is often associated with small-cell lung cancer and primarily affects middle-aged and older adults. Although there is no cure for LEMS, medications such as 3,4-Diaminopyridine (3,4-DAP) can help improve symptoms. Additionally, immunosuppressants and physiotherapy can assist in

managing the condition. The symptoms include muscle weakness, difficulty standing, dry mouth, and excessive fatigue.

- **Friedreich's Ataxia (FA)** is a genetic disorder that impacts the nervous system and leads to difficulties with muscle coordination. It typically manifests in childhood or adolescence and affects both genders equally. While there is no cure for FA, physical therapy and cardiac medications can help manage its symptoms. Patients often face challenges such as lack of coordination, speech difficulties, scoliosis, and heart disease.

- **Chronic Inflammatory Demyelinating Polyneuropathy (CIDP)** is an autoimmune disorder that causes damage to the nerves and leads to muscle weakness. Although it can affect individuals of any age, it is more common in adults, particularly males. While there is no cure for CIDP, treatments such as intravenous immunoglobulin (IVIG), corticosteroids, and plasma exchange can lead to remission in some patients. Common symptoms include numbness, tingling, progressive weakness, and a loss of reflexes, which can significantly impair mobility over time.

- **Duchenne Muscular Dystrophy (DMD)** is a severe genetic disorder that primarily affects young boys, leading to progressive muscle degeneration due to a deficiency of dystrophin, an essential muscle protein. The disease results in a gradual loss of mobility, respiratory failure, and cardiac complications, often leading to early mortality. Although there is currently no cure, corticosteroids and emerging gene therapies may help slow the progression of the disease. Common symptoms of DMD include frequent falls, difficulty walking, muscle weakness, and breathing difficulties.

- **Becker Muscular Dystrophy (BMD)** is a milder form of Duchenne Muscular Dystrophy (DMD), with symptoms typically appearing later in adolescence or adulthood. It leads to progressive muscle weakness, primarily affecting the legs and pelvis, but this occurs at a slower pace compared to DMD. Although there is no cure for BMD, treatment options such as physical therapy, steroids, and mobility aids can help manage symptoms. Patients may face difficulties in walking, experience muscle cramps, and encounter cardiac complications in the later stages of the condition.

- **Facioscapulohumeral Muscular Dystrophy (FSHD)** is a genetic disorder that leads to weakness in the muscles of the face, shoulders, and upper arms. Symptoms usually begin in early adulthood and progress slowly over time. Although there is no cure for FSHD, physical therapy, exercise, and the use of assistive devices can help enhance mobility. Common symptoms include difficulty lifting the arms, weakness in the facial muscles, and scapular winging.

- **Congenital myopathies** are a group of inherited muscle disorders that affect individuals from birth. These disorders typically lead to low muscle tone (hypotonia), weakness, and delayed motor development. The severity of symptoms varies depending on the specific subtype of the condition. Although there is no cure for congenital myopathies, treatment focuses on managing symptoms. This may include physical therapy, respiratory support, and nutritional care.

- **Dermatomyositis** is an inflammatory muscle disease marked by muscle weakness and a distinctive skin rash. It can affect both children and adults, but it is more common in women. This condition is often linked to autoimmune responses and, in some instances, cancer. Treatment typically involves corticosteroids, immunosuppressants, and physical therapy. Common symptoms include muscle weakness, difficulty swallowing, fatigue, and changes in skin color.

- **Polymyositis** is a chronic inflammatory disorder that leads to progressive muscle weakness, especially in the shoulders, hips, and neck. This condition primarily affects adults and is believed to be autoimmune in nature. Although there is no cure for polymyositis, treatments such as corticosteroids, immunosuppressants, and physical therapy can help manage symptoms. Patients often face challenges with activities like climbing stairs, lifting objects, and swallowing.

- **Metabolic myopathies** refer to a group of disorders resulting from enzyme deficiencies that impair muscle metabolism. For example, McArdle’s disease causes symptoms like exercise intolerance, muscle cramps, and weakness due to the body's inability to effectively utilize energy sources. Treatment primarily involves dietary adjustments, exercise management, and symptom control. Toxic myopathies occur when drug use, alcohol, or toxins damage muscle fibers. Common causes include statins,

corticosteroids, and excessive alcohol consumption. Symptoms may include muscle pain, weakness, and, in severe cases, muscle breakdown known as rhabdomyolysis. Treatment typically involves stopping the use of the harmful substance and providing supportive care. Periodic Paralysis refers to a group of rare genetic disorders that result in temporary episodes of muscle weakness or paralysis caused by fluctuations in potassium levels. The two main types are hypokalemic periodic paralysis and hyperkalemic periodic paralysis. Treatment typically involves dietary adjustments, potassium supplements, and medications to help stabilize ion channel function.

- **Stiff-Person Syndrome (SPS)** is a rare neurological disorder characterized by severe muscle stiffness and painful spasms, which are often triggered by stress or sudden movements. This condition is associated with autoimmune dysfunction and can significantly affect mobility. Treatment options include muscle relaxants, immunotherapy, and physical therapy. Common symptoms of SPS include rigid muscles, difficulty walking, and increased sensitivity to stimuli.

- **Mitochondrial myopathies** are caused by defects in mitochondria, which are the energy-producing structures within cells. These defects can lead to muscle weakness, difficulty with exercise, and complications affecting multiple organs. Although there is no cure for mitochondrial myopathies, supportive treatments such as physical therapy, nutritional supplements, and adjustments to exercise can help improve quality of life.

- **Kennedy's Disease,** also known as **Spinal and Bulbar Muscular Atrophy (SBMA)**, is a genetic disorder that mainly affects males. It leads to progressive muscle weakness, tremors, and difficulties with swallowing and speaking. The condition is associated with mutations in the androgen receptor gene. Although there is currently no cure, physical therapy and other supportive treatments can help manage the symptoms. Common complications of the disease include muscle atrophy and hormonal imbalances.

- **Tay-Sachs Disease (Late-Onset Form)** is a rare inherited neurological disorder that causes progressive muscle weakness, motor decline, and cognitive impairments. It occurs due to enzyme deficiencies that impact lipid metabolism in nerve cells. While

there is no cure for this condition, supportive therapies such as physical therapy and respiratory support can help manage symptoms.

1.8 Organizations and Institutions Supporting People with Neuromuscular Disorders

Around the world, numerous organizations are dedicated to improving the quality of life for individuals affected by neuromuscular disorders. These organizations take a holistic approach, utilizing various strategies, including providing essential medical assistance, offering financial support to families, launching public awareness campaigns to educate communities, and funding innovative research aimed at developing new treatments and potential cures.

Among the key contributors in this field, several organizations excel in different areas, each providing unique resources and expertise. Some focus on directly assisting patients by offering vital medical resources, such as adaptive technologies and rehabilitation services, while others prioritize advocacy and legislative initiatives to push for improved healthcare policies.

Additionally, many nonprofit organizations work actively to raise awareness about neuromuscular disorders, striving to inform both the public and healthcare professionals about the challenges faced by those living with these conditions. By fostering collaboration among researchers, healthcare providers, and patients, these organizations are making significant progress toward improving outcomes and enhancing the overall well-being of affected individuals and their families.

- **Muscular Dystrophy Association (MDA)** funds innovative research, provides healthcare resources, and supports families through specialized clinics and educational programs. The organization also emphasizes public awareness and advocacy to promote policy changes that enhance healthcare accessibility for individuals with neuromuscular conditions.

- **National Institute of Neurological Disorders and Stroke (NINDS)** is a division of the National Institutes of Health (NIH) that conducts and supports fundamental biomedical research focused on neurological disorders, including neuromuscular diseases. The institute provides research grants, collaborates with scientists globally, and helps develop innovative diagnostic tools and treatment strategies.

- **World Federation of Neurology (WFN)** is committed to advancing global education and research in neurology. It encourages collaboration among specialists worldwide

and plays a vital role in sharing knowledge, supporting scientific progress, and organizing international conferences. These initiatives aim to enhance the management of neuromuscular disorders.

- **European Neuromuscular Centre (ENMC)** is a collaborative research network focused on enhancing knowledge and treatment for neuromuscular disorders. ENMC organizes workshops, promotes knowledge sharing, and funds research projects within this field. By uniting scientists, clinicians, and patient organizations, ENMC helps connect research with practical clinical applications.
- **Patient Advocacy Groups**: Many non-profit organizations, including Cure SMA, the ALS Association, and the Charcot-Marie-Tooth Association, focus on specific neuromuscular diseases. These organizations fund scientific research, provide resources for patients, organize support networks, and advocate for improved medical care policies on a global scale.

1.9 Providing Data for Global Research and Treatment Advancements

Data collection and analysis play a crucial role in refining the diagnosis, treatment, and long-term management of neuromuscular disorders. These complex conditions, which affect the nerves that control voluntary muscles, require precise and comprehensive data to improve patient outcomes. Numerous global initiatives have been established to increase the availability of high-quality datasets. By pooling information from diverse sources, these initiatives promote significant advancements in both medical practice and research. As a result, healthcare professionals can develop more effective treatment plans, while researchers gain valuable insights into the underlying mechanisms of these disorders, ultimately leading to improved therapies and support for affected individuals.

1.9.1. Clinical Registries and Biobanks

Neuromuscular disease registries are vital resources that meticulously collect and organize extensive patient information. This includes genetic data, which reveals hereditary factors, along with a wide array of clinical characteristics that detail the symptoms and progression of the diseases. Additionally, these registries track treatment outcomes, providing insights into the effectiveness of various therapies and interventions. By consolidating this wealth of information, researchers can identify significant patterns, explore correlations between genetic

and clinical factors, and unravel the complex mechanisms underlying disease progression. Among the prominent registries that serve this crucial role are:

- **TREAT-NMD Global Registry** is a global network of patient registries that gathers clinical and genetic data on neuromuscular diseases. Its purpose is to facilitate research studies and clinical trials. This registry supports pharmaceutical companies and research institutions in developing targeted therapies and assessing the effectiveness of treatments.
- **The UK Biobank** is a large repository containing genetic and health data from thousands of individuals with neuromuscular disorders, supporting advanced research in disease mechanisms, biomarker discovery, and drug development.
- **Rare Disease Clinical Research Network (RDCRN)** is supported by the NIH. This initiative maintains extensive patient registries and promotes collaboration among researchers, clinicians, and pharmaceutical companies to accelerate the development of new therapies for rare neuromuscular disorders.

1.9.2. Contributions to Medical Advancements

Data gathered from comprehensive clinical studies, detailed patient registries, and in-depth genetic research has played a pivotal role in advancing the creation of innovative treatments and therapeutic interventions. These efforts have led to the development of targeted medications and personalized therapies that address specific conditions more effectively, improving patient outcomes and enhancing the overall quality of care.

- **Gene Therapy**: Advancements in gene-editing technologies, such as CRISPR, have led to revolutionary treatments for conditions like Duchenne Muscular Dystrophy (DMD) and Spinal Muscular Atrophy (SMA). Gene therapy focuses on correcting genetic mutations at their source, providing long-term solutions for conditions that were previously untreatable.
- **Stem Cell Research**: Clinical trials are currently exploring the use of stem cells for muscle regeneration in patients with neuromuscular diseases. These stem cell-based therapies have the potential to restore muscle function and improve the quality of life for those affected.

- **Precision medicine** involves personalized treatment strategies informed by genetic profiling and biomarker analysis. These tailored therapies target the specific molecular mechanisms behind each patient's condition, enhancing treatment effectiveness while minimizing adverse effects.

1.9.3. Enhancing Medical and Treatment Applications

The convergence of artificial intelligence (AI), machine learning, and big data analytics in the realm of neuromuscular disorder research has significantly transformed diagnostic precision and treatment outcomes. This innovative integration allows for the meticulous analysis of complex datasets, unveiling patterns and insights that were previously obscured. By leveraging advanced algorithms, researchers can enhance the accuracy of diagnoses, leading to more targeted and effective treatment strategies tailored to individual patient needs. Key contributions of this technological synergy include improved predictive models, personalized therapies, and a deeper understanding of the pathophysiology of various neuromuscular conditions, ultimately paving the way for groundbreaking advancements in patient care.

- **Early Diagnosis and Prediction Models**: AI-driven tools analyze patient data, medical histories, and genetic profiles to detect early signs of neuromuscular disorders. Machine learning models help predict disease progression and optimize treatment planning.
- **Wearable Health Technologies**: Advanced wearable devices continuously monitor muscle function, movement patterns, and disease progression, providing real-time data to physicians and researchers. These innovations allow for more precise tracking of treatment outcomes and disease management.
- **Telemedicine and Remote Monitoring**: The emergence of digital health platforms has enhanced access to specialized care for neuromuscular patients. Telemedicine services enable patients to receive expert consultations and participate in remote research studies, effectively reducing geographical barriers to healthcare and clinical trial participation.

The collaborative efforts of medical researchers, healthcare providers, patient advocacy groups, and technological innovators are driving progress in neuromuscular disorder research. These organizations contribute to better patient outcomes, enhanced treatment options, and a deeper understanding of these complex diseases by providing valuable data and supporting medical advancements. The integration of cutting-edge research methodologies with real-world data

applications ensures that individuals affected by neuromuscular disorders receive the best possible care, paving the way for future breakthroughs in neurology and personalized medicine.

The ongoing expansion of global research networks, along with advancements in genetic engineering, AI-based diagnostics, and novel therapeutic approaches, promises a future where neuromuscular disorders can be more effectively managed and, in some cases, potentially cured. The commitment to data-driven research and collaborative innovation will continue to transform the landscape of neuromuscular healthcare, offering hope and improved quality of life for patients worldwide.

1.10 Treatment Provided by Institutions for Neuromuscular Disorders

Neuromuscular disorders represent a diverse group of medical conditions that primarily impact the nerves and muscles, resulting in a range of symptoms including muscle weakness, diminished strength, and significant physical limitations. These disorders can stem from various underlying causes, including genetic mutations, autoimmune responses, or complications from other diseases, each contributing to the complexity of treatment. Medical institutions have developed a variety of treatment modalities aimed at managing symptoms, preserving function, and enhancing the quality of life for affected individuals. Among the primary treatment options are:

1.10.1.Physical Therapy

Physical therapy is a fundamental treatment for many neuromuscular disorders. It encompasses a variety of exercises designed to strengthen muscles, improve coordination, and maintain mobility. Specialized programs are tailored to each patient, focusing on slowing the progression of the disease and enhancing functional independence.

- **Best for:**Conditions such as muscular dystrophy, amyotrophic lateral sclerosis (ALS), and multiple sclerosis.
- **Benefits:** Regular physical therapy helps prevent muscle atrophy, improve balance, and support daily activities. By maintaining functional independence and reducing disability, patients can perform daily tasks more effectively, thereby enhancing their overall quality of life**.**

1.10.2.Medications (Tablets & Injections)

Pharmacological treatments are essential in managing symptoms, slowing disease progression, and alleviating pain in neuromuscular disorders. Commonly prescribed medications include corticosteroids, immunosuppressants, and neuroprotective agents, tailored to the specific disorder and its severity.

- **Best for:** Myasthenia gravis (treated with acetylcholinesterase inhibitors), Duchenne muscular dystrophy (treated with corticosteroids), and peripheral neuropathies.
- **Improvement:** These medications help reduce inflammation, enhance muscle strength, and slow disease progression, which can lead to prolonged mobility and improved overall well-being. In many cases, they enable patients to engage in daily activities more effectively, while also reducing pain and weakness.

1.10.3. Surgical Interventions

Surgical interventions are often considered when neuromuscular disorders lead to severe structural complications or do not respond to other treatments. These surgical procedures may involve correcting deformities, relieving nerve compression, or implanting assistive devices, such as spinal stimulators.

- **Best suited for:** Conditions like spinal muscular atrophy (SMA), carpal tunnel syndrome, myasthenia gravis (in severe cases), and nerve entrapment syndromes.
- **Potential improvements:** Surgery can provide long-term relief from chronic pain, restore partial mobility, and enhance nerve function. This can enable patients to regain some independence in their daily tasks and significantly improve their overall quality of life.

1.10.4. Gene Therapy

Gene therapy is an innovative and evolving treatment option for neuromuscular disorders caused by genetic mutations. By directly targeting the genetic root cause, gene therapy aims to correct or replace the defective genes responsible for muscle degeneration, offering new hope to patients with certain conditions.

- **Best for:** Duchenne muscular dystrophy and spinal muscular atrophy.

- **Improvement:** Gene therapy has the potential to stop or slow the progression of these disorders by correcting the underlying genetic defects. This treatment could lead to improved muscle function, delayed onset of the disease, and a better prognosis for affected individuals.

1.10.5. Stem Cell Therapy

Stem cell therapy is an innovative approach to treating neuromuscular disorders. It involves the transplantation of stem cells that can develop into muscle cells, which may replace damaged muscle tissue. This therapy shows promise for regenerating injured tissues and restoring muscle function.

- **Best for:** Duchenne muscular dystrophy, ALS (amyotrophic lateral sclerosis), and spinal cord injuries.
- **Benefits:** Stem cell therapy has the potential to regenerate muscle tissue, enhance strength, and slow the progression of these diseases. Although it is still in the early stages of research, stem cell therapy offers the hope of long-term recovery and improvement in muscle function.

1.10.6.Respiratory Support

For patients with neuromuscular disorders that affect the muscles responsible for breathing, respiratory support is essential. Non-invasive ventilation (NIV) or mechanical ventilators can assist with breathing and help prevent respiratory failure, which is common in the advanced stages of certain neuromuscular conditions.

- **Best for:** ALS, Duchenne muscular dystrophy, myasthenia gravis.
- **Benefits:** Respiratory support helps prevent complications such as respiratory failure, improves oxygenation, and can significantly prolong life expectancy. It also enhances comfort and increases the patient's ability to perform daily tasks.

1.10.7.Physiotherapy & Occupational Therapy

In addition to physical therapy, both physiotherapy and occupational therapy play a crucial role in improving motor functions, preventing muscle stiffness, and enhancing a patient’s

ability to perform daily activities. These therapies may involve the use of assistive devices such as braces, wheelchairs, and orthotic supports to aid mobility.

- **Best for:**Parkinson's disease, ALS, multiple sclerosis, and Guillain-Barré syndrome.
- **Improvements:** These therapies help improve muscle control, reduce stiffness, and support good posture and functional independence. By focusing on daily activities and enhancing mobility, patients can enjoy a greater ability to live independently and comfortably.

The innovative combination of these seven treatments—**physical therapy, medications, surgical interventions, gene therapy, stem cell therapy, respiratory support, and physiotherapy/occupational therapy**—offers a comprehensive and holistic strategy for managing neuromuscular disorders. This multi-faceted approach not only targets the physical manifestations of these conditions but also emphasizes the functional aspects of daily living. By integrating these therapies, individuals can experience a significant reduction in symptoms, alongside marked improvements in their mobility, independence, and overall quality of life.

The efficacy of these diverse treatments is deeply rooted in the principles of early diagnosis, tailored care plans, and ongoing management. A swift and accurate diagnosis allows for the initiation of appropriate interventions, while personalized care plans cater to the unique needs and circumstances of each individual. Consistent monitoring and adjustments to the treatment regimen ensure that those living with neuromuscular disorders can lead enriching and fulfilling lives, despite the inherent challenges that their conditions present. By addressing both physical limitations and emotional well-being, this comprehensive framework supports patients in achieving their personal goals and enhancing their everyday experiences.

1.11 Diagnostic Tests for Neuromuscular Disorders: Understanding the Purpose and Process

Neuromuscular disorders comprise a diverse range of conditions that impact the functioning of muscles, nerves, and the intricate communication pathways between them. These disorders can manifest through a variety of symptoms, including muscle weakness, tingling or numbness, coordination challenges, and in more severe cases, significant physical impairments that can affect daily activities and overall quality of life.

The diagnosis of neuromuscular disorders necessitates a thorough and multidimensional approach. Healthcare professionals typically begin with a detailed medical history and a physical examination to assess symptoms and their impact on mobility and function. This initial evaluation is often supplemented by specialized diagnostic tests, such as electromyography (EMG), which measures the electrical activity of muscles, and nerve conduction studies, which assess how well electrical impulses travel through nerves.

These diagnostic tools are crucial for identifying the specific type of neuromuscular disorder present, as they help clinicians determine not only the severity of the condition but also its potential underlying causes. Disorders could be due to genetic factors, autoimmune responses, infections, metabolic issues, or other pathologies. Understanding the precise nature of the disorder allows for the formulation of tailored treatment plans that may include physical therapy, medications, lifestyle adjustments, or, in some cases, surgical interventions.

By combining clinical assessments with advanced diagnostic techniques, healthcare providers can deliver a comprehensive overview of a patient's condition, thereby facilitating effective management strategies that aim to improve function, enhance quality of life, and mitigate the progression of the disorder.

1.11.1. Electromyography (EMG)

Electromyography (EMG) is a specialized diagnostic test designed to assess the intricate electrical activity of muscles and evaluate the health of the motor neurons that control them. The procedure typically involves the careful insertion of a fine, needle-like electrode into the muscle tissue, allowing for the precise recording of electrical signals generated during muscle contractions. As the muscle activates, these electrodes capture and transmit the electrical impulses, which are subsequently analyzed to gain insights into the muscle's functional status.

EMG proves to be particularly invaluable in diagnosing a variety of neuromuscular disorders, including **muscular dystrophies**, **amyotrophic lateral sclerosis (ALS)**, **myasthenia gravis**, and **peripheral neuropathy**. By examining the recorded electrical signals, healthcare professionals can identify any abnormalities in muscle function and nerve communication. This detailed assessment is crucial in uncovering the underlying causes of muscle weakness, abnormal movements, and other neuromuscular symptoms. Ultimately, EMG serves as a powerful tool that aids physicians in developing effective treatment plans tailored to each patient's condition.

1.11.2. Nerve Conduction Studies (NCS)

Nerve conduction studies (NCS) are diagnostic tests designed to evaluate the speed and strength of electrical signals as they traverse through peripheral nerves. This assessment is essential for understanding nerve function and is frequently conducted alongside electromyography (EMG) for a comprehensive evaluation.

During an NCS, small, adhesive electrodes are strategically placed on the skin over the specific nerve being examined. A gentle electrical current is then introduced through these electrodes, which stimulates the nerve. The resulting electrical responses are carefully recorded, allowing for the measurement of conduction velocity and response times.

This test is instrumental in pinpointing instances of nerve damage or impairment in nerve conduction. Such insights are crucial for diagnosing a variety of neurological disorders, including **peripheral neuropathy**, **Guillain-Barré syndrome**, **carpal tunnel syndrome**, and **diabetic neuropathy**. By accurately determining whether nerve damage or dysfunction is behind certain neuromuscular symptoms — such as weakness, numbness, or pain — NCS plays a pivotal role in guiding treatment options and improving patient outcomes.

1.11.3. Muscle Biopsy

A muscle biopsy is a medical procedure that entails the extraction of a small specimen of muscle tissue for detailed laboratory examination. This diagnostic test offers a comprehensive view of the muscle's architecture, allowing healthcare professionals to identify any signs of inflammation, degeneration, or abnormalities within the muscle fibers. It is particularly useful when other diagnostic tests yield inconclusive results or when a specific muscle disorder is suspected.

Muscle biopsies play a crucial role in diagnosing various conditions, including **muscular dystrophies**, **inflammatory myopathies**, and **mitochondrial myopathies**. In the laboratory, the tissue sample is meticulously examined under a microscope, where pathologists look for distinctive changes that might elucidate the causes of muscle weakness or atrophy.

Though it is considered an invasive procedure, a muscle biopsy is generally performed under local anesthesia to minimize discomfort for the patient. The information gleaned from this test can significantly contribute to understanding the underlying mechanisms of neuromuscular disorders and guide appropriate treatment strategies. By evaluating the characteristics of the

muscle tissue, doctors can develop a more targeted approach to managing these complex conditions.

1.11.4. Genetic Testing

Genetic testing is a sophisticated process that involves the thorough analysis of a patient's DNA to uncover mutations or genetic abnormalities linked to various neuromuscular disorders. This vital testing method proves particularly beneficial for diagnosing inherited conditions such as **Duchenne muscular dystrophy**, **spinal muscular atrophy (SMA)**, **Charcot-Marie-Tooth disease**, and **myotonic dystrophy**.

The procedure typically begins with the collection of a blood or saliva sample, which serves as the source of DNA. Once extracted, the DNA is meticulously examined for specific mutations that might elucidate the symptoms a patient is experiencing.

By confirming a diagnosis through genetic testing, healthcare professionals can not only identify the precise neuromuscular disorder impacting the individual but also gain insight into the inheritance patterns of the condition. This information is essential for guiding families in understanding their risks and implications. Furthermore, it allows for the development of tailored treatment plans that are best suited to the patient's unique genetic profile.

Overall, genetic testing stands as a crucial diagnostic tool, particularly for patients with a family history of neuromuscular disorders or those who display symptoms indicative of a genetic condition. Its relevance extends beyond mere diagnosis, offering a path toward informed decision-making and enhanced treatment strategies.

1.11.5. Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique used to create detailed images of the body's internal structures, including muscles and nerves. In neuromuscular disorders, MRI is used to detect abnormalities in the muscles, nerves, and spinal cord. The test works by using a magnetic field and radio waves to produce high-resolution images of the soft tissues. MRI is particularly helpful for diagnosing **muscle atrophy**, **inflammatory myopathies**, and **spinal cord disorders** that affect nerve function. It can also detect changes in muscle tissue that may be indicative of diseases like **muscular dystrophy** or **polymyositis**. MRI helps doctors visualize structural changes and make more accurate diagnoses.

1.11.6. Spinal Tap (Lumbar Puncture)

A spinal tap, also known as a lumbar puncture, is a vital medical procedure designed to collect cerebrospinal fluid (CSF) from the spinal canal, which lies within the lower back. This clear fluid acts as a protective cushion for both the brain and spinal cord, playing a crucial role in maintaining central nervous system health. The primary purpose of extracting CSF is to analyze it for any signs of infection, inflammation, or other abnormalities that may be linked to various neuromuscular disorders.

During the procedure, a skilled healthcare professional carefully inserts a thin, sterile needle between the vertebrae in the lower back. This delicate maneuver is performed while the patient is typically in a curled position to widen the spaces between the vertebrae. Once the needle reaches the spinal canal, a sample of the cerebrospinal fluid is drawn out and collected for further examination.

In the laboratory, the fluid is meticulously analyzed for elevated levels of proteins, an increased white blood cell count, or other specific markers that could suggest the presence of conditions such as **Guillain-Barré syndrome**, **multiple sclerosis**, or **myasthenia gravis**. This analysis is particularly crucial in diagnosing disorders affecting both the central nervous system and peripheral nerves, as it provides invaluable insights into the underlying causes of various neuromuscular symptoms, enabling healthcare providers to formulate effective treatment plans tailored to the individual's needs.

1.11.7. Blood Tests

Blood tests play a crucial role in the evaluation of markers associated with neuromuscular disorders, providing valuable insights into underlying neuromuscular health. To conduct these tests, blood samples are collected to measure a range of biochemical substances, one of the most significant being **creatine kinase (CK)**. This enzyme is released into the bloodstream following muscle damage; therefore, elevated levels of CK can serve as a vital indicator of muscle injury, making it particularly relevant for conditions such as muscular dystrophy and myositis.

In addition to CK, blood tests can assess the presence of specific antibodies that may indicate autoimmune disorders, revealing a further layer of complexity in diagnosing these conditions. For example, in myasthenia gravis, particular antibodies that target the neuromuscular junction can be detected, providing critical information that aids in diagnosis.

Furthermore, blood tests are instrumental in uncovering metabolic disturbances and electrolyte imbalances, such as imbalances in potassium or calcium, which can significantly contribute to neuromuscular symptoms. These tests not only help to identify the presence of specific disorders but also assist physicians in understanding the overall biochemical environment of the patient, enabling a more tailored and effective treatment approach. Overall, blood tests are an essential component of the diagnostic process, guiding healthcare professionals in determining the underlying causes of neuromuscular symptoms and developing appropriate management strategies.

Diagnostic tests are crucial in the nuanced process of diagnosing neuromuscular disorders and shaping effective treatment strategies. Each test serves as a window into the intricate health of the muscles, nerves, and the vital communication pathways that connect them. Tests such as electromyography (EMG) enable doctors to assess the electrical activity of muscles, revealing abnormalities that may indicate neuromuscular disease. Nerve conduction studies (NCS) measure the speed and strength of signals traveling through the nerves, providing insights into potential nerve damage or dysfunction.

Muscle biopsies involve the careful extraction of muscle tissue for microscopic examination, allowing for the identification of specific diseases at a cellular level. Genetic testing plays a transformative role by uncovering inherited conditions that might underlie the symptoms. Advanced imaging techniques like magnetic resonance imaging (MRI) offer detailed views of both muscles and nerves, aiding in the identification of structural abnormalities.

Furthermore, spinal taps allow for the evaluation of cerebrospinal fluid, which can provide critical information about neurological conditions. Blood tests are also instrumental, revealing markers that can indicate inflammatory or metabolic disorders affecting the neuromuscular system.

By employing a combination of these diagnostic tools, healthcare providers can piece together a comprehensive picture of the patient’s condition. This thorough approach not only facilitates timely diagnoses but also enables the formulation of tailored care plans, ultimately enhancing the patient's overall health and significantly improving their quality of life.

1.12 Neuromuscular Disorder Projects and Their Alignment with Sustainable Development Goals (SDGs)

Projects focused on neuromuscular disorders primarily align with Sustainable Development Goal 3 **(SDG 3): Good Health and Well-Being.** This goal emphasizes the importance of ensuring access to quality healthcare and improving patient outcomes. Such projects contribute to early diagnosis, effective treatments, and rehabilitation, helping patients lead healthier lives. Furthermore, advancements in medical research, assistive technologies, and innovative therapies support universal health coverage (Target 3.8) and help reduce the disease burden.

Additionally, these projects relate to **SDG 9: Industry, Innovation, and Infrastructure**. Research in neuromuscular disorders promotes biomedical innovation, genetic therapies, and AI-driven diagnostics. The development of advanced healthcare infrastructure and medical devices enhances treatment options and accessibility.

SDG 10: Reduced Inequalities is also relevant, as individuals with neuromuscular disorders often face challenges in accessing healthcare and social services. Projects that promote affordable treatment, inclusive policies, and disability support help bridge this healthcare gap.

Moreover, these initiatives contribute to **SDG 4: Quality Education by ensuring that individuals** with neuromuscular disorders have access to inclusive learning environments. The use of assistive technologies and special education programs improves educational opportunities for those affected.

Finally, **SDG 8: Decent Work and Economic Growth plays** a role in promoting equal employment opportunities for people with neuromuscular conditions. Research on adaptive workplaces, vocational training, and employment policies aids individuals with disabilities in participating in the workforce and achieving financial independence.

REVIEW OF LITERATURE

CHAPTER II

REVIEW OF LITERATURE

2.1 Sales, Déborah Santos, Hammerle, Mariana Beiral et al., 2025, *Dysphagia and its Impact on Quality of Life in Rare Neuromuscular Disorders*. This comprehensive study investigates the occurrence of dysphagia, a condition marked by difficulties in swallowing, within a specific population of patients diagnosed with rare neuromuscular disorders (NMDs). It also explores the subsequent effects this condition has on their overall quality of life (QoL). The research involved a diverse group of 103 individuals, each diagnosed with various rare NMDs, and employed the Eating Assessment Tool-10 (EAT-10) to assess the risk of dysphagia. In addition, the study utilized the SWAL-QoL survey to gauge the quality of life specifically related to swallowing.The results unveiled a significant prevalence of dysphagia, with a striking 52.4% of participants reporting difficulties in swallowing. This condition was found to have a profound impact on multiple aspects of their quality of life. Key domains influenced included the desire to eat, the ability to select appropriate foods, communication capabilities, mental health status, and social functioning. Notably, the domain related to sleep appeared to be largely unaffected by dysphagia. A strong correlation emerged between the severity of dysphagia and a deteriorating quality of life, with significant statistical differences observed across various classifications of dysphagia severity ($p < 0.001$ for most assessed domains). Interestingly, the study revealed that there was no substantial difference in swallowing-related quality of life between patients who were classified as 'sitters' and those identified as 'walkers.' This finding suggests that the impact of dysphagia crosses the boundaries of mobility status, affecting all individuals regardless of their physical capabilities.In conclusion, this study highlights dysphagia as a widespread and pressing issue among individuals with rare NMDs, significantly compromising their quality of life. The findings underscore the urgent need for early detection and the development of customized management strategies aimed at improving outcomes for these patients.

2.2 Deenen, Johanna C.W. et al., 2024, *Population-based Incidence Rates of 15 Neuromuscular Disorders: A Nationwide Capture-Recapture Study in the Netherlands*. This extensive research delivers a meticulous estimate of the annual incidence rates for fifteen distinct neuromuscular disorders (NMDs) within the Netherlands, employing a sophisticated capture-recapture methodology that ensures a robust examination of these conditions. The incidence rates displayed considerable variability, illustrating the diverse nature of NMDs; for

instance, glycogenosis type 5 was recorded at a notably low incidence rate of just 0.03 per 100,000 individuals, while myotonic dystrophy type 1 emerged as a more common affliction, with a higher rate of 0.9 per 100,000 people. In summary, the cumulative incidence of neuromuscular disorders across the population reached 4.1 per 100,000, emphasizing the significance of these conditions. Crucially, the study unveiled that nine of the reported incidence rates were newly identified, thus providing invaluable insights that can aid the medical community in understanding these rare disorders more comprehensively. In contrast, several other rates confirmed existing data while surpassing previously established estimates, highlighting the evolving landscape of NMD research. These findings accentuate the urgent need for an automated data collection system, which would significantly enhance the monitoring, reporting, and understanding of these rare disorders. This is particularly pertinent given that the combined incidence of NMDs is markedly higher than that of some more prevalent neurological conditions, such as multiple sclerosis, thereby underscoring the necessity for heightened awareness and resources dedicated to these impactful health issues.

2.3 Margeta, Marta et al., 2023, Neuromuscular Disease: 2023 *Update*. This review thoroughly highlights ten pivotal advancements in the realm of neuromuscular disease research that emerged in 2022, delving into a diverse array of topics including neuromuscular biology, the identification of emerging diseases, the exploration of etiology, advancements in diagnostics, and innovative therapeutic approaches. It provides an insightful discussion on the complications arising from COVID-19, particularly how they intersect with various conditions, such as DNAJB4-associated myopathy, NMNAT2-deficient hereditary axonal neuropathy, Guillain-Barré syndrome, sporadic inclusion body myositis, and amyotrophic lateral sclerosis (ALS).The paper emphasizes the significant strides made in genetic testing methodologies for muscular dystrophies, showcasing how these advancements are enhancing diagnostic accuracy and patient outcomes. Furthermore, it highlights the promising potential of SARM1 inhibitors in the prevention of Wallerian degeneration, offering new avenues for therapeutic intervention. In addition, the review brings attention to the innovative incorporation of convolutional neural networks (CNNs) in diagnostic practices. This integration of cutting-edge artificial intelligence analysis is not only elevating our understanding of neuromuscular diseases but also significantly improving the management strategies employed by healthcare professionals, thereby enhancing patient care in this complex field.

2.4 Wilson, Lindsay A., Macken, William L. et al., 2023, *Neuromuscular Disease Genetics in Under-Represented Populations: Increasing Data Diversity*. Neuromuscular diseases (NMDs) affect approximately 15 million people worldwide, yet genetic diagnosis is often limited, especially in low-to-middle-income countries (LMICs), where 86% of research has focused on individuals of European ancestry. To address this disparity, a pioneering study formed a cloud-based partnership among 18 centers across Brazil, India, South Africa, Turkey, Zambia, the Netherlands, and the UK, recruiting 6,001 participants over 43 months. Using whole exome sequencing and a custom bioinformatics pipeline, researchers achieved a "solved" or "possibly solved" diagnosis for 56% of participants, with nearly 29% of identified disease-causing variants being novel. This initiative enhances genetic counseling and care pathways, while also improving eligibility for gene-specific clinical trials, promoting a more inclusive approach to genomic research in NMDs.By increasing the representation of diverse populations, the project lays the groundwork for a deeper understanding of the genetic basis of these diseases. It emphasizes the need to address health disparities and ensure equitable access to genomic resources. Ultimately, this international collaboration aims to drive innovations in the diagnosis and treatment of neuromuscular conditions, benefiting a broader spectrum of affected individuals globally.

2.5 Omar, Abdillahi, Marwaha, Komal et al., 2023, *Physiology, Neuromuscular Junction*
This article provides a comprehensive examination of the neuromuscular junction (NMJ), a critical synapse where motor neurons transmit action potentials to muscles, enabling contraction. It highlights the NMJ's significance as a target for various neuromuscular diseases and pharmacological agents. The article delves into the mechanisms of acetylcholinesterase inhibitors, including irreversible organophosphates like malathion and parathion, which cause acetylcholine accumulation and receptor overstimulation, leading to toxicity that can be reversed with atropine and pralidoxime. It categorizes neuromuscular blockers into depolarizing agents (e.g., succinylcholine), which induce sustained depolarization leading to paralysis, and non-depolarizing agents (e.g., tubocurarine, rocuronium), which act as competitive acetylcholine antagonists and are used in anesthesia. The article also discusses direct cholinergic agonists such as bethanechol, carbachol, and pilocarpine, which are used to treat conditions like postoperative ileus, glaucoma, and asthma diagnosis. Additionally, it highlights the medical applications of botulinum toxin in treating disorders like dystonia, blepharospasm, and achalasia by inhibiting acetylcholine release. The discussion extends to genetic factors affecting NMJ function, such as pseudocholinesterase deficiency and ryanodine

receptor mutations, which can prolong succinylcholine effects and trigger malignant hyperthermia. Overall, the article provides essential insights into NMJ physiology, pharmacology, and its implications in clinical medicine.

2.6 Wilson, Lindsay A., Macken, William L. et al., 2023, *Neuromuscular Disease Genetics in Under-Represented Populations: Increasing Data Diversity*. This study emphasizes the considerable disparity in the availability of genetic data for neuromuscular diseases (NMDs), especially in low-to-middle-income countries (LMICs), where access to DNA-based diagnostic tools remains limited. The majority of existing genetic research has primarily focused on individuals of European ancestry, resulting in a significant gap in our understanding of the genetics of NMDs in diverse populations. In response to this critical issue, a pioneering transcontinental cloud-based collaboration was formed, bringing together 18 research centers from countries such as Brazil, India, South Africa, Turkey, Zambia, the Netherlands, and the UK. This partnership aims to enhance genetic diversity in NMD studies, ensuring a more inclusive approach to genetic research. Over an extensive period of 43 months, the initiative successfully recruited 6,001 participants, employing advanced techniques such as whole exome sequencing and targeted genetic analysis. Remarkably, pathogenic or likely pathogenic variants were identified in 56% of the cases analyzed. Among these, an impressive 29% of the variants were novel, contributing significantly to the global body of knowledge regarding NMD genetics. This collaborative effort not only enhances the accuracy of genetic diagnoses and the quality of genetic counseling but also paves the way for the development of targeted gene-specific therapies that can transform the treatment landscape for NMDs. Furthermore, it serves as a powerful example of how global partnerships can tackle genetic data inequalities and improve healthcare outcomes for underserved populations. To statistically validate the findings, robust methods such as ANOVA may have been utilized to analyze variations in genetic markers across different geographical and ethnic populations, ensuring the reliability of the results.

2.7 Tayfur, Beyza, Charuphongsa, Chedsada et al., 2023, *Neuromuscular Joint Function in Knee Osteoarthritis: A Systematic Review and Meta-Analysis*. This study aimed to provide a comprehensive understanding of neuromuscular alterations in individuals with knee osteoarthritis (KOA) to inform targeted rehabilitation strategies. A systematic review and meta-analysis were conducted using studies retrieved from five databases up to October 2020, selecting 7 high-quality and 22 moderate-quality studies based on a modified Downs and Black checklist. A total of 1146 KOA patients and 1353 age- and sex-matched controls were analyzed,

focusing on neuromuscular function, including muscle strength, voluntary activation, cortical and spinal-reflex excitability, and torque-related outcomes. Findings revealed that KOA patients had quadriceps and hamstring strength deficits and increased hamstring-to-quadriceps strength ratios across different severities. Women showed lower quadriceps strength in early KOA and reduced voluntary activation in end-stage KOA, whereas men did not. Additionally, quadriceps force control ability was reduced, but no significant changes were observed in rapid force production or muscle size, except for the vastus medialis. The study concludes that neuromuscular deficits in KOA involve strength, activation, muscle size, and force control impairments, with women potentially being more affected than men, highlighting the need for sex-specific rehabilitation approaches.

2.8 Lace, Baiba, Micule, Ieva et al., 2022, *Overview of Neuromuscular Disorder Molecular Diagnostic Experience for the Population of Latvia*. This study delves into the critical role of genetic testing as the primary diagnostic instrument for neuromuscular disorders (NMDs) in Latvia, highlighting its effectiveness in identifying hereditary conditions and determining disease prevalence. Employing advanced next-generation sequencing (NGS) technologies, the research successfully confirmed diagnoses in 153 distinct cases, achieving an impressive detection rate of 37%. Among the most frequently identified childhood-onset NMDs were spinal muscular atrophy (SMA) and dystrophinopathies, which exhibited notable birth prevalence rates of 1.01 per 10,000 and 2.08 per 10,000 male newborns, respectively. Additionally, the study calculated the point prevalence for several adult-onset NMDs, revealing significant figures such as facioscapulohumeral muscular dystrophy at a rate of 0.079 per 10,000 individuals, limb-girdle muscular dystrophy at 0.078 per 10,000, and myotonic dystrophy type 1 at 0.047 per 10,000. To support these findings, comprehensive statistical analyses were conducted utilizing demographic data sourced from the Central Statistical Bureau of Latvia, which facilitated an evaluation of both birth prevalence and the frequency of these disorders within the adult population. The study ultimately concludes that genetic testing, particularly through the innovative use of NGS, emerges as an invaluable and effective approach for the early and accurate diagnosis of NMDs, essential not only for enhancing patient management strategies but also for steering future research initiatives and developing effective genetic screening programs aimed at improving overall health outcomes.

2.9 Saburova, Ekaterina A., Vasiliev, Alexander N. et al., 2017, *Human APP Gene Expression Alters Active Zone Distribution and Spontaneous Neurotransmitter Release at the Drosophila Larval Neuromuscular Junction*. This study investigates the impact of human amyloid precursor protein (APP) expression on synaptic organization and neurotransmitter release at the neuromuscular junction (NMJ) of Drosophila melanogaster. Through the use of transgenic lines that overexpress human APP, the research reveals an increase in synaptic bouton formation, yet a reduction in the number of active zones (AZs) per bouton, as indicated by changes in Bruchpilot (Brp) protein distribution. These structural alterations correlate with a decrease in the frequency of miniature excitatory junction potentials (mEJPs), suggesting lower spontaneous neurotransmitter release probability. Interestingly, while the total AZ count per NMJ remains constant, APP appears to affect vesicular exocytosis rather than AZ formation. Further analysis indicates APP's role in synaptic vesicle turnover and its interaction with key presynaptic proteins like synaptotagmin-1 and synaptobrevin, supporting the idea that APP modifies the structural and functional properties of AZs, thus contributing to synaptic dysfunction. Statistical methods such as ANOVA are likely employed to validate differences in neurotransmitter release frequencies across experimental conditions.

2.10 Woodcock, Ian R., Fraser, Lorna et al., 2016, *The Prevalence of Neuromuscular Disease in the Paediatric Population in Yorkshire, UK; Variation by Ethnicity and Deprivation Status*. This study sought to investigate the prevalence of neuromuscular diseases (NMDs) among children aged 16 and under in the Yorkshire region of the UK, with a particular focus on how these rates varied across different ethnic groups and levels of socioeconomic deprivation. The research involved a thorough retrospective review of medical records from 261 confirmed NMD cases identified at a single medical center in 2010, set against a backdrop of a pediatric population totaling 707,961. The findings revealed an overall prevalence of neuromuscular diseases at a rate of 36.9 per 100,000 children. Notably, dystrophin-related muscle diseases emerged as the most prevalent type, affecting 16.9 per 100,000 males. The study uncovered significant ethnic disparities; specifically, South Asian children displayed an alarmingly high prevalence of NMDs at 91.2 per 100,000, starkly contrasted with the rate of 28.7 per 100,000 found among their White counterparts. Furthermore, it was noted that non-dystrophin-related NMDs were occurring four times more frequently within the South Asian demographic. In addition to the ethnic discrepancies, the research identified a concerning linear relationship between increased socioeconomic deprivation and higher rates of NMD prevalence. These findings indicate that not only are neuromuscular diseases disproportionately

higher among South Asian children, but they also affect those from socioeconomically disadvantaged backgrounds more severely. The study emphasizes the urgent need for targeted healthcare interventions and appropriate allocation of resources to address these significant health disparities.

2.11 Deenen, Johanna C.W., Horlings, Corinne G.C. et al., 2015, *The Epidemiology of Neuromuscular Disorders: A Comprehensive Overview of the Literature*
This study offers an in-depth examination of the epidemiology of neuromuscular disorders (NMDs) through a thorough analysis of literature published since 1990, highlighting significant advancements in diagnostic techniques, particularly the incorporation of genetic testing, which has greatly improved the accuracy of diagnoses for these complex conditions. The research presents incidence rates for ten distinct NMDs, revealing that most disorders have rates ranging between 1 and 10 cases per 100,000 individuals, while conditions like Charcot-Marie-Tooth disease and post-polio syndrome exhibit notably higher prevalence rates, exceeding 10 cases per 100,000. In contrast, congenital muscular dystrophies tend to present with lower prevalence rates, underscoring the diverse impact of these disorders. Furthermore, the study identifies substantial gaps in the epidemiological data for specific NMDs, particularly chronic inflammatory demyelinating polyneuropathy and Pompe's disease, indicating a critical need for targeted research in these areas. Ultimately, the conclusion emphasizes that while each individual NMD may be rare, their collective prevalence is comparable to that of Parkinson's disease, highlighting the significance of NMDs in the broader context of public health research and the ongoing necessity for focused attention and resources dedicated to understanding and addressing these conditions.

2.12 Silva, M. Eduarda, Mendonça, Teresa et al., 2005, *Statistical Analysis of Neuromuscular Blockade Response: Contributions to an Automatic Controller Calibration*.
This comprehensive study delves into the complexities of neuromuscular blockade, particularly focusing on the efficacy and response to various muscle relaxant drugs commonly utilized in surgical interventions. It introduces an innovative methodology aimed at optimizing automatic controllers responsible for managing neuromuscular blockade. This optimization process is tailored specifically to the distinct physiological characteristics of each patient, thereby enhancing individualized care.By employing sophisticated statistical techniques, such as principal component analysis (PCA) and Walsh–Fourier spectral analysis, the research meticulously extracts critical features from the neuromuscular response data collected following an initial bolus dose administered at the onset of anesthesia. This in-depth analysis

significantly improves the controller's capability for real-time autocalibration through the application of multiple linear regression models, ensuring that it adapts efficiently to the inherent variations present among different patients.To further enrich the findings, Analysis of Variance (ANOVA) is utilized to assess the significance of differences in average neuromuscular responses across diverse patient demographic groups. This analytical approach provides a nuanced understanding of the variability in neuromuscular responses, which is crucial for tailoring anesthetic strategies effectively.The study also rigorously validates several characterization methodologies by comparing both observed and simulated data, thereby illuminating the relationship between muscle relaxant administration and patient neuromuscular responses. Ultimately, the research aspires to enhance the precision of automated muscle relaxant delivery systems and, in doing so, improve overall surgical outcomes, paving the way for safer and more effective perioperative care.

2.13 Woodcock, Ian R., Fraser, Lorna et al., 2016, *The Prevalence of Neuromuscular Disease in the Paediatric Population in Yorkshire, UK; Variation by Ethnicity and Deprivation Status*. This study sought to investigate the prevalence of neuromuscular diseases (NMDs) among children aged 16 and under in the Yorkshire region of the UK, with a particular focus on how these rates varied across different ethnic groups and levels of socioeconomic deprivation. The research involved a thorough retrospective review of medical records from 261 confirmed NMD cases identified at a single medical center in 2010, set against a backdrop of a pediatric population totaling 707,961.The findings revealed an overall prevalence of neuromuscular diseases at a rate of 36.9 per 100,000 children. Notably, dystrophin-related muscle diseases emerged as the most prevalent type, affecting 16.9 per 100,000 males. The study uncovered significant ethnic disparities; specifically, South Asian children displayed an alarmingly high prevalence of NMDs at 91.2 per 100,000, starkly contrasted with the rate of 28.7 per 100,000 found among their White counterparts. Furthermore, it was noted that non-dystrophin-related NMDs were occurring four times more frequently within the South Asian demographic.In addition to the ethnic discrepancies, the research identified a concerning linear relationship between increased socioeconomic deprivation and higher rates of NMD prevalence. These findings indicate that not only are neuromuscular diseases disproportionately higher among South Asian children, but they also affect those from socioeconomically disadvantaged backgrounds more severely. The study emphasizes the urgent need for targeted healthcare interventions and appropriate allocation of resources to address these significant health disparities.

RESEARCH METHODOLOGY

CHAPTER III

RESEARCH METHODOLOGY

3.1 OBJECTIVES

- To understand the interplay between treatment adherence, disease severity, and patient outcomes in neuromuscular disorders.
- To investigate the influence of clinical and treatment factors on disease progression and patient strength in neuromuscular disorders.
- To evaluate different machine learning models for predicting the early-stage diagnosis of neuromuscular disorders and identify the most effective model for clinical use.
- To assess the best-performing model using ROC and AUC values for determining the stages of neuromuscular disorder patients.
- To predict major health issues in patients with neuromuscular disorders for more accurate diagnosis and treatment.
- To validate the best model by analyzing ROC and AUC values for predicting major health issues in neuromuscular disorder patients.

3.2 DATA TYPE

The data used in this study is sourced from the **National Institute of Health (NIH)**, collected under the **Neuromuscular Patient Record** up to 2024. This dataset includes patient records from several countries. It underwent multiple preprocessing steps, including the removal of missing (unentered) variables, outlier filtering, and duplicate elimination. After these refinements, the final dataset comprises **276 patient responses** and **35 variables**, making it suitable for this study.

3.3 KEY FEATURES OF THE DATASET

3.3.1. DEMOGRAPHIC INFORMATION

- **Age :** The age of the patient at the time data is collected, which can provide insights into how age may influence the progression of their condition.

- **Gender :** The biological sex of the patient, which may be relevant in understanding variations in disease presentation and treatment response.
- **BMI (Body Mass Index):** A calculated number derived from the patient's height and weight, serving as an important indicator of body composition and overall health.

3.3. 2. DISEASE CHARACTERISTICS

- **Disorder:** The specific type of neuromuscular disorder that has been diagnosed, which can guide treatment decisions and prognostic outlooks.
- **Duration (Years):** The total number of years since the patient received their diagnosis, reflecting the chronicity of their condition.
- **Severity Score:** A numerical representation that quantifies the severity of the disease, allowing for objective assessment and comparison over time.
- **Disease Activity Score:** A composite score that assesses various aspects of disease activity, providing a holistic view of the patient's current status.
- **Genetic Mutation Score:** A score that indicates the presence and implications of genetic mutations associated with the patient's disorder, which can inform treatment strategies.

3.3.3. TREATMENT INFORMATION

- **Treatment:** The specific type or types of interventions being utilized to manage the patient's condition, ranging from medication to physical therapy.
- **Treatment Start Year:** The year in which the patient commenced their treatment, offering context regarding the timing of interventions relative to the disease's progression.
- **Dosage:** The specific amount of medication or therapy prescribed to the patient, which is critical for ensuring effective treatment while minimizing side effects.
- **Number of Treatments:** A count of different therapeutic approaches the patient is receiving concurrently, providing insight into the complexity of their medical management.
- **Adherence to Treatment:** A measure of how consistently the patient follows the prescribed treatment regimen, which is vital for achieving optimal health outcomes.

3.3.4. CLINICAL ASSESSMENTS & LAB MEASUREMENTS

- **Muscle Strength:** An evaluation of the strength of the patient's muscles, important for assessing functional capabilities and tracking changes over time.
- **Nerve Conduction Velocity (NCV):** A test that measures how quickly electrical signals move through the nerves, serving as a key indicator of nerve function and integrity.
- **Respiratory Volume:** An assessment of lung capacity and functionality, critical for understanding respiratory health in patients with neuromuscular disorders.
- **Creatine Kinase Levels (CK Levels):** A blood test measuring levels of CK, an enzyme that indicates muscle damage, helping to monitor the impact of the disease over time.
- **Blood Pressure Systolic:** The measurement of the pressure in the arteries when the heart beats, providing valuable information about cardiovascular health.
- **Blood Pressure Diastolic:** The measurement of the pressure in the arteries when the heart is at rest between beats, which is essential for understanding overall cardiovascular function.

3.3.5. DISEASE PROGRESSION & OUTCOMES

- **Time to Disease Progression:** An estimate of the duration before the patient's condition is expected to worsen, helping to anticipate future healthcare needs.
- **Time to Death:** The interval from the initial diagnosis to the time of death, if applicable, which is important for understanding prognosis and planning.
- **Rate of Progression:** The speed at which the disease deteriorates, which can influence treatment choices and patient management strategies.
- **Early Stage:** A status indicating whether the disease is currently considered to be in an early stage (Yes/No), which can affect treatment options and outcomes.
- **Cured:** A status confirming whether the patient has achieved complete recovery, providing insight into the effectiveness of treatment and management strategies.
- **Died with Disease:** Indicates whether the patient passed away while still diagnosed with the disorder, relevant for understanding disease impact on survival.

- **Died during Treatment:** Indicates whether the patient passed away while undergoing treatment, which can inform discussions about treatment efficacy and patient safety.

3.3.6. TREATMENT & FUNCTIONAL IMPROVEMENTS

- **Muscle Strength Improvement:** An assessment of progress in muscle strength as a result of treatment, reflecting the effectiveness of the therapeutic interventions.
- **Functional Mobility Improvement:** A measure of the enhancements in the patient's ability to move and perform daily activities, indicative of overall functional health.
- **Respiratory Function Improvement:** Observations of positive changes in lung function, important for overall wellness and quality of life.
- **Quality of Life Improvement:** An overall evaluation of enhancements in the patient's well-being, encompassing physical, emotional, and social factors.

3.3.7. RISK FACTORS & INTERACTIONS

- **Major Issue Type:** The most significant health issue associated with the patient's disorder, which can guide targeted interventions and monitoring.
- **Severity Interaction:** An exploration of how different severity-related factors interact, providing deeper insights into the complexities of the patient's condition.
- **Treatment Interaction:** An examination of how different treatment types affect one another, offering crucial information for optimizing therapeutic strategies.

3.4 RESEARCH METHODOLOGY

This study employs statistical analysis along with supervised and unsupervised machine learning techniques to explore variable relationships, identify patterns, and develop predictive models for neuromuscular disorders. Statistical methods are used to assess variable associations and data distributions, while dimensionality reduction techniques help manage high-dimensional data efficiently. Predictive modeling enhances early diagnosis and prognosis, leading to more accurate disease assessment. Finally, model performance evaluation ensures the reliability and effectiveness of predictive models in clinical applications. This integrated approach provides valuable insights into disease progression, treatment effectiveness, and patient outcomes.

- Exploratory Data Analysis
- Chi-Square
- Analysis Of Variance
- Random Forest
- Support Vector Machine (SVM)
- Neural Network
- Logistic Regression
- Extreme Gradient Boosting (XG Booster)
- Categorical Booster (CatBooster)
- Generalized Linear Model via Elastic Net (GLMNET)
- Decision Tree
- Factor Analysis for Mixed Data (FAMD)
- Area Under the Curve (AUC)
- Receiver Operating Characteristic (ROC)

3.4.1 EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. Its main goals are to gain insights into the data, uncover patterns, detect anomalies, and formulate hypotheses for further investigation.

Steps in EDA include:

1. **Data Collection:** Gathering the dataset from various sources, such as databases, files, or APIs.

2. **Data Cleaning:** Preprocessing the data by handling missing values, removing duplicates, and correcting errors to ensure data quality.

3. **Descriptive Statistics:** Calculating summary statistics such as mean, median, standard deviation, and quartiles to describe the central tendency and spread of the data.

4. Data Visualization: Creating visual representations of the data using plots and charts such as line plot , box plots, scatter plots, and heatmaps to explore relationships and distributions within the data.

5. Exploratory Modeling: Building simple models or applying statistical tests to understand relationships between variables and make preliminary predictions.

6. Pattern Recognition: Identifying patterns, trends, and outliers in the data that may require further investigation.

7. Hypothesis Generation: Formulating hypotheses or questions based on initial observations for further analysis.

EDA is an essential first step in any data analysis project as it helps analysts understand the structure and characteristics of the data, identify potential problems, and guide subsequent modeling and analysis decisions.

3.4.2 CHI SQUARE

The Chi-Square test (χ^2 test) is a statistical method used to determine whether there is a significant association between two categorical variables. It compares the observed frequencies of different categories with the expected frequencies under the assumption of independence. The test is widely used in medical and epidemiological studies to analyze relationships between factors such as treatment adherence and disease severity. The Chi-Square statistic is calculated using the formula:

$$\chi 2 = \sum \frac{(O - E)^2}{E}$$

where O represents the observed frequency, and E represents the expected frequency. If the calculated p-value is less than 0.05, we reject the null hypothesis, indicating a significant association between the variables. In this study on neuromuscular disorders, the Chi-Square test can be applied to assess whether treatment adherence influences disease progression, whether genetic mutations are linked to early-stage detection, or whether specific treatment types affect major health issues. The results help in understanding key factors affecting patient outcomes and guiding clinical decision-making.

3.4.3 ANALYSIS OF VARIANCE (ANOVA)

Analysis of Variance (ANOVA) is a statistical technique used to compare the means of multiple groups to determine if there is a significant difference between them. It helps analyze whether at least one group mean is significantly different from the others by examining the variation within and between groups. The ANOVA test is based on the F-statistic, which is calculated as the ratio of between-group variance to within-group variance:

$$F = \frac{Between - group\ variance}{Within\ -\ group\ variance}$$

If the p-value obtained from the test is less than 0.05, it suggests a significant difference between at least one of the group means. ANOVA is classified into different types: One-Way ANOVA, which compares the means of one independent variable with multiple groups; Two-Way ANOVA, which examines the effect of two categorical variables and their interaction; and Repeated Measures ANOVA, which analyzes changes in the same group over time. The test assumes that the dependent variable is continuous, the data follows a normal distribution, and the groups have equal variances. If ANOVA results are significant, post-hoc tests such as Tukey's HSD are performed to identify which specific groups differ. In the context of neuromuscular disorder research, ANOVA can be used to compare muscle strength improvement across different treatment types, analyze whether patients from different countries respond differently to the same treatment, or assess the impact of treatment adherence on disease severity scores. This analysis helps in understanding treatment effectiveness and guiding clinical decisions.

3.4.4 RANDOM FOREST

Random Forest is an ensemble machine learning algorithm that builds multiple decision trees and combines their outputs to improve accuracy and reduce overfitting. It works by creating a "forest" of decision trees, where each tree is trained on a random subset of the data using a technique called bootstrap aggregation (bagging). The final prediction is made by aggregating the outputs of all trees—either by majority voting for classification or averaging for regression.

Two key hyperparameters in Random Forest are ntree and mtry:

1. **ntree (Number of Trees):**
    - This parameter controls the number of decision trees in the forest.
    - A higher number of trees generally improves model performance but increases computational cost.
    - The default value in many implementations is 500, but optimal values depend on the dataset.

2. **mtry (Number of Features to Consider at Each Split):**

   - This parameter determines how many features are randomly selected at each split of a decision tree.
   - For classification problems, the default is usually $\sqrt{p}$, where P is the total number of features.
   - For regression problems, the default is p/3.

Random Forest is effective for both classification and regression tasks, handling missing values well and providing insights into feature importance. However, it can be computationally expensive for large datasets and less interpretable compared to simpler models.

3.4.5 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression tasks. It works by identifying the optimal hyperplane that best separates different classes in the feature space. The support vectors are the data points that lie closest to the hyperplane, and SVM maximizes the margin between these points and the decision boundary to improve generalization.

For linearly separable data, SVM finds a straight-line (or plane in higher dimensions) boundary. For non-linearly separable data, it utilizes the kernel trick to project the data into a higher-dimensional space where it can be separated more effectively.

3.4.5.1 Key Hyperparameters in SVM

1. **C (Regularization Parameter):**

   - Controls the trade-off between achieving a large margin and minimizing misclassification.
   - A small C allows more misclassifications but results in a wider margin, promoting generalization.
   - A large C reduces misclassifications but may lead to overfitting.

2. **Kernel Type (kernel):**

   - Determines how data is mapped into a higher-dimensional space for better separation.
   - Common kernel types include:
     - **Linear Kernel**:$K(x_i, x_j) = x_i \cdot x_j K(xi, xj) = xi \cdot xj$ (suitable for linearly separable data).
     - **Polynomial Kernel**: $K(xi, xj) = (xi \cdot xj + r)\hat{}d$ (captures complex relationships).
     - **Radial Basis Function (RBF) Kernel**: $K(xi, xj) = e - \gamma \,||\, xi - xj \,||\, 2K(x_i, x_j) = e^{-\gamma||x_i - x_j||^2} K(xi, xj) = e - \gamma \,||\, xi - xj \,||\, 2$ (widely used for non-linear data).
     - **Sigmoid Kernel**:

       $K(xi, xj) = \tan h(\alpha xi \cdot xj + c)\, K(x_i, x_j) = \tanh(\alpha x_i \cdot x_j + c)\, K(xi, xj) = \tanh(\alpha xi \cdot xj + c)$(similar to neural networks).

3. **Gamma (γ) (for RBF and Polynomial Kernels):**

   - Determines how much influence a single training point has on the model.
   - A small γ results in a smoother decision boundary, leading to better generalization.
   - A large γ makes the model more sensitive to individual data points, increasing the risk of overfitting.

4. **Degree (d) (for Polynomial Kernel):**

   - Specifies the degree of the polynomial kernel function.
   - Higher values allow capturing more complex patterns in the data.

5. **Epsilon (ε) (for SVM Regression):**

   - Defines the margin within which predictions are considered correct.

- A smaller **ε** makes the model more sensitive, while a larger value allows more tolerance.

3.4.6 NEURAL NETWORK

A **Neural Network** is a computational model inspired by the human brain, consisting of interconnected layers of neurons that process and learn patterns from data. It is widely used in various machine learning tasks due to its ability to model complex, non-linear relationships. A neural network typically consists of **three main layers**:

1. **Input Layer:** Receives raw data and passes it to the network for processing.
2. **Hidden Layers:** Perform computations through multiple neurons, each applying an activation function to capture non-linearity. The number of hidden layers and neurons impacts model complexity.
3. **Output Layer:** Produces the final prediction, whether a classification label, probability score, or numerical value.

Training Process:

- The network learns by adjusting **weights** through an optimization algorithm like **Stochastic Gradient Descent (SGD) or Adam** to minimize the **loss function** (e.g., Mean Squared Error for regression, Cross-Entropy for classification).
- **Backpropagation** is used to update weights by computing the gradient of the loss function with respect to each weight and adjusting them accordingly.

Hyperparameters:

- **Learning Rate:** Controls the step size of weight updates.
- **Number of Layers & Neurons:** Defines model depth and complexity.
- **Activation Functions:** Such as ReLU, Sigmoid, or Tanh, influence how neurons pass information.
- **Batch Size & Epochs:** Determines how many samples are processed before weight updates and how many times the model iterates over the dataset.
- **Regularization (L1, L2, Dropout):** Prevents overfitting by penalizing large weights or randomly deactivating neurons during training.

Neural networks are effective for pattern recognition, classification, and regression tasks. However, they require large amounts of data and computational power to train effectively while balancing generalization and performance.

3.4.7 LOGISTIC REGRESSION

Logistic Regression is a statistical and machine learning model used for binary and multi-class classification problems. Unlike linear regression, which predicts continuous values, logistic regression estimates the probability of an outcome belonging to a particular class using the logistic (sigmoid) function. Logistic regression models the relationship between independent variables (**X**) and the probability (**p**) of a dependent variable (**Y**) belonging to class 1. The model is expressed as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

where:

- p is the predicted probability of the outcome being 1,
- β_0 is the intercept,
- $\beta_1, \beta_2, \ldots \beta_n$ are the regression coefficients,
- $X_1, X_2, \ldots, X_n$ are the independent variables,
- e is the base of the natural logarithm.

Decision Rule:

- If $p \geq 0.5$, the model classifies the outcome as **1** (positive class).
- If $p < 0.5$, the model classifies the outcome as **0** (negative class).

Logistic regression is advantageous due to its simplicity, interpretability, and efficiency in handling classification problems, especially when the relationship between independent and dependent variables follows a linear trend in the log-odds scale. However, it has limitations, such as its assumption of no multicollinearity among predictors, difficulty in capturing non-linear relationships unless transformations or interaction terms are introduced, and sensitivity to imbalanced data, which may require resampling or class-weight adjustments. Extensions of logistic regression include Multinomial Logistic Regression, which is used for multi-class

classification problems, and Ordinal Logistic Regression, which is applied when the dependent variable has ordered categories, making it a versatile model for various applications.

3.4.8 EXTREME GRADIENT BOOSTING (XGBOOST)

Extreme Gradient Boosting (XGBoost**)** is a powerful, optimized machine learning algorithm based on gradient boosting, designed for both classification and regression tasks. It improves model performance by building an ensemble of decision trees in a sequential manner, where each tree corrects the errors of the previous one. XGBoost is widely used due to its efficiency, scalability, and ability to handle missing values and feature interactions effectively.

Working Mechanism

XGBoost minimizes a given loss function by sequentially adding weak learners (decision trees) while applying **gradient descent optimization** to update predictions. The key objective is to reduce the residual errors from previous iterations while regularizing the model to prevent overfitting. The general form of the objective function is:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k} \Omega(f_k)$$

where:

- $l(y_i, \hat{y_i})$ is the loss function measuring the difference between actual and predicted values,
- $\Omega(fk)$ is a regularization term controlling the complexity of trees to prevent overfitting,
- θ represents the parameters of the model.

Important Hyperparameters

- **n_estimators (Number of Trees):** The total number of boosting rounds (trees) in the model.
- **learning_rate (Shrinkage Factor):** Controls the contribution of each tree; smaller values improve generalization but require more trees.
- **max_depth (Tree Depth):** Limits the depth of trees to prevent overfitting.
- **subsample (Row Sampling):** Specifies the fraction of samples used to train each tree to introduce randomness.

- **colsample_bytree (Feature Sampling):** Determines the fraction of features used in each tree to reduce correlation among trees.
- **gamma (Minimum Loss Reduction):** Controls tree pruning by requiring a minimum loss reduction for further partitioning.
- **lambda and alpha (L2 and L1 Regularization):** Prevent overfitting by penalizing large coefficients.

XGBoost is a powerful and efficient gradient boosting algorithm known for its scalability and high predictive accuracy. It incorporates regularization techniques, including both L1 (Lasso) and L2 (Ridge) penalties, to control model complexity and reduce overfitting. One of its key advantages is automatic handling of missing data, where the model learns optimal splits without requiring imputation. It supports parallel processing using multi-threading, significantly enhancing computational efficiency. Unlike traditional methods, XGBoost employs tree pruning by setting a maximum depth rather than relying on greedy pruning, improving generalization. Additionally, its sparsity-aware mechanism effectively manages sparse datasets with missing or zero values. These features make XGBoost highly efficient for large datasets, supporting distributed computing for faster training. Due to its robustness and accuracy, it is widely used in real-world applications such as fraud detection, recommendation systems, and medical diagnosis.

3.4.9 CATEGORICAL BOOSTING (CATBOOST)

CatBoost is a gradient boosting algorithm specifically designed for handling categorical features efficiently without requiring extensive preprocessing. Unlike other boosting methods that require encoding categorical variables (e.g., one-hot encoding), CatBoost utilizes ordered boosting and target-based encoding to manage categorical data effectively while reducing overfitting. It employs an innovative ordered boosting technique, which prevents target leakage by processing categories in a way that avoids using future information when encoding training samples.

CatBoost is highly efficient in handling high-cardinality categorical features, making it ideal for applications with a large number of categorical variables. It also supports GPU acceleration, enabling fast training even on large datasets. Additionally, the algorithm incorporates built-in regularization, which reduces overfitting and enhances model generalization.

Key Features of CatBoost:

- **Native Handling of Categorical Data:** Eliminates the need for extensive preprocessing like one-hot encoding.
- **Ordered Boosting:** Prevents target leakage by encoding categorical variables in a way that respects temporal order.
- **High Efficiency:** Supports GPU acceleration for faster model training.
- **Robust to Overfitting:** Uses built-in regularization techniques to improve generalization.

Due to its superior handling of categorical data and strong predictive performance, CatBoost is widely used in real-world applications, including finance, e-commerce, recommendation systems, and medical diagnosis.

3.4.10 GENERALIZED LINEAR MODEL VIA ELASTIC NET (GLMNET)

GLMNET is a regularized regression technique that extends the Generalized Linear Model (GLM) by incorporating the Elastic Net penalty, which combines L1 (Lasso) and L2 (Ridge) regularization. This combination provides a balance between feature selection and coefficient shrinkage, making GLMNET highly effective in handling high-dimensional datasets with multicollinearity.

The GLM framework allows the model to handle different types of response variables (e.g., binary, count, continuous) by using appropriate link functions (e.g., logit for logistic regression, log for Poisson regression). The Elastic Net regularization helps in cases where features are highly correlated, as Lasso alone may arbitrarily select one feature and ignore the others. The mixing parameter α controls the balance between L1 and L2 penalties, where $\alpha = 1$ corresponds to Lasso, $\alpha = 0$ corresponds to Ridge, and values between 0 and 1 create a hybrid model.

Key Features of GLMNET:

- **Handles High-Dimensional Data:** Works well with datasets where the number of features exceeds the number of observations.
- **Regularization for Feature Selection:** L1 helps in selecting important variables, while L2 prevents excessive shrinkage.

- **Efficient Optimization:** Uses a coordinate descent algorithm, making it computationally efficient for large datasets.
- **Flexible for Various Distributions:** Supports different GLM families like Gaussian, Binomial, and Poisson regression.

GLMNET is widely applied in predictive modeling, medical research, finance, and genetics, where feature selection and model interpretability are critical.

3.4.11.DECISION TREE

A **Decision Tree** is a supervised learning algorithm used for both **classification** and **regression** tasks. It works by recursively splitting the dataset into subsets based on feature values, forming a tree-like structure where each internal node represents a decision based on a feature, branches denote possible outcomes, and leaf nodes represent final predictions.

The splitting process is guided by measures such as **Gini impurity** or **entropy** (for classification) and **mean squared error (MSE)** (for regression). The algorithm selects the best feature at each step that maximizes the separation of data points into pure subsets. The tree expands until it reaches a stopping criterion, such as a maximum depth or a minimum number of samples per leaf.

Key Features:

- **Simple and Interpretable:** Provides an easy-to-understand decision-making process.
- **Handles Both Categorical and Numerical Data:** Can be applied to a wide range of datasets.
- **No Need for Feature Scaling:** Does not require normalization or standardization.
- **Captures Non-Linear Relationships:** Capable of modeling complex interactions between variables.

Limitations:

- **Prone to Overfitting:** A deep tree may fit the training data too well and generalize poorly.
- **Sensitive to Small Variations:** Small changes in data can lead to entirely different tree structures.

- **Less Effective for Highly Correlated Features:** May not perform well when multiple features provide similar information.

To overcome these limitations, pruning techniques (such as reduced-error pruning and cost-complexity pruning) and ensemble methods like Random Forest and Gradient Boosting are commonly used to improve decision tree performance.

3.4.12 FACTOR ANALYSIS FOR MIXED DATA (FAMD)

Factor Analysis for Mixed Data (FAMD) is a dimensionality reduction technique designed to handle datasets that contain both continuous (numeric) and categorical variables. It is particularly useful in exploratory data analysis, where understanding relationships between mixed data types is crucial. FAMD is an extension of Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA), combining their strengths to process heterogeneous data efficiently.

How FAMD Works:

- **Standardization:** Continuous variables are standardized (mean = 0, variance = 1), while categorical variables are transformed into indicator variables and given appropriate weights.
- **Singular Value Decomposition (SVD):** Like PCA, FAMD decomposes the data matrix to extract principal components that best explain the variance in both numerical and categorical data.
- **Balanced Contribution:** Unlike traditional PCA, FAMD ensures that both variable types contribute equally to the component construction, preventing dominance by continuous variables.

Key Features:

- **Handles Mixed Data Types:** Suitable for datasets containing both categorical and continuous variables.
- **Reduces Dimensionality:** Extracts principal components to simplify high-dimensional datasets.
- **Preserves Data Structure:** Maintains the relationships between numeric and categorical variables.

Applications of FAMD:

- Customer segmentation in marketing (e.g., combining demographics and purchasing behavior).
- Medical studies integrating patient symptoms (categorical) and lab measurements (continuous).
- Social science research involving survey data with both numerical scores and categorical responses.

FAMD is particularly useful in preprocessing for clustering, classification, or predictive modeling, where mixed data is a challenge.

3.4.13 CONFUSION MATRIX AND ITS METRICS

A confusion matrix is a performance evaluation tool used in classification models to compare predicted outcomes with actual outcomes. It provides a summary of the number of correct and incorrect predictions, categorized by class. The confusion matrix is especially useful in imbalanced datasets where accuracy alone may be misleading.

Structure of a Confusion Matrix

Actual / Predicted	Positive Prediction	Negative Prediction
Positive	True Positive	False Negative
Negative	False Positive	True Negative

- **True Positives (TP):** Correctly predicted positive cases.
- **False Negatives (FN):** Incorrectly predicted as negative when they are actually positive.
- **False Positives (FP):** Incorrectly predicted as positive when they are actually negative.
- **True Negatives (TN):** Correctly predicted negative cases.

Key Performance Metrics Derived from the Confusion Matrix

1. **Accuracy:** Measures the overall correctness of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Works well when classes are balanced but may be misleading in imbalanced datasets.

2. **Precision (Positive Predictive Value - PPV):** Measures the proportion of correctly predicted positive cases among all positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

High precision means fewer false positives.

3. **Recall (Sensitivity or True Positive Rate - TPR):** Measures the proportion of actual positives correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

High recall means fewer false negatives.

4. **F1-Score:** Harmonic mean of precision and recall, balancing both metrics.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Useful when precision and recall need to be balanced.

5. **Specificity (True Negative Rate - TNR):** Measures the proportion of actual negatives correctly identified.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Important in cases where false positives need to be minimized (e.g., fraud detection).

6. **False Positive Rate (FPR):** Measures the proportion of false positives among actual Lower values indicate better performance in distinguishing negative cases.

$$\text{FPR} = \frac{FP}{FP + TN}$$

7. **False Negative Rate (FNR):** Measures the proportion of false negatives among actual positives.

$$\text{FNR} = \frac{FN}{FN + TP}$$

Lower values indicate better performance in identifying positive cases

8. **Kappa (Cohen's Kappa):**

Cohen's Kappa (κ) is a statistical measure used to evaluate the level of agreement between two raters or classifiers, beyond what would be expected by chance. It is especially useful in classification problems where multiple raters or models label the same set of instances. The equation for Cohen's Kappa is:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Where:

- P_o = Observed agreement (proportion of times both raters agree).
- P_e = Expected agreement (agreement due to chance).
- K = 1 → Perfect agreement.
- K = 0 → Agreement is due to chance.
- K < 0 → Agreement is worse than random chance.
  It is commonly used in machine learning and inter-rater reliability studies to validate classification models.

9. **No Information Rate (NIR)**

The No Information Rate (NIR) represents the accuracy of a naive classifier that always predicts the most frequent class in the dataset. It serves as a baseline to compare the actual model's accuracy.

$$NIR = \frac{\max(\text{class frequencies})}{\text{total observations}}$$

If a model's accuracy is lower than or close to NIR, it indicates that the model is not performing significantly better than a simple majority-class prediction.

10. **McNemar's Test P-Value**

McNemar's test is a statistical test used to compare two paired classification models to determine if their predictions significantly differ. It is particularly useful in evaluating whether a new model provides a significant improvement over an existing one.

The McNemar's test statistic is computed as:

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

Where:

- b= Number of cases misclassified by model 1 but correctly classified by model 2.
- c = Number of cases misclassified by model 2 but correctly classified by model 1.

The test follows a chi-square ($\chi 2$) distribution with 1 degree of freedom, and the p-value indicates whether the difference in classification performance is statistically significant. A p-value < 0.05 suggests a significant difference between the models.

The confusion matrix provides a detailed view of a model's classification performance beyond simple accuracy. Choosing the right metric (precision, recall, F1-score, specificity, etc.) depends on the problem's nature, such as whether false positives or false negatives are more critical.

3.4.14 Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)

The Receiver Operating Characteristic (ROC) curve is a graphical representation of a classification model's performance at different threshold values. It plots the True Positive Rate (TPR) (also called sensitivity) against the False Positive Rate (FPR). A model with a perfect classification would have a curve that passes through the top-left corner, indicating high sensitivity with minimal false positives.

The Area Under the Curve (AUC) quantifies the overall performance of the model by measuring the area under the ROC curve. An AUC value closer to 1 indicates a highly effective model, while an AUC of 0.5 suggests random guessing. The AUC helps compare models objectively, as higher AUC values indicate better discrimination between classes.

These metrics are particularly useful in imbalanced datasets where accuracy alone may be misleading. A well-performing model balances sensitivity and specificity, ensuring optimal classification across various decision thresholds.

3.5 SOFTWARES USED FOR ANALYSIS

In this study, **Microsoft Excel, R Programming, and Jamovi** were used to perform data preprocessing, statistical analysis, machine learning modeling, and data visualization. Each tool serves a specific purpose, contributing to a comprehensive and structured approach to analysis.

3.5.1 MICROSOFT EXCEL

Microsoft Excel is a widely used tool for data management, offering functionalities for organizing, cleaning, and processing raw datasets. It provides built-in statistical functions for calculating measures such as mean, median, standard deviation, and correlation. Pivot tables help in summarizing and exploring relationships between variables, while conditional formatting allows for easy pattern detection in datasets. Excel's Solver tool is useful for optimization problems, and the Data Analysis ToolPak provides additional statistical capabilities, including regression analysis and hypothesis testing. Though Excel has limitations in handling large datasets and advanced statistical modeling, it remains an essential tool for quick data exploration, visualization, and reporting.

3.5.2 R PROGRAMMING

R is a powerful and flexible programming language specifically designed for statistical computing and data science. It supports a wide range of statistical methods, including descriptive analysis, hypothesis testing, regression modeling, classification, clustering, and time series analysis. The extensive package ecosystem in R features tools such as ggplot2 for advanced data visualization, dplyr and tidyr for data manipulation, caret and mlr3 for machine learning modeling, glmnet for regularized regression techniques, and factoextra for multivariate techniques like PCA and clustering. These tools facilitate dimensionality reduction, predictive modeling, and hypothesis testing, making R a preferred choice for statistical and machine learning applications. Additionally, its ability to integrate with other programming languages like Python and SQL enhances its versatility.

3.5.3 JAMOVI

Jamovi is an open-source statistical software that simplifies data analysis through a point-and-click graphical interface while maintaining the power of R in the backend. It supports a variety of statistical tests, including t-tests, ANOVA, chi-square tests, correlation analysis, logistic regression, and principal component analysis (PCA). The software is particularly useful for users who prefer not to write code but still require advanced statistical functionalities. Jamovi's real-time analysis and instant visualization features help in interpreting results efficiently. It also supports extensions through R scripting, making it adaptable for more complex statistical modeling.

While Excel is useful for data entry, organization, and preliminary exploration, R is superior in handling large datasets, conducting advanced statistical modeling, and building machine learning models. Jamovi acts as an intuitive alternative for statistical tests without coding, making it accessible for users unfamiliar with programming. The combination of these tools ensures a structured and efficient workflow for data-driven research, from basic statistical testing to advanced predictive modeling.

ANALYSIS AND INTERPRETATION

CHAPTER IV

ANALYSIS AND INTERPRETATION

4.1 COMPARING EARLY STAGE CONDITION OF THE PATIENT AND THEIR ADHERENCE TO TREATMENT

NULL HYPOTHESIS

H_0 : There is no association between early stage and adherence to treatment.

ALTERNATIVE HYPOTHESIS

H_1 : There is an association between early stage and adherence to treatment.

LEVEL OF SIGNIFICANCE

$\alpha = 0.05$

TEST STATISTICS

Table 4.1.1 Observed Value of Early Stage and Adherence to Treatment.

		Early Stage		
		No	**Yes**	**Total**
Adherence To Treatment	**No**	83	48	131
	Yes	85	60	145
	Total	168	108	276

Table 4.1.2 Test Statistics

Test Statistics	Values
Chi-Square Value	0.46499
df	1
P-Value	0.4953

INTERPRETATION

The P-value is 0.46499, which is greater than the 0.05 level of significance. Therefore, we accept the null hypothesis: there is no association between early stage and adherence to treatment.

Table 4.1.3 Expected Value for Early Stage and Adherence to Treatment.

		Early Stage		
		No	**Yes**	**Total**
Adherance To Treatment	**No**	70.7391	51.2609	131
	Yes	88.2609	56.7391	145
	Total	168	108	276

All expected values of early stage and adherence to treatment are greater than 5, making the chi-square test appropriate for this association checking problem.

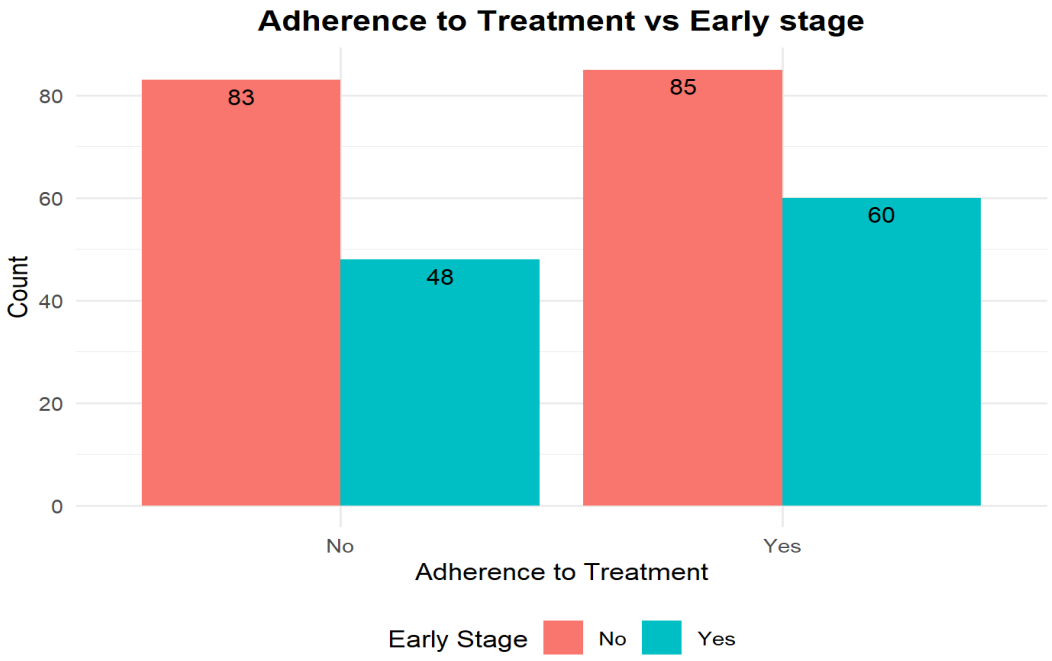

Fig 4.01 Adherence to treatment Vs early stage

The diagram above presents data on patient adherence to treatment in relation to their disease stage. It shows that 83 patients did not adhere to treatment and were not in the early stage, while 40 patients did not adhere to treatment but were in the early stage. Additionally, 85 patients adhered to treatment and were not in the early stage, and 60 patients adhered to treatment and were in the early stage.

4.2 COMPARING SEVERITY SCORE OF THE PATIENT AND THEIR CURE STATUS

NULL HYPOTHESIS:

H_0 : There is no association between severity score and cure status.

ALTERNATIVE HYPOTHESIS:

H_1 : There is an association between between severity score and cure status.

LEVEL OF SIGNIFICANCE:

$\alpha = 0.05$

TEST STATISTICS:

Table 4.2.1 Observed Value of Severity Score and Cure Status.

	Severity Score				
Cure Status		**Low**	**Medium**	**High**	**Total**
	No	15	38	21	74
	Yes	44	108	50	202
	Total	59	146	71	276

Table 4.2.2 Test statistics

Test Statistics	Values
Chi-Square Value	0.3805
df	2
P-Value	0.8268

INTERPRETATION

The P-value is 0.8268, which is greater than the 0.05 level of significance. Therefore, we accept the Null Hypothesis, which states there is no association between severity score and cure status.

Table 4.2.3 Expected Value For Severity Score and Cure Status.

		Severity Score			
		Low	**Medium**	**High**	**Total**
Cure Status	**No**	15.8188	39.1449	19.0362	74
	Yes	43.1812	106.855	51.9638	202
	Total	59	146	71	276

All expected values for severity score and cure status are greater than 5, making the chi-square test suitable for checking this association.

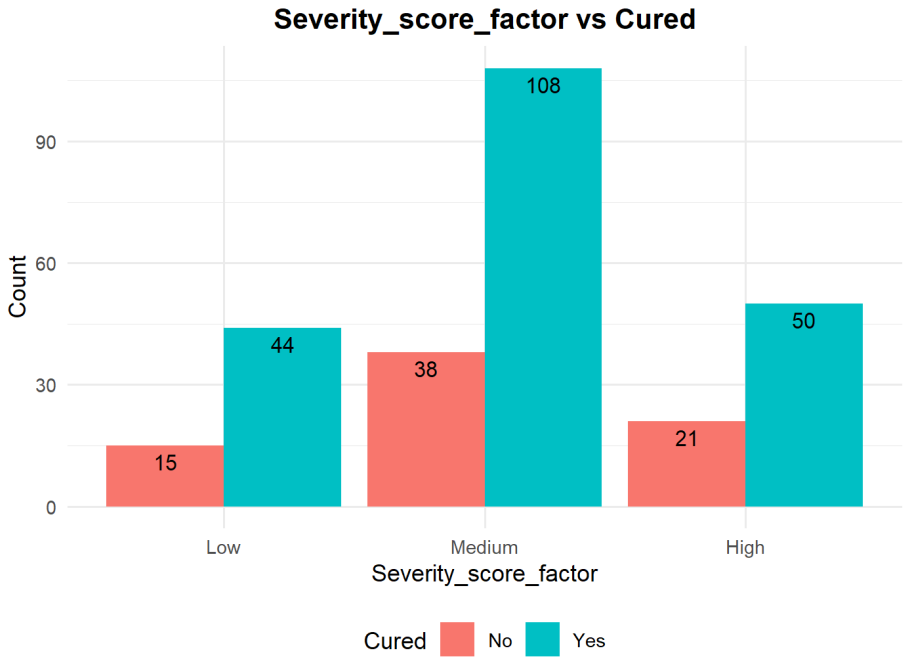

Fig 4.02 Severity score factor vs cured

In the diagram above, there are 15 patients with a low severity score who have not been cured. Additionally, 44 patients with a low severity score have been cured. There are 35 patients with a medium severity score who have not been cured, while 108 patients with a medium severity score have been cured. Furthermore, 21 patients have a high severity score and have not been cured, whereas 50 patients with a high severity score have been cured.

4.3 COMPARING TREATMENT OF THE PATIENT AND DIED WHILE TREATMENT

NULL HYPOTHESIS:

H_0 : There is no association between treatment and died while treatment.

ALTERNATIVE HYPOTHESIS:

H_1 : There is an association between treatment and died while treatment.

LEVEL OF SIGNIFICANCE:

$\alpha = 0.05$

TEST STATISTICS:

Table 4.3.1 Observed Values of Treatment and Died While Treatment

	Died While Treatment			
Treatment		No	Yes	Total
	Corticosteroids	48	8	56
	Physical Therapy	50	6	56
	Pyridostigmine + IVIG	46	5	51
	Riluzole + PT	38	7	45
	Thymectomy	58	10	68
	Total	240	36	276

Table 4.3.2 Test Statistics

Test Statistics	Values
Chi-Square Value	1.232
df	4
P-Value	0.8728

INTERPRETATION

The P-value is 0.8728, which is greater than the 0.05 level of significance. Therefore, we accept the Null Hypothesis, indicating that there is no association between treatment and mortality during treatment.

Table 4.3.3 Expected Values for Treatment and Died While Treatment

	Died While Treatment			
Treatment		**No**	**Yes**	**Total**
	Corticosteroids	48.6956	7.3044	56
	Physical Therapy	48.6957	7.3044	56
	Pyridostigmine + IVIG	44.3478	6.6522	51
	Riluzole + PT	39.1304	5.8696	45
	Thymectomy	59.1304	8.8696	68
	Total	240	36	276

Since every expected value for both treatment and death while receiving treatment exceeds 5, the chi-square test is suitable for this association-checking problem.

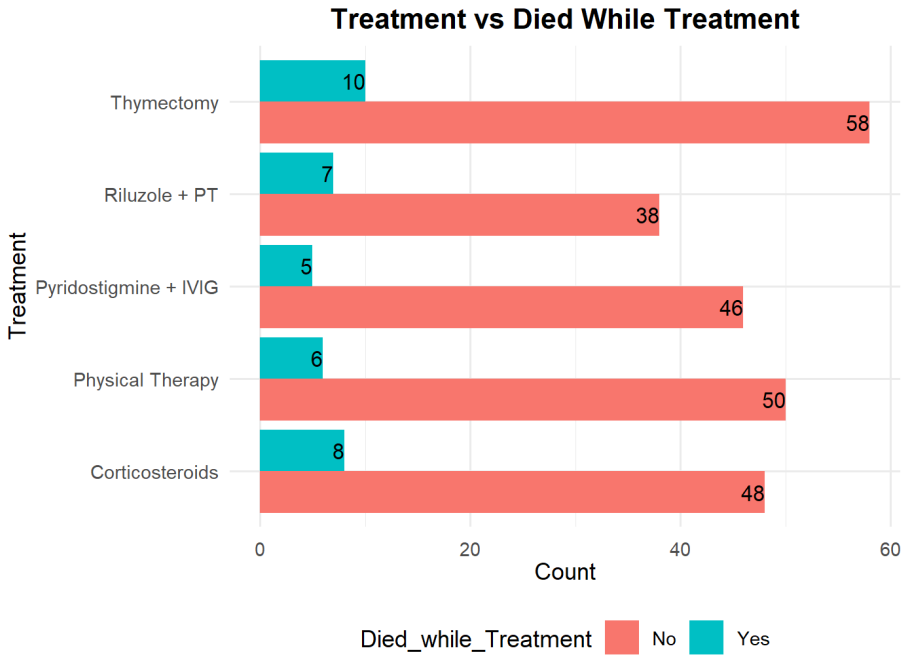

Fig 4.03 Treatment Vs Died while treatment

In the above diagram, there were 68 patients who underwent Thymectomy treatment, of which 10 patients died during the treatment while the others survived. Among the 45 patients treated with Riluzole, 7 patients died while receiving the treatment, while the remainder did not. Out of 51 patients who received a combination of Pyridostigmine and IVIG, 5 patients died during the treatment, while the others did not. For the 56 patients who underwent physical therapy, 6 patients died during treatment, while the rest survived. Lastly, out of 56 patients receiving corticosteroid treatment, 8 patients died during the treatment, while the others did not.

4.4 COMPARING MAJOR ISSUE AREA AND QUALITY OF LIFE IMPROVEMENT FOR THE PATIENT

NULL HYPOTHESIS:

H_0 : There is no association between major issue area and quality of life improvement.

ALTERNATIVE HYPOTHESIS:

H_1 : There is an association between major issue area and quality of life improvement.

LEVEL OF SIGNIFICANCE:

$\alpha = 0.05$

TEST STATISTICS:

Table 4.4.1 Observed Values Of Major Issue Area and Quality of Life Improvement

	Quality of Life Improvement					
Major Issue Area		**Mild**	**Moderate**	**None**	**Significant**	**Sum**
	Joints (Nerve and Muscle Area)	31	25	39	9	104
	Muscles	28	11	17	4	60
	Muscles and Nerves	15	5	14	6	40
	Nerves	29	14	24	5	72
	Total	103	55	94	24	276

Table 4.4.2 Test statistics

Test Statistics	Values
Chi-Square Value	8.5478
df	9
P-Value	0.48

INTERPRETATION

The P-value is 0.48, which is greater than the 0.05 level of significance. Therefore, we accept the Null Hypothesis, indicating that there is no association between the major issue area and quality of life improvement.

Table 4.4.3 Expected Values Of Major Issue Area and Quality of Life Improvement

		Quality of Life Improvement				
		Mild	**Moderate**	**None**	**Significant**	**Sum**
Major Issue Area	**Joints (Nerve and Muscle Area)**	38.8116	20.7246	35.4203	9.0435	104
	Muscles	22.3913	11.9565	20.4348	5.2174	60
	Muscles and Nerves	14.9275	7.971	13.6232	3.4783	40
	Nerves	26.8696	14.3478	24.5217	6.2609	72
	Total	103	55	94	24	276

One of the expected values for the major issue area and quality of life improvement is less than 5. Therefore, we need to compute the F-test (P value- 0.465) for further confirmation so there is no association between the major issue area and quality of life improvement.

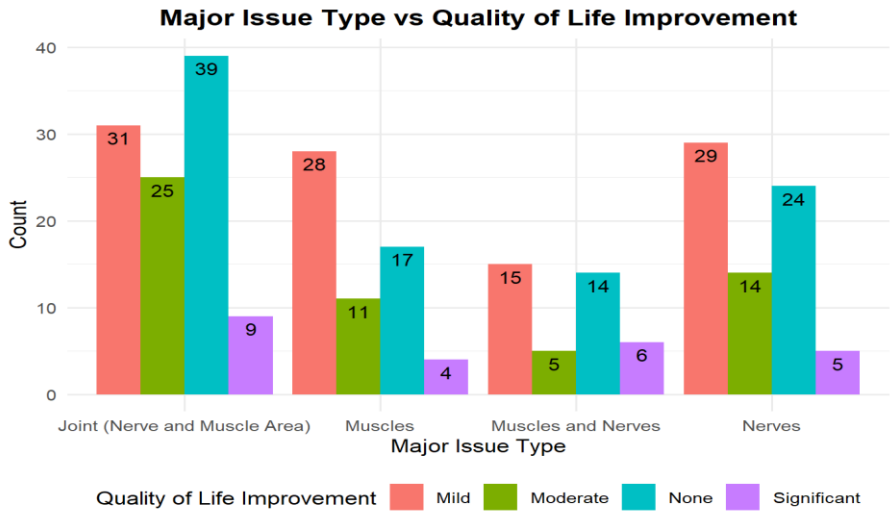

Fig 4.04 Major issue type vs quality of life improvement

The diagram summarizes findings from 104 patients with major issues in joints, muscles, and nerves. Within this group, 31 patients experienced mild improvement, 25 had moderate improvement, 39 reported no improvement, and 9 saw significant improvement. For the 60 patients specifically facing muscle issues, 28 had mild improvement, 11 showed moderate improvement, 17 reported no improvement, and 4 experienced significant improvement. Among the 40 patients with combined muscle and nerve issues, 15 had mild improvement, 5 had moderate improvement, 14 reported no improvement, and 6 experienced significant improvement. Finally, in the group of 72 patients with nerve issues, 29 reported mild improvement, 14 had moderate improvement, 24 had no improvement, and 5 had significant improvement.

4.5 COMPARING MAJOR ISSUE AREA AND AND THEIR ADHERENCE TO TREATMENT

NULL HYPOTHESIS:

H_0 : There is no association between major issue area and adherence to treatment .

ALTERNATIVE HYPOTHESIS:

H_1 : There is an association between major issue area and adherence to treatment .

LEVEL OF SIGNIFICANCE:

$\alpha = 0.05$

TEST STATISTICS:

Table 4.5.1 Observed Values Of Major Issue Area and Adherence to Treatment

	Adherence to Treatment			
Major issue Area		**No**	**Yes**	**total**
	Joint (Nerve and Muscle Area)	46	58	104
	Muscles	30	30	60
	Muscles and Nerves	23	17	40
	Nerves	32	40	72
	Total	131	145	276

Table 4.5.2 Test statistics

Test Statistics	Values
Chi-Square Value	2.4697
df	3
P-Value	0.4808

Table :10

INTERPRETATION

The P-value is 0.4808 , which is greater than the 0.05 level of significance. Therefore, we accept the Null Hypothesis, indicating that there is no association between the major issue area and adherence to treatment.

Table 4.5.3 Expected Values For Major Issue Area and Adherence to Treatment

	Adherence to Treatment			
Major issue Area		**No**	**Yes**	**total**
	Joint(Nerve and Muscle Area)	49.3623	54.6377	104
	Muscles	28.4783	31.5217	60
	Muscles and Nerves	18.9855	21.0145	40
	Nerves	34.1739	37.8261	72
	Total	131	145	276

Since every expected value for Major Issue Area and Adherence to Treatment exceeds 5, the chi-square test is suitable for this association-checking problem.

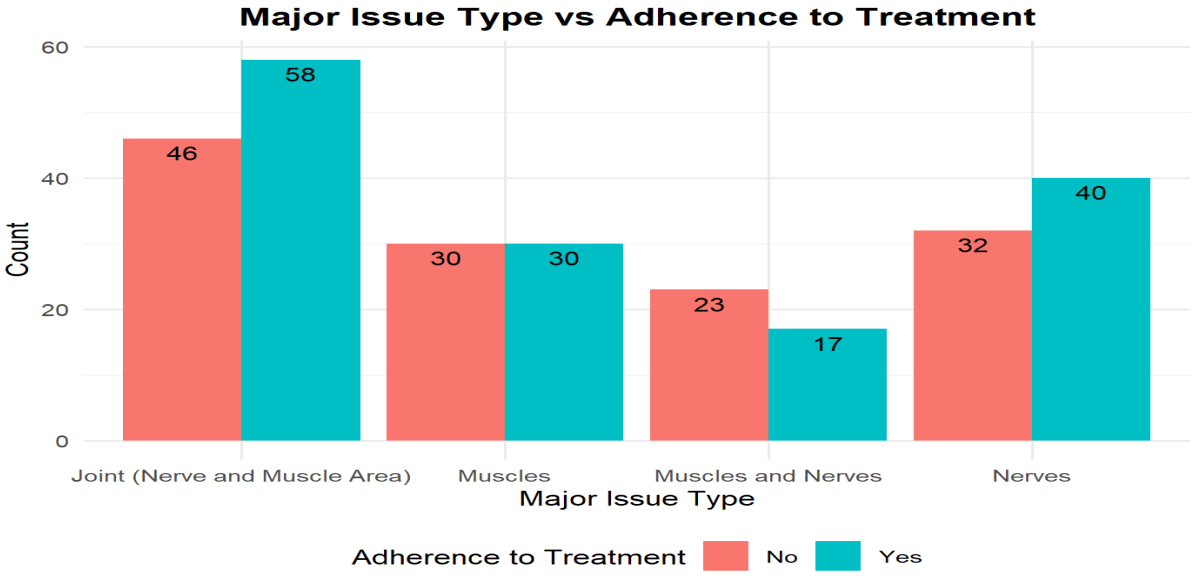

Fig 4.05 Major issue type Vs Adherence to treatment

The diagram clearly illustrates the treatment adherence among patients with major issues related to joints, nerves, and muscles. Out of 104 patients with significant joint, nerve, and muscle problems, 46 patients are not adhering to treatment, while 58 are. Among the 60 patients with major muscle issues, 30 are not adhering to treatment, whereas the remaining 30 are. For the 40 patients experiencing issues with both muscle and nerve, 23 are not adhering to treatment, and 17 are. Finally, of the 72 patients with major nerve issues, 32 are not adhering to treatment, while 40 patients are adhering. This information highlights the importance of addressing treatment adherence in managing these conditions effectively.

4.6 TO ASSESS THE EFFECTS OF MAJOR ISSUE TYPE AND SEVERITY SCORE ON MUSCLE STRENGTH

NULL HYPOTHESIS:

H_{01} :There is no significant difference between the major issue types in muscle strength.

H_{02}: There is no significant difference between the severity score levels in muscle strength.

H_{03}: There is no interaction between Major Issue Type and Severity score.

ALTERNATIVE HYPOTHESIS:

H_{11}: There is a significant difference between the major issue types in muscle strength.

H_{12}: There is a significant difference between the severity score levels in muscle strength.

H_{13}: There is an interaction between Major Issue Type and Severity score.

Before analysis, testing the normality of the muscle strength is crucial to determine whether to use a parametric or non-parametric test. If the data is normally distributed, parametric tests are preferred; otherwise, non-parametric methods are more suitable.

NULL HYPOTHESIS:

H_0: The muscle strength follows a Normal distribution

ALTERNATIVE HYPOTHESIS:

H_1: The muscle strength doesn`t follows a Normal distribution

Table 4.6.1 Normality Test On Muscle Strength

Test Statistic	Value
Shapiro Wilk value	0.9699
P-Value	1.524

The p-value of the Shapiro-Wilk test is 0.1524, which is greater than 0.05. Therefore, we fail to reject the null hypothesis, indicating that muscle strength follows a normal distribution. Hence, we proceed with a parametric test.

TABLE 4.6.2 ANOVA TABLE FOR THE EFFECT OF MAJOR ISSUE TYPE AND SEVERITY SCORE ON MUSCLE STRENGTH

Source of variation	df	Sum Sq	Mean Sq	F Value	P-Value
Major issue Type	3	2580	860	1.378	0.25
Severity Score Factor	2	497	248.3	0.398	0.672
Interaction	3	951	317	0.508	0.677
Residuals	267	166650	624.2		

Since all p-values are greater than 0.05, we fail to reject the null hypothesis for each factor. This means that neither Major Issue Type nor Severity Score has a statistically significant effect on Muscle Strength. The mean Muscle Strength is similar across different groups, indicating no meaningful variation.

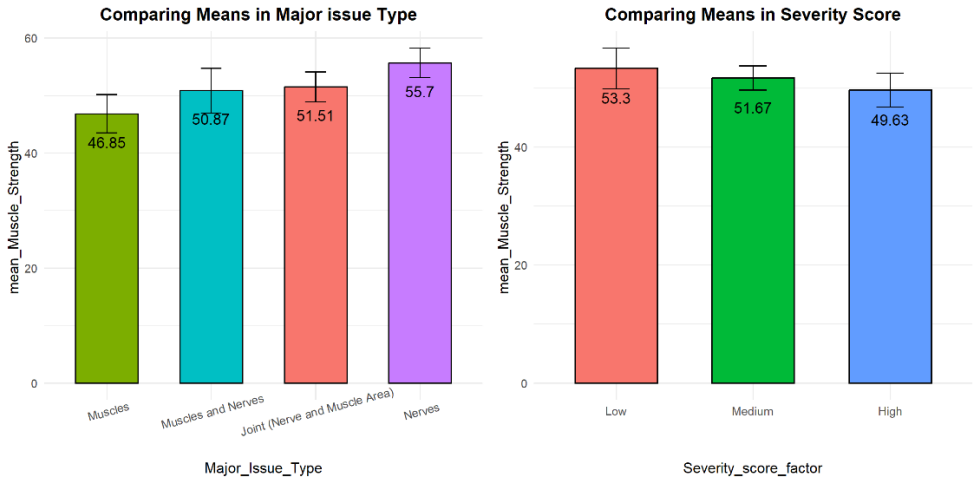

Fig 4.06 Comparing means in major issue type and severity score

From fig 4.06 the group means may differ, the differences are too small to be significant at the 5% level in the ANOVA test, suggesting that variations are within the range of random error. Further analysis, such as mean comparisons or effect sizes, might help identify impactful groups.

INTERPRETATION:

All p-values are greater than 0.05, so we fail to reject the null hypothesis, indicating that Major Issue Type, Severity Score Factor, and their interaction do not significantly affect Muscle Strength. Although differences exist—muscle issue patients tend to have lower mean strength compared to nerve issue patients, and those with lower severity scores show higher mean strength—these variations are too small to be meaningful at the 5% significance level. A deeper analysis, like effect size evaluation or confidence intervals, may offer additional insights.

4.7 TO ASSESS THE EFFECTS OF GENDER AND DISORDER ON MUSCLE STRENGTH

NULL HYPOTHESIS:

H_{01} :There is no significant difference between the gender in muscle strength.

H_{02}: There is no significant difference between the disorder in muscle strength.

H_{03}: There is no interaction between gender and disorder.

ALTERNATIVE HYPOTHESIS:

H_{11}: There is a significant difference between the gender in muscle strength.

H_{12}: There is a significant difference between the disorder in muscle strength.

H_{13}: There is an interaction between gender and disorder.

Before analysis, testing the normality of the muscle strength is crucial to determine whether to use a parametric or non-parametric test. If the data is normally distributed, parametric tests are preferred; otherwise, non-parametric methods are more suitable.

NULL HYPOTHESIS:

H_0: The muscle strength follows a Normal distribution

ALTERNATIVE HYPOTHESIS:

H_1: The muscle strength doesn`t follows a Normal distribution

Table 4.7.1 Normality Test On Muscle Strength

Test Statistic	Value
Shapiro Wilk	0.9699
P-Value	1.524

The p-value of the Shapiro-Wilk test is 0.1524, which is greater than 0.05. Therefore, we fail to reject the null hypothesis, indicating that muscle strength follows a normal distribution. Hence, we proceed with a parametric test

TABLE 4.7.2 ANOVA TABLE FOR THE EFFECT OF GENDER AND DISORDER ON MUSCLE STRENGTH

Source of variation	df	Sum Sq	Mean Sq	F Value	P-Value
Gender	1	810	809.7	1.279	0.259
Disorder	12	7134	594.5	0.939	0.508
Interaction	11	3894	354	0.559	0.861
Residuals	251	158840	632.8		

Since all p-values are greater than 0.05, we fail to reject the null hypothesis for each factor. This means that neither Gender nor Disorder has a statistically significant effect on Muscle Strength. The mean Muscle Strength is similar across different groups, indicating no meaningful variation.

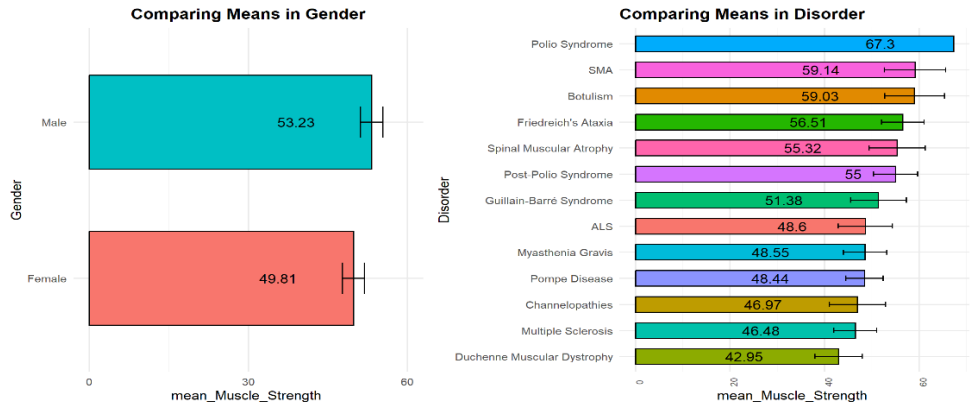

Fig 4.07 Comparing means in Gender and disorder

From fig 4.07 the group means may differ, the differences are too small to be significant at the 5% level in the ANOVA test, suggesting that variations are within the range of random error. Further analysis, such as mean comparisons or effect sizes, might help identify impactful groups.

INTERPRETATION:

All p-values are greater than 0.05, so we fail to reject the null hypothesis for both Gender and Disorder, indicating neither has a significant effect on Muscle Strength. The observed variations are likely due to random error. However, mean values suggest patterns: males generally have higher Muscle Strength than females, and patients with Duchenne Muscular Dystrophy show the lowest mean Strength among disorder groups. These differences are not statistically significant at the 5% level.

4.8 TO ASSESS THE EFFECTS OF DOSAGE AND ADHERENCE TO TREATMENT ON RATE OF PROGRESSION

NULL HYPOTHESIS:

H_{01} :There is no significant difference between the dosage in rate of progression.

H_{02}: There is no significant difference between the adherence to treatment in rate of progression.

H_{03}: There is no interaction between adherence to treatment and dosage.

ALTERNATIVE HYPOTHESIS:

H_{11}: There is a significant difference between the dosage in rate of progression.

H_{12}: There is a significant difference between the adherence to treatment in rate of progression.

H_{13}: There is an interaction between adherence to treatment and dosage.

Before analysis, testing the normality of the rate of progression is crucial to determine whether to use a parametric or non-parametric test. If the data is normally distributed, parametric tests are preferred; otherwise, non-parametric methods are more suitable.

NULL HYPOTHESIS:

H_0: The rate of progression follows a Normal distribution

ALTERNATIVE HYPOTHESIS:

H_1: The rate of progression doesn`t follows a Normal distribution

Table 4.8.1 Normality Test On Rate of Progression

Test Statistic	Value
Shapiro Wilk	0.9906
P-Value	0.07291

The p-value of the Shapiro-Wilk test is 0.1524, which is greater than 0.05. Therefore, we fail to reject the null hypothesis, indicating that rate of progression follows a normal distribution. Hence, we proceed with a parametric test.

TABLE 4.8.2 ANOVA TABLE FOR THE EFFECT OF DOSAGE AND ADHERENCE TO TREATMENT ON RATE OF PROGRESSION

Source of variation	df	Sum Sq	Mean Sq	F Value	P-Value
Dosage	2	0.05	0.0273	0.032	0.969
Adherence to Treatment	1	0.64	0.6359	0.74	0.39
Interaction	2	1.92	0.962	1.12	0.328
Residuals	270	231.89	0.8589		

Since all p-values are greater than 0.05, we fail to reject the null hypothesis for each factor. This means that neither dosage nor adherence to treatment has a statistically significant effect on rate of progression. The mean rate of progression is similar across different groups, indicating no meaningful variation.

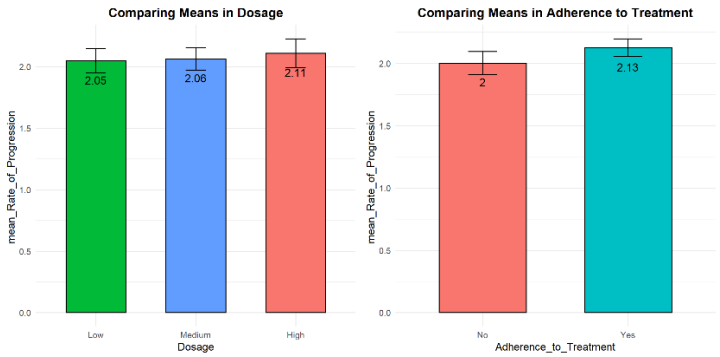

Fig 4.08 Comparing means in Dosage and Adherence to treatment

From fig 4.08 the group means may differ, the differences are too small to be significant at the 5% level in the ANOVA test, suggesting that variations are within the range of random error. Further analysis, such as mean comparisons or effect sizes, might help identify impactful groups.

INTERPRETATION:

All p-values are greater than 0.05, which means we fail to reject the null hypothesis regarding both dosage and adherence to treatment. This indicates that neither factor has a significant effect on the rate of progression, with the observed variations likely resulting from random error. However, the mean values suggest some patterns: patients who adhere to treatment generally exhibit a higher mean rate of progression compared to those who do not. Similarly, patients on a high dosage tend to have a higher rate of progression than those on a low dosage. It's important to note that these differences are not statistically significant at the 5% level.

4.9 CORRELATION PLOT

A correlation plot helps to determine the relationship between two continuous variables by displaying their correlation values. This analysis is conducted using all numerical variables in the dataset. The diagram below illustrates the correlation values among the variables in the dataset.

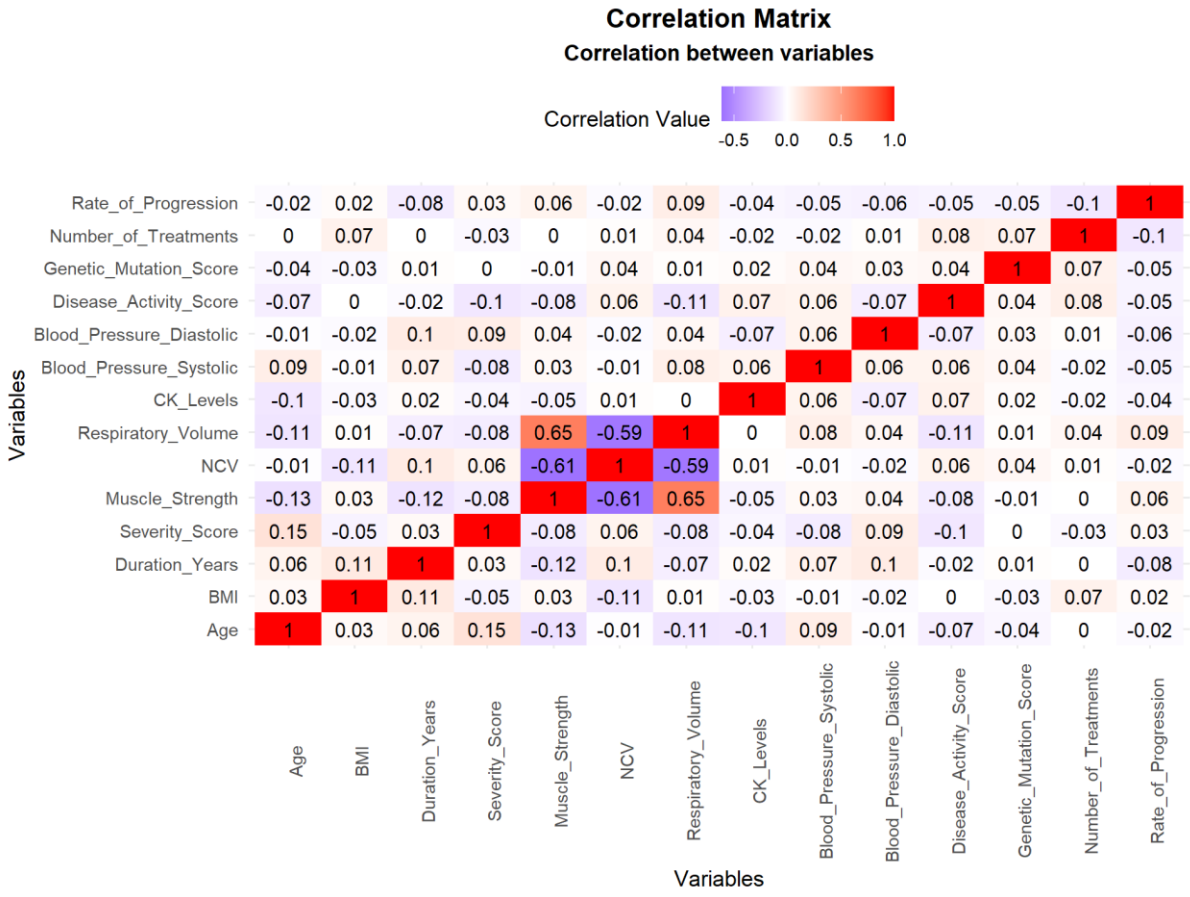

Fig 4.09 Correlation Matrix

INTERPRETATION

The correlation matrix reveals that disease severity strongly correlates with duration, respiratory function, and muscle strength, while NCV and respiratory volume show significant negative associations with key health indicators. Blood pressure variables are internally consistent but weakly related to severity. These insights highlight the critical role of respiratory and muscular factors in disease progression**.**

4.10 PREDICTING EARLY-STAGE NEUROMUSCULAR DISORDERS: A COMPARATIVE STUDY OF MACHINE LEARNING MODELS:

The prediction of whether a patient is in the early stage of a neuromuscular disorder is determined by several factors, including **Age, Gender, BMI, Duration of Disease (Years), Severity Score, Disease Activity Score, Genetic Mutation Score, Rate of Progression, Muscle Strength, Nerve Conduction Velocity (NCV), Respiratory Volume, Creatine Kinase (CK) Levels, Systolic and Diastolic Blood Pressure, Adherence to Treatment, Dosage, and Number of Treatments.**

To predict whether a patient is in the early stage of a neuromuscular disorder, 17 variables are used as independent variables, while early stage classification serves as the dependent variable in the analysis.

The dataset is split into two parts: **70% for the training set** and **30% for the testing set** to evaluate the performance of the classification models.

The training set consists of 192 samples, including 75 patients in the early stage of their condition, while the remaining patients are not in the early stage. Similarly, the testing set has 84 samples, with 33 patients in the early stage and the others not in this stage.

In this analysis, we use **Random Forest, Logistic Regression, Support Vector Machine (SVM), CatBoost, and Neural Networks** to classify whether a patient is in the early stage of a neuromuscular disorder.

4.10.1 LOGISTIC REGRESSION

Table 4.10.1.1 Training Data

Prediction	Reference		
		No	**Yes**
	No	114	6
	Yes	3	69

Table 4.10.1.2 Test statistics

Accuracy	0.9531
95% C.I	(0.9129,0.9783)
No Information Rate	0.6094
P Value[Acc > NIR]	<2e^-16
Kappa	0.9008
Mcnemar`s Test P Value	0.505

The logistic regression model shows excellent predictive performance on the training set, achieving an accuracy of **95.31%.** The **95%** confidence interval of (0.9129, 0.9783) confirms its reliability. With a Kappa score of **0.9008,** there is strong agreement between predicted and

actual classifications. The model has a sensitivity of **97.44%** and specificity of 92.00%, resulting in only 6 false negatives and 3 false positives. Its positive predictive value is 95.00%, while the negative predictive value is 95.83%, indicating highly reliable predictions. The McNemar's test P-value of **0.505** shows no significant bias in misclassification. Overall, the model performs significantly better than random guessing (P-value < 2e-16), making it a strong candidate for predicting early-stage neuromuscular disorders.

Table 4.10.1.3 Testing Data

Prediction	**Reference**		
		No	**Yes**
	No	48	7
	Yes	3	26

Table 4.10.1.4 Test statistics

Accuracy	0.881
95% C.I	(0.7919,0.9414)
No Information Rate	0.6071
P Value[Acc > NIR]	2.78e^-8
Kappa	0.745
Mcnemar`s Test P Value	0.3428

The logistic regression model achieved an accuracy of **88.1%** on the testing set, indicating strong generalizability, supported by a 95% confidence interval of **(0.7919, 0.9414).** A Kappa score of 0.745 reflects substantial agreement between predicted and actual values, while sensitivity of 94.12% shows effective identification of "No" cases, and specificity of 78.79% indicates moderate performance for "Yes" cases. The model also demonstrates high positive (87.27%) and negative predictive values (89.66%), with a balanced accuracy of 86.45%. McNemar's test yielded a P-value of **0.3428**, suggesting no significant bias in misclassification, and overall, the model significantly outperforms random guessing (P-value < 2.78e-08), proving its effectiveness in predicting early-stage neuromuscular disorders.

Table 4.10.1.5 Accuracy and Kappa value

Training Accuracy	0.9531
Testing Accuracy	0.881
Training Kappa	0.9008
Testing Kappa	0.745

The logistic regression model demonstrates strong performance, with a **training accuracy of 95.31%** and a **testing accuracy of 88.1%**, indicating effective learning and good generalization to unseen data. The **Kappa statistic**, which measures agreement beyond chance, is **0.9008 for training** and **0.745 for testing**, suggesting high reliability in classification. The

slight drop in accuracy and Kappa from training to testing implies some overfitting, but overall, the model remains robust in predicting whether a patient is in the early stage of a neuromuscular disorder.

4.10.2 RANDOM FOREST

In the Random Forest model, mtry is tuned using Grid Search to optimize accuracy, improve performance, and prevent overfitting, ensuring good generalization to unseen data.

Table 4.10.2.1 Accuracy and Kappa value

mtry	Accuracy	Kappa
2	0.9845	0.9669
12	0.9792	0.956
1	0.8854	0.7451

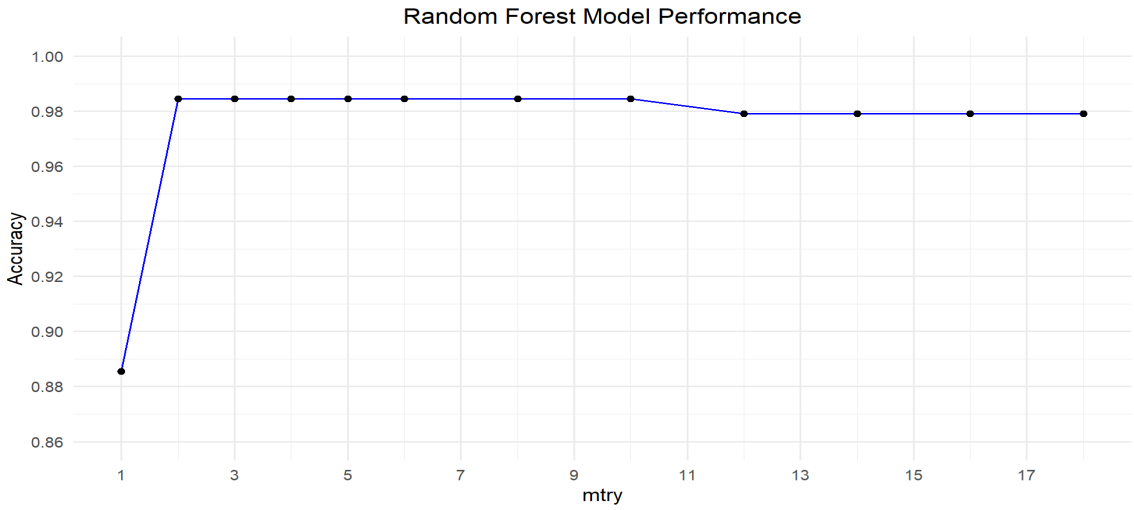

Fig 4.10.2.01 Random Forest Model performance

The table and figure clearly show that the mtry value of 2 has more accuracy in the training set. Therefore, we choose the hyperparameter value of mtry as 2 to achieve better accuracy.

Table 4.10.2.2 Random Forest Model

Random Forest Model	
Type	Classification
Number of Trees	50
No of Variables tried at each split	2
OOB estimate of Error rate	3.12%

The **Random Forest Model** is used for **classification** with **50 decision trees**. At each split, **2 variables** are randomly selected to determine the best split. The **Out-of-Bag (OOB) error rate**, which estimates the model's generalization error, is **3.12%**, indicating strong predictive performance and low error.

Table 4.10.2.3 Training data

		Reference	
		No	**Yes**
Prediction	**No**	117	0
	Yes	0	75

Table 4.10.2.4 Test statistics

Accuracy	1
95% C.I	(0.981,1)
No Information Rate	0.6094
P Value[Acc > NIR]	<2.2e^-16
Kappa	1
Mcnemar`s Test P Value	NA

The confusion matrix for the Random Forest training set demonstrates perfect classification with **100% accuracy**, correctly predicting all 117 "No" cases and 75 "Yes" cases without any misclassifications. A Kappa statistic of **1** indicates complete agreement between predicted and actual values, achieved by tuning hyperparameters, particularly the optimal mtry value of **2**. Key metrics such as Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value are all 1.0000, confirming perfect classification, while a Balanced Accuracy of 1.0000 further supports this performance. Additionally, the very small P-value (< 2.2e-16) suggests that the model's accuracy is significantly better than random guessing.

Table 4.10.2.5 Testing set

		Reference	
		No	**Yes**
Prediction	**No**	51	3
	Yes	0	30

Table 4.10.2.6 Test statistics

Accuracy	0.9643
95% C.I	(0.8992,0.9926)
No Information Rate	0.6071
P Value[Acc > NIR]	1.71e^-14
Kappa	0.9239
Mcnemar`s Test P Value	0.2482

The Random Forest model, optimized with mtry set to 2, achieved an impressive accuracy of **96.43% on the testing set**. It correctly identified 51 "No" cases and 30 "Yes" cases, with only 3 "Yes" cases misclassified as "No." The model showed perfect sensitivity (1.0000) for "No" cases and a specificity of 90.91% for "Yes" cases. The Kappa statistic of **0.9239** indicates strong agreement, while the positive predictive value of 0.9444 and negative predictive value of

1.0000 confirm the model's reliability. Overall, the Random Forest model demonstrates excellent generalization to unseen data and high predictive performance.

4.10.3 SUPPORT VECTOR MACHINE (SVM)

In SVM, hyperparameters C and sigma are optimized using grid search to improve accuracy. C balances error and simplicity, while sigma affects the RBF kernel's flexibility. This tuning enhances classification performance.

Table 4.10.3.1 Accuracy and Kappa value

Combination Number	C	Sigma	Accuracy	Kappa
32	100	0.001	0.9476	0.8884
37	1000	0.0001	0.9476	0.8884
27	10	0.01	0.9425	0.8782
38	1000	0.001	0.9425	0.8782
26	10	0.001	0.9115	0.8085

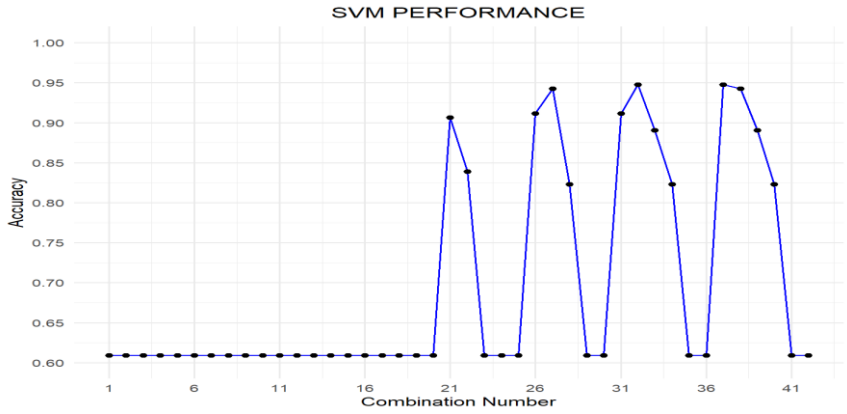

Fig 4.10.3.01 SVM performance

By analyzing the table and figure above, it is clear that hyperparameter combination number 32 yields the highest accuracy. This combination features a C value of 100 and a sigma value of 0.001, resulting in greater accuracy compared to the other combinations.

Table 4.10.3.2 SVM Model

SVM Model	
SVM- Type	C-Classification
SVM-Kernal	radial
Cost	100
Number of SupportVectors	72
Gamma	0.001

This SVM model uses a C-Classification type with an RBF kernel. The Cost (C) parameter is set to 100, focusing on minimizing classification errors, resulting in a more complex decision boundary. The Gamma value is 0.001, which helps prevent overfitting. There are 72 support vectors that define the decision boundary, and hyperparameters were tuned using grid search for optimal accuracy.

Table 4.10.3.3 Training set

Prediction	**Reference**		
		No	**Yes**
	No	117	5
	Yes	0	70

Table 4.10.3.4 Test statistics

Accuracy	0.974
95% C.I	(0.9403,0.9915)
No Information Rate	0.6094
P Value[Acc > NIR]	<2.2e^-16
Kappa	0.96446
Mcnemar`s Test P Value	0.07364

This confusion matrix shows the performance of the SVM model on the training set after optimizing the C and gamma hyperparameters using grid search. The model achieved an accuracy of 97.4%, indicating strong classification performance. It correctly identified all actual "No" cases (sensitivity of 1.0000) and most "Yes" cases (specificity of 0.9333), with only a few misclassifications. The Kappa statistic of 0.9446 indicates a high agreement between predicted and actual values. The p-value from McNemar’s test (0.07364) reveals no significant differences in misclassification rates between classes. A balanced accuracy of 0.9667 suggests the model maintains a good balance between sensitivity and specificity, demonstrating effective classification of the training data.

Table 4.10.3.5 Testing set

Prediction	Reference		
		No	Yes
	No	48	7
	Yes	3	26

Table 4.10.3.6 Test statistics

Accuracy	0.881
95% C.I	(0.7919,0.9414)
No Information Rate	0.6071
P Value[Acc > NIR]	<2.7e^-16
Kappa	0.745
Mcnemar`s Test P Value	0.3428

The confusion matrix summarizes the testing performance of the fine-tuned SVM model, which achieved an accuracy of **88.1%.** While it shows strong generalization, its sensitivity (0.9412) indicates that most "No" cases were correctly classified, but specificity (0.7879) reveals some "Yes" cases were misclassified as "No." The Kappa statistic (**0.745**) shows substantial agreement, and the McNemar's test p-value (**0.3428**) suggests no significant difference in misclassification rates. With a balanced accuracy of **0.8645**, the model performs well, maintaining high sensitivity but at the expense of some specificity, indicating potential for further tuning.

4.10.4 NEURAL NETWORK

Hyperparameters like size (hidden units) and decay (regularization) in the Neural Network model are optimized through grid search to improve classification accuracy. This tuning helps balance underfitting and overfitting, enhancing generalization and overall performance.

Table 4.10.4.1 Accuracy and Kappa value

Combination Number	Size	Decay	Accuracy	Kappa
5	5	0.1	0.9318	0.8536
10	15	0.1	0.9159	0.8236
15	25	0.1	0.8975	0.7806
20	35	0.1	0.8646	0.7043
25	45	0.1	0.8544	0.6965

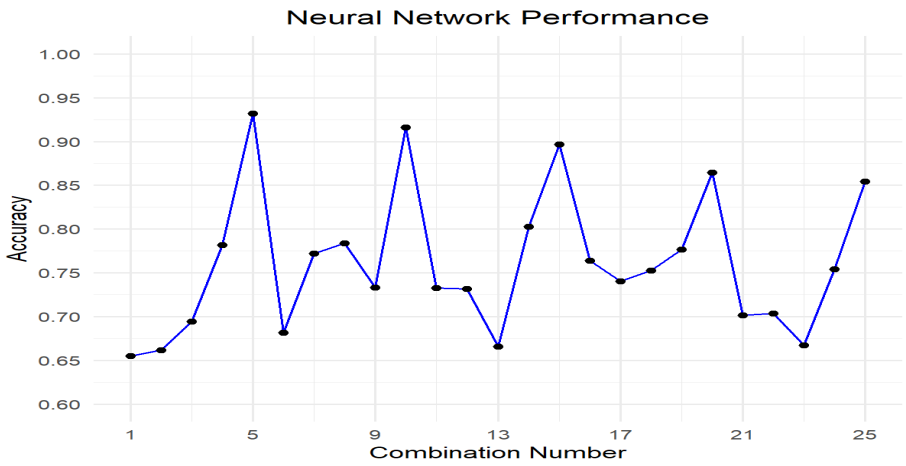

Fig 4.10.4.01 Neural Network Performance

By analyzing the table and figure above, it is clear that hyperparameter combination number 5 yields the highest accuracy. This combination features a Size value of 5 and a Decay value of 0.1, resulting in greater accuracy compared to the other combinations.

The 18-5-1 neural network is designed with 18 input neurons, 5 hidden neurons, and 1 output neuron that predicts the Early Stage. The model contains a total of 101 weights, which includes biases. It was trained using entropy fitting to optimize classification performance. A decay factor of 0.1 was applied to help prevent overfitting by regularizing the weight values.

Table 4.10.4.2 Training set

Prediction	**Reference**		
		No	**Yes**
	No	117	2
	Yes	0	73

Table 4.10.4.3 Test statistics

Accuracy	0.9896
95% C.I	(0.9629,0.9987)
No Information Rate	0.6094
P Value[Acc > NIR]	<2.2e^-16
Kappa	0.978
Mcnemar`s Test P Value	0.4795

The confusion matrix for the neural network on the training dataset indicates exceptional classification performance. It accurately classified 117 instances of "No" and 73 of "Yes," with only two misclassifications (predicting "No" instead of "Yes") and zero false positives. The model achieved an overall accuracy of **98.96%**, supported by a Kappa statistic of **0.978**, indicating nearly perfect agreement with actual values. It also boasts 100% sensitivity (recall) for "No" cases and **97.33%** specificity for differentiating between classes, resulting in a

balanced accuracy of 98.67%. These findings demonstrate that the neural network effectively learned from the training data, establishing it as a highly reliable classifier.

Table 4.10.4.4 Testing set

Prediction	Reference		
		No	**Yes**
	No	49	5
	Yes	2	28

Table 4.10.4.5 Test statistics

Accuracy	0.9167
95% C.I	(0.8358,0.9658)
No Information Rate	0.6071
P Value[Acc > NIR]	1.56e^-12
Kappa	0.8225
Mcnemar`s Test P Value	0.4497

The confusion matrix for the neural network model on the testing dataset shows strong classification performance, achieving an **accuracy of 91.67%** with a **95% confidence interval of (0.8358, 0.9658)**. The **Kappa statistic of 0.8225** indicates a high level of agreement between predicted and actual values. The model exhibits **high sensitivity (96.08%)**, meaning it correctly identifies most "No" cases, and **specificity (84.85%)**, ensuring good classification of "Yes" cases. Additionally, the **positive predictive value (90.74%)** and **negative predictive value (93.33%)** confirm the reliability of the model's predictions. The **balanced accuracy of 90.46%** highlights its effectiveness across both classes without bias, and the **p-value (<0.0000000001557)** indicates that the model significantly outperforms random guessing. Overall, the neural network model demonstrates strong generalization to unseen data, making it a robust classifier.

4.10.5 CATEGORICAL BOOSTING

In CatBoost, categorical variables like **"Gender," "Adherence_to_Treatment," and "Dosage"** are handled natively without one-hot encoding. Instead, CatBoost uses **Ordered Target Statistics and CatBoost Encoding**, which efficiently capture category relationships while preventing data leakage. By specifying these variables as categorical, the model improves predictive performance, preserves feature interactions, and reduces memory usage, making it ideal for handling high-cardinality categorical data.

Table 4.10.5.1 CATBOOST Model

CatBoost model	
Trees	1000
Loss Function	Logloss

The **CatBoost model** is trained with **1,000 trees** using the **Logloss** function, which is commonly used for binary classification. Logloss helps minimize prediction errors by penalizing incorrect high-confidence predictions, improving model accuracy.

Table 4.10.5.2 Training set

		Reference	
		No	**Yes**
Prediction	**No**	117	0
	Yes	0	75

Table 4.10.5.3 Test statistics

Accuracy	1
95% C.I	(0.981,1)
No Information Rate	0.6094
P Value[Acc > NIR]	<2.2e^-16
Kappa	1
Mcnemar`s Test P Value	NA

The Confusion Matrix for the CatBoost training set shows that the model achieved perfect classification performance with an accuracy of 100%. This means all predictions were correct, resulting in no false positives or false negatives. The Kappa statistic is 1, indicating a perfect agreement between predicted and actual values. Both sensitivity and specificity are 1.0000, which means the model correctly identified all positive and negative cases.

The P-value (< 0.22 × 10^-15) suggests that the model's accuracy is significantly better than random chance. However, the perfect accuracy on the training set may indicate overfitting, so it is important to carefully analyze the model's performance on the test set.

Table 4.10.5.4 Testing set

		Reference	
		No	**Yes**
Prediction	**No**	51	2
	Yes	0	31

Table 4.10.5.5 Test statistics

Accuracy	0.9762
95% C.I	(0.9166,0.9971)
No Information Rate	0.6071
P Value[Acc > NIR]	<9.5e^-17
Kappa	0.9495
Mcnemar`s Test P Value	0.4795

The Confusion Matrix for the CatBoost test set shows a highly accurate classification with an accuracy of 97.62% and a 95% confidence interval (91.66% – 99.71%), indicating strong model performance. The Kappa statistic (0.9495) suggests a very high agreement between predicted and actual labels. Sensitivity is 1.0000, meaning the model correctly identified all positive cases, while specificity is 0.9394, showing a strong ability to identify negative cases

as well. The P-value (< 0.000000000000000948) confirms that the model's performance is significantly better than random classification. While the model generalizes well, its near-perfect sensitivity suggests it might be slightly overfitted, so further validation may be useful.

4.10.6 COMPARISON OF MODELS

When comparing the analyses above, select the one that demonstrates the highest accuracy performance.

Table 4.10.6 Performance among the models

Model	Training Accuracy	Training Kappa	Testing Accuracy	Testing kappa
Logistic Regression	0.9531	0.9008	0.881	0.745
Random Forest	1	1	0.9643	0.9239
Support Vector Machine(SVM)	0.974	0.9645	0.881	0.745
Neural Network	0.9896	0.978	0.9167	0.8225
Catboost	1	1	0.9762	0.9495

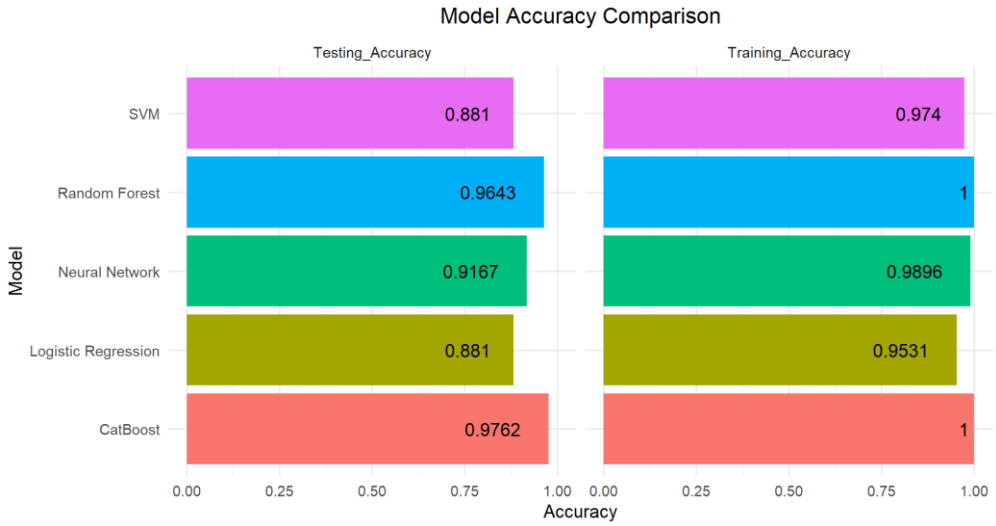

Fig 4.10.6.01 Comparison of Model Accuracy

The table and figure clearly shows that Catboost has higher accuracy on both the training and testing sets, so we conclude that Catboost is the best model among those evaluated.

4.10.7 Computing AUC and ROC for Catboost classification

- **Training set AUC = 1**
- **Testing Set AUC = 1**

An **AUC (Area Under the Curve)** of 1 for both training and test sets indicates perfect discrimination between classes, meaning the model correctly classifies every instance with 100% sensitivity and specificity. This ideal performance places the **ROC (Receiver Operating Characteristic) curve in the top-left corner**. While this level of accuracy is valuable in critical classification applications, it's essential to analyze the dataset further to understand the factors behind this perfect separation and ensure consistent predictive capability across different datasets.

INTERPRETATION

The results show that CatBoost outperforms all models, achieving perfect accuracy on both training and testing sets with an AUC of 1, indicating it can fully discriminate between positive and negative cases. This remarkable performance is crucial for early detection of conditions, allowing healthcare professionals to implement timely interventions and optimize treatment plans, ultimately improving patient outcomes. However, further analysis is needed to verify its robustness across different datasets and ensure its effectiveness in real-world scenarios.

4.11 PREDICTING MAJOR AFFECTED AREAS IN NEUROMUSCULAR PATIENTS: A MACHINE LEARNING APPROACH

To identify major concerns in neuromuscular patients—whether they stem from nerves, muscles, joints, or a combination—we employ various machine learning algorithms. These analyze independent variables such as Severity Score, Rate of Progression, Muscle Strength, Nerve Conduction Velocity (NCV), Respiratory Volume, CK Levels, Age, BMI, Disease Duration, Number of Treatments, Blood Pressure, Disease Activity Score, Genetic Mutation Score, and treatment adherence. This approach enables us to pinpoint the primary affected area, leading to personalized treatment strategies that enhance effectiveness and improve patient quality of life.

The dataset includes a mix of unordered categorical variables (e.g., Treatment, Disorder) and ordered categorical variables (e.g., Adherence to Treatment). To streamline analysis, we use **Factor Analysis for Mixed Data** (FAMD), which effectively manages both numerical variables (like Duration_Years and Blood Pressure) and categorical ones (like Gender and Disorder). By selecting the most informative components, FAMD boosts the predictive model's efficiency and interpretability, facilitating targeted interventions for neuromuscular patients.

Before applying dimensional reduction, there are some tests that need to be conducted.

Table 4.11.1 Kaiser- Meyer-Olkin Test

Kaiser-Meyer-Olkin	
Overall MSA	0.5

Table 4.11.2 Barlett Test

Bartlett Test of Homogeneity of Variances	
Barlett`s K Squared	20513
df	18
Pvalue	<2.2e-17

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (0.5) indicates that the dataset is at the threshold for factor analysis, suggesting it can be performed but may benefit from refining variable selection. The Bartlett's Test of Sphericity ($p < 2.2e\text{-}16$) confirms significant correlations among variables, making factor analysis appropriate. Given the mix of numerical and categorical variables, Factor Analysis for Mixed Data (FAMD) is a suitable dimensionality reduction technique to preserve variable relationships while reducing dimensionality effectively.

Table 4.11.3 Eigen values and Cumulative variance

Component	Eigen Value	Explained Variance	Cumulative Variance
1	2.8287	28.319	28.319
2	2.2972	22.997	51.316
3	1.7793	17.813	69.129
4	1.5796	15.813	84.942
5	1.5041	15.058	100

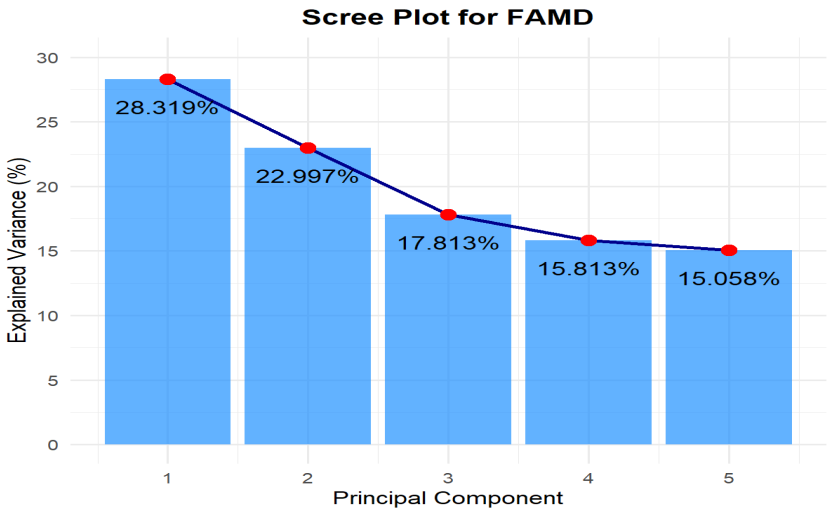

Fig 4.11.01 Scree plot for FAMD

With the help of above table and figure , The FAMD variance table shows that the first two components capture 51.32% of the dataset's variance, while the first four together explain 84.94%, indicating that most of the data structure is retained. With all five components accounting for 100% of the variance, FAMD effectively reduces dimensionality while preserving key information, making it suitable for further analysis.

In **Factor Analysis for Mixed Data (FAMD)**, the **distance matrix** represents the dissimilarities between observations based on their principal component scores. Since FAMD handles both numerical and categorical variables, it computes distances in a way that balances their contributions.

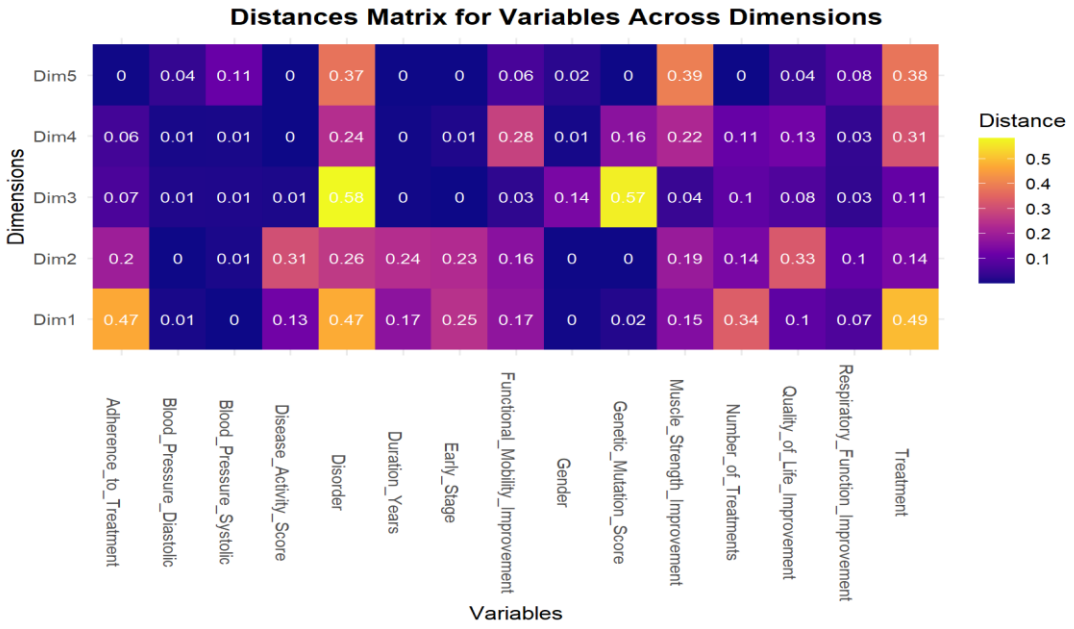

Fig 4.11.02 Distance Matrix

The FAMD distance matrix highlights the contribution of variables across five dimensions. Disorder, Genetic_Mutation_Score, and Treatment show the highest influence, while Early_Stage and Quality_of_Life_Improvement contribute moderately. Gender and Blood Pressure variables have minimal impact. This helps in identifying key variables for dimensionality reduction while preserving important information for classification and prediction.

Now, The dataset is split into two parts: **70% for the training set** and **30% for the testing set** to evaluate the performance of the classification models.

The training set includes 192 samples: 72 patients with major joint issues, 42 with major muscle issues, 40 with issues in both muscle and nerves, and 50 with major nerve issues. Additionally, a second training set consists of 84 samples: 32 patients with major joint issues, 18 with major muscle issues, 12 with issues in both muscle and nerves, and 22 with major nerve issues.

In this analysis, we use **Random Forest, Generalised Linear model with Regulation , XG boost ,Support Vector Machine (SVM),Decision Tree** to classify whether a patient is in the early stage of a neuromuscular disorder.

4.12 RANDOM FOREST

In the Random Forest model, mtry is tuned using Grid Search to optimize accuracy, improve performance, and prevent overfitting, ensuring good generalization to unseen data.

Table 4.12.1 Accuracy Kappa value

mtry	Accuracy	Kappa
4	0.9637	0.949
5	0.9634	0.9487
11	0.9634	0.9487
6	0.9583	0.9415
21	0.9583	0.9415

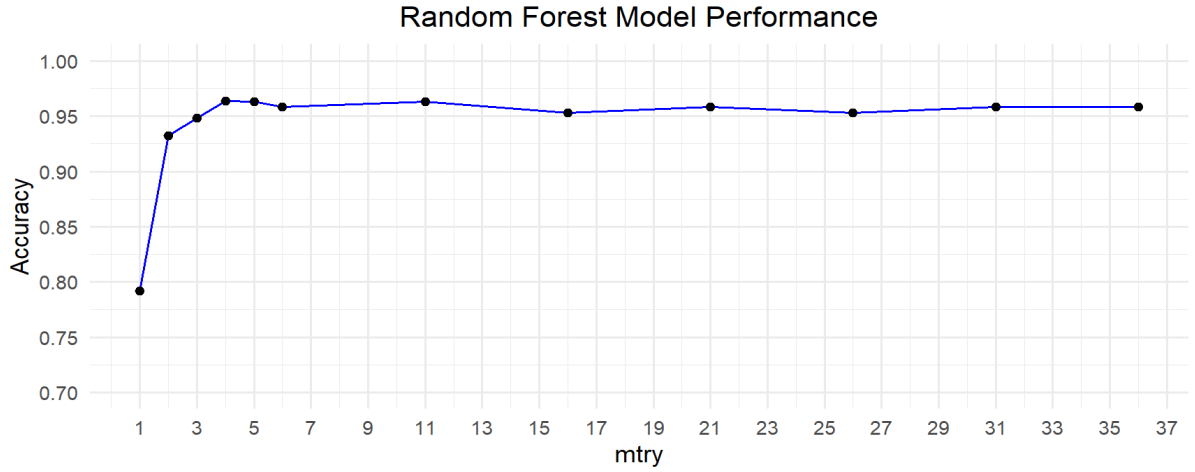

Fig 4.12.01 Random Forest model performance

The table and figure clearly show that the mtry value of 4 has more accuracy in the training set. Therefore, we choose the hyperparameter value of mtry as 4 to achieve better accuracy

Table 4.12.2 Random Forest Model

Random Forest Model	
Type	Classification
Number of Trees	50
No of the Variables tried at each split	4
OOB estimate of Error rate	7.29%

The Random Forest model, with 50 trees and 4 variables per split, achieves an OOB error rate of 7.29%, indicating good predictive performance. Feature importance can also be analyzed.

Table 4.12.3 Training set

		Reference			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	72	0	0	0
	Muscles	0	42	0	0
	Both	0	0	28	0
	Nerves	0	0	0	50

Table 4.12.4 Test statistics

Accuracy	1
95% C.I	(0.981,1)
No Information Rate	0.375
P Value [Acc > NIR]	<2.2e^-16
Kappa	1
Mcnemar`s Test P Value	NA

The confusion matrix for the training set of the Random Forest model shows perfect classification across all four categories: Joint (Nerve and Muscle Area), Muscles, Muscles and Nerves, and Nerves, with no misclassifications. The overall **accuracy is 100%**, with a Kappa value of 1, indicating perfect agreement. All performance metrics—sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)—are 1.000, confirming flawless classification. This suggests the model has learned the training data extremely well, but further evaluation on a test set is needed to check for overfitting.

Table 4.12.5 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	32	1	4	1
	Muscles	0	17	0	0
	Both	0	0	8	0
	Nerves	0	0	0	21

Table 4.12.6 Test statistics

Accuracy	0.9286
95% C.I	(0.851,0.9733)
No Information Rate	0.381
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.381
Mcnemar`s Test P Value	NA

The confusion matrix for the testing set of the Random Forest model shows an overall accuracy of 92.86%, indicating strong generalization. The Kappa value (0.8987) suggests excellent agreement beyond chance. Sensitivity and specificity are high across all classes, though "Muscles and Nerves" has slightly lower sensitivity (0.667), meaning some instances were misclassified. Despite a few misclassifications, the model maintains high positive predictive values (PPV) and negative predictive values (NPV), indicating reliable predictions. This

performance suggests the model generalizes well, with minor room for improvement in distinguishing "Muscles and Nerves."

Generalized Linear Model with Regularization

In the Generalized Linear Model with Regularization model , The alpha is tuned using Grid Search to optimize accuracy, improve performance, and prevent overfitting, ensuring good generalization to unseen data.

Table 4.12.7 Train Accuracy and Test Accuracy

Alpha	Train Accuracy	Test Accuracy
0.8	0.9531	0.8691
1	0.9531	0.8691
0.9	0.9531	0.8571
0.1	0.9531	0.8333
0.2	0.9531	0.8333

Choosing an alpha value of 0.8 significantly increases the accuracy for both training and testing data.

After finding the best alpha value for the model, there is a need to choose the best lambda (minimum value) for the lambda parameter.

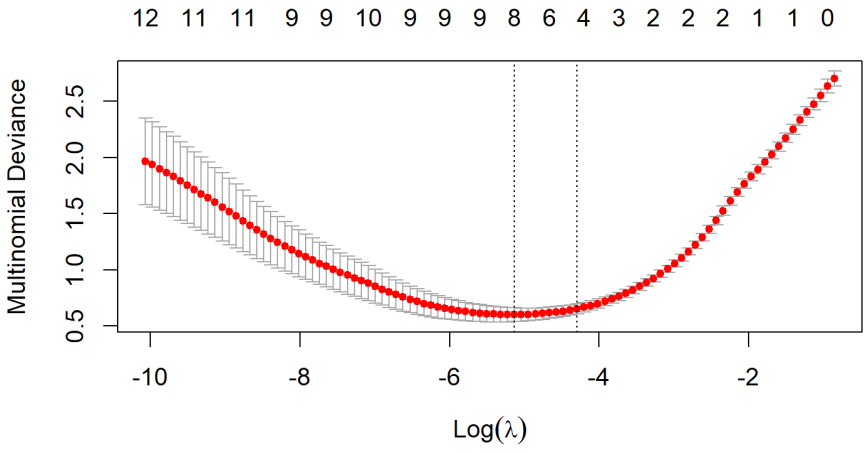

Fig 4.12.02

The diagram clearly shows that the best minimum lambda for alpha is 0.8: 0.005904

Model Summary

Table 4.12.8 GLMNET model

GLMNET Model	
Family	Multinomial
alpha	0.8
lambda	0.005904
df	14
%Dev	85.2

The GLMNET model is a multinomial logistic regression with Elastic Net regularization, using an alpha of 0.8, which leans towards Lasso for feature selection. The optimal lambda is 0.005904, with 14 degrees of freedom, indicating key contributing predictors. The model explains 85.2% of the deviance, showing strong predictive performance while balancing complexity and regularization.

Table 4.12.9 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	69	1	5	0
	Muscles	0	41	1	0
	Both	3	0	22	0
	Nerves	0	0	0	50

Table 4.12.10 Test statistics

Accuracy	0.9479
95% C.I	(0.9063,0.9747)
No Information Rate	0.375
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.9275
Mcnemar`s Test P Value	NA

The confusion matrix for the training set of the GLMNET model shows high classification accuracy (94.79%) across four classes: Joint (Nerve and Muscle Area), Muscles, Muscles and Nerves, and Nerves. The model performs exceptionally well for the Nerves class, achieving 100% sensitivity and specificity. The other classes also show strong performance, with sensitivity ranging from 78.57% to 97.62% and specificity above 95% for all. The overall Kappa value of 0.9275 indicates excellent agreement beyond chance. These results suggest that the model is highly reliable in classifying neuromuscular conditions.

Table 4.12.11 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	26	0	4	2
	Muscles	0	18	0	0
	Both	5	0	8	0
	Nerves	1	0	0	20

Table 4.12.12 Test statistics

Accuracy	0.8571
95% C.I	(0.7638,0.9239)
No Information Rate	0.381
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.802
Mcnemar`s Test P Value	NA

The GLMNET model on the testing set achieved an overall accuracy of 85.71% with a Kappa value of 0.802, indicating strong agreement between predictions and actual classifications. The model performed exceptionally well for the Muscles class, achieving 100% sensitivity and specificity, meaning all Muscles cases were correctly classified. The Nerves class also showed strong performance with 90.91% sensitivity and 98.39% specificity, ensuring minimal misclassifications. The Joint (Nerve and Muscle Area) class had 81.25% sensitivity and 88.46% specificity, indicating some misclassifications, primarily as Muscles and Muscles and Nerves. However, the Muscles and Nerves class had the weakest performance, with 66.67% sensitivity and 93.06% specificity, suggesting challenges in distinguishing it accurately. Overall, the model generalizes well to unseen data but could be improved in classifying Muscles and Nerves more accurately.

4.13 Extreme Gradient Boosting (XGBoost)

In XGBoost, the hyperparameters are max_depth, eta, and nround. These parameters are tuned using the grid search method to obtain the best model for the training set.

Table 4.13.1 XG Boost Model

XG Boost Model	
Objective	Multi:softmax
Number of Class	4
Max Depth	3
Eta	0.01
nround	50
Eval Metric	mlogloss

The XGBoost model is configured for multi-class classification using the softmax objective function, handling four classes. The model has a maximum depth of 3, which controls tree

complexity, and a learning rate (eta) of 0.01, ensuring gradual updates for better convergence. The training process runs for 50 rounds to optimize performance, using multi-class log loss (mlogloss) as the evaluation metric, which helps measure the model's predictive accuracy and confidence.

Table 4.13.2 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	72	0	5	0
	Muscles	0	42	0	0
	Both	0	0	23	0
	Nerves	0	0	0	50

Table 4.13.3 Test statistics

Accuracy	0.974
95% C.I	(0.9403,0.9915)
No Information Rate	0.375
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.9637
Mcnemar`s Test P Value	NA

The XGBoost model's training set achieves 97.4% accuracy with a Kappa of 0.9637, indicating excellent classification performance. Sensitivity and specificity are 100% for "Muscles" and "Nerves", while "Joint (Nerve and Muscle Area)" and "Muscles and Nerves" also show strong predictive reliability. The balanced accuracy exceeds 90% for all classes, ensuring minimal misclassification. Overall, the model performs exceptionally well, demonstrating near-perfect classification with high precision and recall.

Table 4.13.4 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	32	1	4	1
	Muscles	0	17	0	0
	Both	0	0	8	0
	Nerves	0	0	0	21

Table 4.13.5 Test statistics

Accuracy	0.9286
95% C.I	(0.8510,0.9733)
No Information Rate	0.381
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.8987
Mcnemar`s Test P Value	NA

The XGBoost model's testing set achieves 92.86% accuracy with a Kappa of 0.8987, indicating strong agreement between predictions and actual values. The model performs exceptionally well for "Joint (Nerve and Muscle Area)" and "Nerves", with 100% and 95.45% sensitivity, respectively. "Muscles" also shows high sensitivity (94.44%), while "Muscles and Nerves" has lower sensitivity (66.67%), indicating some misclassification. However, specificity is 100% for all classes except "Joint (Nerve and Muscle Area)" (88.46%), ensuring minimal false

positives. Overall, the model generalizes well with high precision, recall, and balanced accuracy across most classes.

4.14 SUPPORT VECTOR MACHINE (SVM)

In SVM, hyperparameters C and sigma are optimized using grid search to improve accuracy. C balances error and simplicity, while sigma affects the RBF kernel's flexibility. This tuning enhances classification performance.

Table 4.14.1 Accuracy and Kappa value

Combination Number	C	Sigma	Accuracy	Kappa
32	100	0.001	0.8450	0.7810
37	1000	0.0001	0.8448	0.7806
27	10	0.01	0.8399	0.7734
39	1000	0.01	0.8291	0.7611
33	100	0.01	0.8187	0.7459

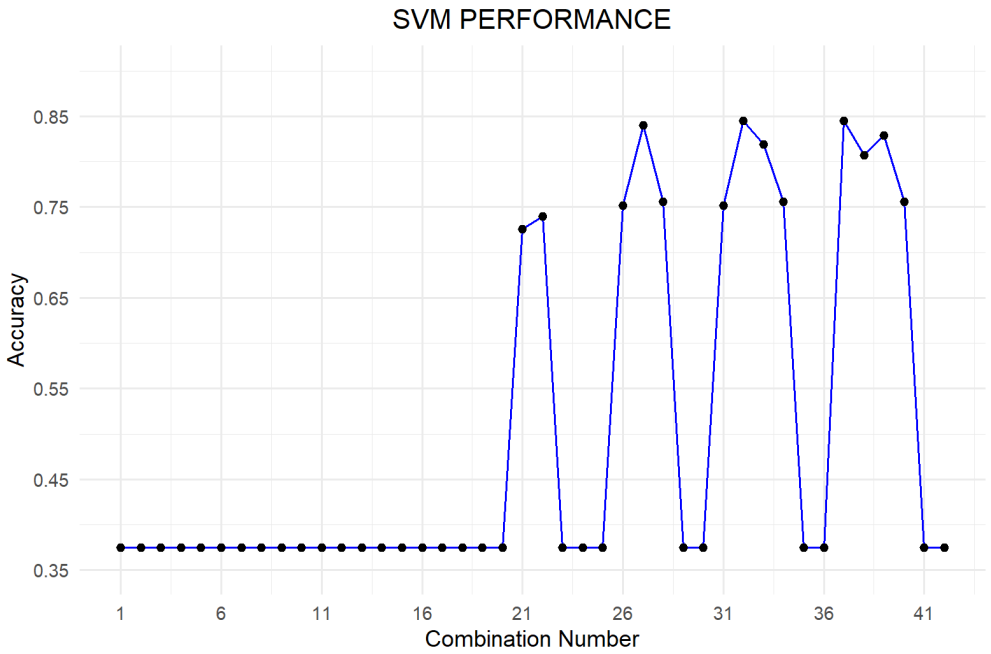

Fig 4.14.01 SVM Performance

By analyzing the table and figure above, it is clear that hyperparameter combination number 32 yields the highest accuracy. This combination features a C value of 100 and a sigma value of 0.001, resulting in greater accuracy compared to the other combinations.

Table 4.14.2 SVM Model

SVM Model	
SVM- Type	C-Classification
SVM-Kernal	radial
Cost	1000
Number of SupportVectors	72
Gamma	0.001

The SVM model is trained using a radial kernel with a cost parameter of 1000 and gamma set to 0.001 for classifying the Major Issue Type into four categories: Joint (Nerve and Muscle Area), Muscles, Muscles and Nerves, and Nerves. The model utilizes 78 support vectors, distributed across the classes as 38, 11, 17, and 12, respectively. This setup suggests a complex decision boundary to separate the classes effectively while maintaining a balance between margin maximization and misclassification.

Table 4.14.3 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	69	0	6	0
	Muscles	1	42	0	0
	Both	2	0	22	0
	Nerves	0	0	0	50

Table 4.13.4 Test statistics

Accuracy	0.9531
95% C.I	(0.9129,0.9783)
No Information Rate	0.375
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.9348
Mcnemar`s Test P Value	NA

The SVM model on the training set achieved a high accuracy of 95.31%, with a strong Kappa value of 0.9348, indicating excellent agreement between predictions and actual labels. It performed exceptionally well for the Muscles and Nerves class, with 100% sensitivity and specificity, ensuring perfect classification. The Muscles class also had 100% sensitivity and 99.33% specificity, making it highly reliable. Joint (Nerve and Muscle Area) had 95.83% sensitivity and 95.00% specificity, with some minor misclassification. However, the Muscles and Nerves class showed slightly lower sensitivity at 78.57%, leading to occasional misclassification. Despite this, the model demonstrates strong predictive power, effectively distinguishing between classes with minimal errors.

Table 4.14.5 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	29	1	3	4
	Muscles	1	17	1	0
	Both	1	0	8	0
	Nerves	1	0	0	18

Table 4.14.6 Test statistics

Accuracy	0.8571
95% C.I	(0.7638,0.9239)
No Information Rate	0.381
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.7986
Mcnemar`s Test P Value	NA

The SVM model on the testing set achieved an accuracy of 85.71%, with a Kappa value of 0.7986, indicating substantial agreement. The Joint (Nerve and Muscle Area) class had a 90.62% sensitivity but a slightly lower 78.38% positive predictive value, suggesting some misclassification. The Muscles class performed well, with 94.44% sensitivity and 96.97% specificity, ensuring strong classification. However, the Muscles and Nerves class had a lower 66.67% sensitivity, indicating challenges in distinguishing it accurately. The Nerves class showed 81.82% sensitivity and 98.39% specificity, meaning some instances were misclassified. Despite these minor misclassifications, the model maintains a strong predictive performance, making it a reliable classifier for distinguishing neuromuscular conditions.

4.15 DECISION TREE

In a Decision Tree, the complexity parameter (cp) is optimized using grid search to enhance accuracy, improving overall classification performance**.**

Table 4.15.1 Accuracy and Kappa Value

CP	Accuracy	Kappa
0.046	0.9483	0.9270
0.041	0.9483	0.9270
0.036	0.9483	0.9270
0.031	0.9483	0.9270
0.026	0.9483	0.9270

Based on the table above, it is clear that the hyperparameter "cp" value of 0.046 provides great accuracy for the decision tree. Therefore, we conclude to use the cp value of **0.046** for this analysis.

Table 4.15.2 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	72	0	8	0
	Muscles	0	42	0	0
	Both	0	0	20	0
	Nerves	0	0	0	50

Table 4.15.3 Test statistics

Accuracy	0.9583
95% C.I	(0.9196,0.9818)
No Information Rate	0.375
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.9416
Mcnemar`s Test P Value	NA

The Decision Tree model on the training set shows excellent classification performance with 95.83% accuracy and a Kappa of 0.9416, indicating strong agreement with actual classifications. The "Muscles" and "Nerves" classes are perfectly classified (100% sensitivity and specificity). The "Joint (Nerve and Muscle Area)" class has 100% sensitivity but some misclassifications, leading to 93.33% specificity. The "Muscles and Nerves" class has the lowest sensitivity (71.43%), though its specificity remains 100%. Overall, the model effectively classifies neuromuscular issues, with minor difficulties in distinguishing "Muscles and Nerves" cases.

Table 4.15.4 Training set

		Referance			
		Joint	**Muscles**	**Both**	**Nerves**
Prediction	**Joint**	32	1	4	1
	Muscles	0	17	0	0
	Both	0	0	8	0
	Nerves	0	0	0	21

Table 4.15.5 Test statistics

Accuracy	0.9286
95% C.I	(0.851,0.9733)
No Information Rate	0.381
P Value [Acc > NIR]	<2.2e^-16
Kappa	0.8987
Mcnemar`s Test P Value	NA

The Decision Tree model on the testing set achieved 92.86% accuracy with a Kappa of 0.8987, indicating strong agreement with actual classifications. The "Muscles" and "Nerves" classes were perfectly classified with 100% specificity and high sensitivity (94.44% and 95.45%, respectively). The "Joint (Nerve and Muscle Area)" class showed 100% sensitivity but had some misclassifications, leading to 88.46% specificity. The "Muscles and Nerves" class had the lowest sensitivity (66.67%), though its specificity remained 100%. Overall, the model performs well but struggles slightly with distinguishing "Muscles and Nerves" cases.

4.16 COMPARISON OF MODELS

When comparing the analyses above, select the one that demonstrates the highest accuracy performance.

Table 4.16.1 Training Accuracy and Testing Accuracy

Model	Training Accuracy	Training Kappa	Testing Accuracy	Testing kappa
Random Forest	1	1	0.9286	0.8987
GLMNET	0.9479	0.9275	0.8571	0.802
XG Booster	0.974	0.9637	0.9286	0.8987
SVM	0.9531	0.9348	0.8571	0.7986
Decision Tree	0.9583	0.9416	0.9286	0.8987

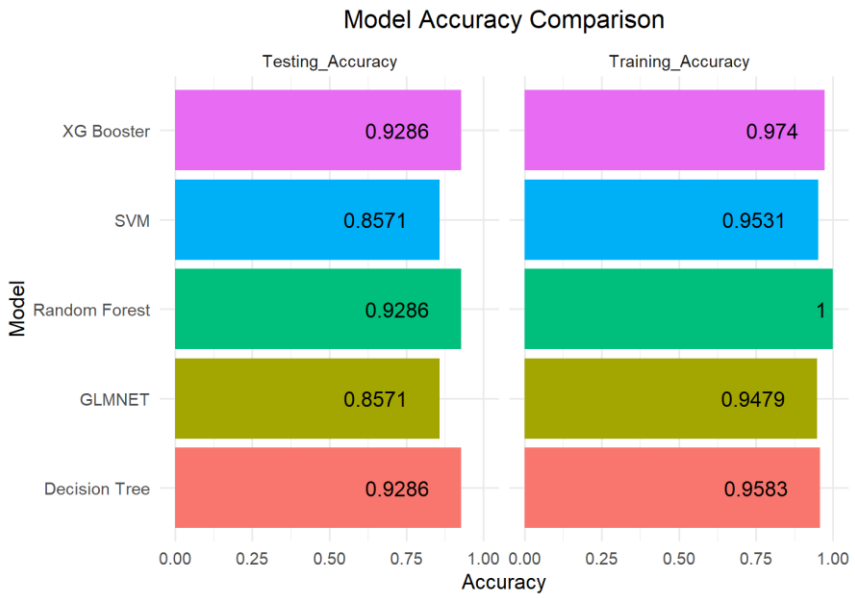

Fig 4.16.01 Comparison of Model Accuracy

The table and figure clearly shows that Random forest has higher accuracy on both the training and testing sets, so we conclude that Random Forest is the best model among those evaluated.

4.17 COMPUTING AUC AND ROC FOR RANDOM FOREST CLASSIFICATION

Table 4.17.1 Train AUC and Test AUC

Major Issue Types	Train AUC	Test AUC
Joint	1	0.9787
Muscles	1	0.9992
Both	1	0.9907
Nerves	1	1
Macro	1	0.9921

The table presents AUC (Area Under the Curve) values for a Random Forest classification model predicting different Major Issue Types in both training and test datasets. The model achieved a perfect AUC of 1.0 for all four classes in the training set, indicating flawless classification on the training data. In the test set, AUC values remain exceptionally high, with "Nerve Issues" achieving a perfect 1.0, while "Joint Issues" has the lowest test AUC of 0.9787, suggesting slight difficulty in distinguishing this class. The overall macro-average AUC is 0.9921, confirming that the model generalizes well and maintains strong predictive performance across all issue types.

INTERPRETATION

The Random Forest model demonstrates exceptional performance, achieving perfect AUC (1.0) in training and high AUC in testing (macro AUC: 0.9921), confirming its strong generalization. By using multiple decision trees and aggregating their predictions, the model ensures robustness and reduces overfitting. While "Joint Issues" has the lowest test AUC (0.9787), the model still classifies all categories with high accuracy. Overall, Random Forest outperforms other models, making it the best choice for classifying Major Issue Types effectively.

FINDINGS,CONCLUSION AND SUGGESTIONS

CHAPTER V

5.1 FINDINGS

- Early-stage patients do not show better treatment adherence than those in later stages. This suggests that being diagnosed early does not necessarily mean patients are more likely to follow their treatment plans.
- The severity of the condition does not determine recovery chances. Even though people with milder conditions might seem to have better outcomes, the overall relationship between severity and recovery is weak.
- The type of treatment does not impact survival during treatment. Regardless of the treatment method used, there is no strong link to survival rates, indicating that other factors may play a bigger role in determining a patient's outcome.
- The main health issue a person faces does not directly improve their quality of life. Even if a specific issue is treated, overall quality of life improvement is not guaranteed, possibly due to other underlying factors affecting the patient's well-being.
- The major health issue does not influence adherence to treatment. Patients with different health concerns do not show significant differences in their ability to follow prescribed treatments.
- Muscle-related issues tend to show lower strength levels than nerve-related issues, and patients with less severe conditions show higher strength levels. However, these differences are too small to be considered meaningful, and further investigation is needed to confirm any true impact.
- Males generally have higher muscle strength than females, and patients with Duchenne Muscular Dystrophy have the lowest muscle strength. However, these differences are not strong enough to be considered statistically significant, meaning that while a pattern exists, it may not be reliable across all cases.
- People who follow their treatment plans and those on higher doses tend to show slightly better progress, but the difference is not strong enough to make a firm conclusion. Other factors, such as lifestyle choices and additional therapies, may also play a role in patient progress.
- The condition tends to worsen over time, affecting breathing ability and muscle strength. Additionally, nerve-related health indicators show a negative impact on

overall well-being. While blood pressure levels remain consistent, they do not strongly correlate with the severity of the condition.

- The CatBoost model demonstrated excellent accuracy in predicting whether a patient has the condition or not. This makes it a useful tool for early detection, allowing doctors to provide timely treatments. However, more testing is needed to confirm if the model works well for different populations.
- The Random Forest model showed high accuracy in identifying cases, outperforming other models. While its accuracy was slightly lower for patients with joint issues, it still performed well across different categories, making it the most reliable model for classifying major health problems.

5.2 SUGGESTIONS

5.2.1 Encourage Treatment Adherence

Patients should be encouraged to follow their treatment plans consistently, as adherence may lead to better health outcomes and improved quality of life**.**

5.2.2 Monitor Disease Progression Regularly

Regular check-ups and monitoring of symptoms can help assess the effectiveness of treatment and make necessary adjustments for better outcomes**.**

5.2.3 Consider Alternative Interventions

If no significant improvements are observed with current treatments, exploring alternative therapies or personalized treatment approaches may be beneficial**.**

5.2.4 Improve Awareness on Health Risks

Providing patients with education on potential health risks and lifestyle modifications can help manage symptoms more effectively and improve their overall well-being**.**

5.2.5 Enhance Support for Patients

Psychological and emotional support should be provided to patients, as it can improve treatment adherence, reduce stress, and enhance quality of life**.**

5.2.6 Optimize Treatment Strategies

A combination of different treatment approaches should be explored to enhance patient outcomes, especially when current methods show limited effectiveness.

5.2.7 Further Research on Factors Affecting Outcomes

Deeper analysis should be conducted to understand the factors influencing disease

progression and treatment response, which can aid in the development of more effective strategies.

5.2.8 Use Advanced Analytical Techniques

Applying more robust analytical methods, such as effect size evaluation and confidence intervals, can provide additional insights into the observed variations in health outcomes.

5.2.9 Schedule Regular Stage Detection and Issue Assessment

Patients should undergo periodic evaluations to detect any changes in their condition and identify major health issues. This can help doctors adjust treatment plans accordingly and ensure better long-term disease management.

5.2.10 Monitor Rate of Disease Progression

Regularly tracking the rate of disease progression is crucial in determining when patients may require more intensive treatment. This helps healthcare professionals intervene at the right time and adjust treatment plans to prevent complications.

5.2.11 Develop Predictive Models for Better Treatment

In the future, constructing predictive models can help determine the most effective treatment plans for patients based on their medical history and disease progression. These models can assist healthcare professionals in making data-driven decisions to optimize patient care.

5.3 CONCLUSION

This study offers an in-depth exploration of the multifaceted factors that influence treatment adherence, the progression of diseases, and the overall recovery outcomes for patients suffering from neuromuscular disorders. The research reveals an intriguing insight: while early diagnosis of these conditions is often considered beneficial, it does not always correlate with improved adherence to prescribed treatment regimens. This suggests a critical need for robust support systems that can assist patients in consistently following their treatment plans effectively.

Furthermore, the severity of a disease does not inherently guarantee a more favorable recovery. This observation highlights the vital necessity for a holistic approach to treatment—one that encompasses not just the physical aspects of the illness, but also lifestyle considerations, psychological well-being, and alternative therapeutic

interventions. It appears that the types of treatment administered do not have a direct effect on survival rates; however, the variability in how individual patients respond underscores the importance of developing personalized treatment plans that are specifically tailored to meet each patient's unique health circumstances.

The study emphasizes the significance of regular monitoring of disease progression, suggesting that continuous assessment of key indicators—such as changes in respiratory function, muscle strength, and nerve health—is essential for timely adjustments to treatment strategies. By ensuring that patients receive periodic evaluations, healthcare providers can better optimize treatment plans and intervene at critical junctures, potentially curbing further deterioration of the patient's condition. Moreover, keeping track of the rate at which the disease progresses allows healthcare professionals to identify moments when more intensive treatments may be required, thereby enhancing patient outcomes.

The impact of machine learning in healthcare is prominently showcased in this study, where advanced models like CatBoost and Random Forest exhibit remarkable accuracy in predicting patient conditions and classifying significant health issues. These sophisticated models facilitate early detection of complications, enable tailored treatment recommendations, and support better disease management by efficiently analyzing complex medical data. By incorporating machine learning into healthcare systems, clinicians are empowered to make informed, data-driven decisions, improve diagnostic precision, and craft targeted treatment plans that elevate the quality of patient care. Nonetheless, further validation of these models across diverse populations remains crucial to ensure their robustness and applicability.

Looking ahead, future initiatives should prioritize the development of advanced predictive models that integrate clinical data, lifestyle factors, and treatment responses to create customized treatment strategies. By harnessing innovative analytical techniques, contemporary machine learning algorithms, and ongoing disease monitoring, healthcare professionals can deliver more precise, timely, and effective interventions. Moreover, enhancing patient awareness, providing psychological support, and implementing strategies to improve treatment adherence will further foster better long-term health outcomes. Ultimately, this study highlights the transformative potential of machine learning within modern healthcare, underscoring its capacity to revolutionize disease management and patient care through accurate, efficient, and personalized solutions

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Sales, D. S., Hammerle, M. B., de Almeida, V. P., de Araujo Davico, C., Pinheiro, P. G., da Silva Souza, R., ... & Peres, K. L. (2025). Dysphagia and its impact on quality of life in rare neuromuscular disorders. *Arquivos de Neuro-Psiquiatria*, *83*(02), 001-006.

[2] Deenen, J. C., Horlings, C. G., Voermans, N. C., van Doorn, P. A., Faber, C. G., van der Kooi, A. J., ... & van Engelen, B. G. (2024). Population-based incidence rates of 15 neuromuscular disorders: a nationwide capture-recapture study in the Netherlands. *Neuromuscular Disorders*, *42*, 27-35.

[3] Margeta, M. (2023). Neuromuscular disease: 2023 update. *Free neuropathology*, *3*, 3-5

[4] Wilson, L. A., Macken, W. L., Perry, L. D., Record, C. J., Schon, K. R., Frezatti, R. S., ... & Hanna, M. G. (2023). Neuromuscular disease genetics in under-represented populations: increasing data diversity. *Brain*, *146*(12), 5098-5109.

[5] Omar, Abdillahi., Marwaha., Komal et al. (2023) . *Physiology, Neuromuscular Junction ,* 4, 15-25

[6] Wilson, L. A., Macken, W. L., Perry, L. D., Record, C. J., Schon, K. R., Frezatti, R. S., ... & Hanna, M. G. (2023). Neuromuscular disease genetics in under-represented populations: increasing data diversity. *Brain*, *146*(12), 5098-5109.

[7] Tayfur, B., Charuphongsa, C., Morrissey, D., & Miller, S. C. (2023). Neuromuscular joint function in knee osteoarthritis: A systematic review and meta-analysis. *Annals of physical and rehabilitation medicine*, *66*(2), 101662.

[8] Lace, B., Micule, I., Kenina, V., Setlere, S., Strautmanis, J., Kazaine, I., ... & Inashkina, I. (2022). Overview of neuromuscular disorder molecular diagnostic experience for the population of Latvia. *Neurology: Genetics*, *8*(3), e685.

[9] Saburova, E. A., Vasiliev, A. N., Kravtsova, V. V., Ryabova, E. V., Zefirov, A. L., Bolshakova, O. I., ... & Krivoi, I. I. (2017). Human APP gene expression alters active

zone distribution and spontaneous neurotransmitter release at the Drosophila larval neuromuscular junction. *Neural Plasticity*, *2017*(1), 9202584.

[10] Woodcock, I. R., Fraser, L., Norman, P., Pysden, K., Manning, S., & Childs, A. M. (2016). The prevalence of neuromuscular disease in the paediatric population in Yorkshire, UK; variation by ethnicity and deprivation status. *Developmental Medicine & Child Neurology*, *58*(8), 877-883.

[11] Deenen, J. C., Horlings, C. G., Verschuuren, J. J., Verbeek, A. L., & van Engelen, B. G. (2015). The epidemiology of neuromuscular disorders: a comprehensive overview of the literature. *Journal of neuromuscular diseases*, *2*(1), 73-85.

[12] Silva, M. E., Mendonça, T., Silva, I., & Magalhães, H. (2005). Statistical analysis of neuromuscular blockade response: contributions to an automatic controller calibration. *Computational statistics & data analysis*, *49*(3), 955-968.

[13] Woodcock, I. R., Fraser, L., Norman, P., Pysden, K., Manning, S., & Childs, A. M. (2016). The prevalence of neuromuscular disease in the paediatric population in Yorkshire, UK; variation by ethnicity and deprivation status. *Developmental Medicine & Child Neurology*, *58*(8), 877-883.

APPENDIX

APPENDIX

The R script used for statistical analysis and Machine learning in my project, originally generated through Jamovi, has been refined and is available in my project's GitHub repository at the following link

https://github.com/Deepaneesh/Neuromuscular-Disorder/blob/main/Complete%20file/Neuromuscular%20code.R

This script includes all the essential code for data processing, exploratory analysis, and modelling.