T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

THE DEPARTMENT OF BIOMEDICAL ENGINEERING

CLASSIFICATION OF GAIT SPEEDS THROUGH TIME-SERIES AND FEATURE ANALYSIS OF SURFACE EMG SIGNALS

MASTER'S THESIS

BASEL KIRRESH

ISTANBUL 2024

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

THE DEPARTMENT OF BIOMEDICAL ENGINEERING

CLASSIFICATION OF GAIT SPEEDS THROUGH TIME-SERIES AND FEATURE ANALYSIS OF SURFACE EMG SIGNALS

MASTER'S THESIS

BASEL KIRRESH

THESIS ADVISOR

ASST. PROF. BORA BÜYÜKSARAÇ

ISTANBUL 2024

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

27/11/2024

MASTER THESIS APPROVAL FORM

Program Name:	Biomedical Engineering
Student's Name and Surname:	Basel Kirresh
Name of The Thesis:	Classification of Gait Speeds Through Time-Series and Feature Analysis of Surface EMG Signals.
Thesis Defense Date:	27/11/2024

This thesis has been approved by the Graduate School which has fulfilled the necessary conditions as Master thesis.

Assoc. Prof. Yücel Batu SALMAN

Institute Director

This thesis was read by us, quality and content as a Master's thesis has been seen and accepted as sufficient.

	Title/Name	**Institution**	**Signature**
Thesis Advisor's	Asst.Prof. Dr BORA BUYUKSARAÇ	Bahçeşehir Üniversitesi	
Member's	Asst. Prof. Dr Burcu TUNÇ ÇAMLIBEL	Bahçeşehir Üniversitesi	
Member's	Asst. Prof. Dr Murat Tümer	Bahçeşehir Üniversitesi	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Basel Kirresh

Signature:

ABSTRACT

CLASSIFICATION OF GAIT SPEEDS THROUGH TIME-SERIES AND FEATURE ANALYSIS OF SURFACE EMG SIGNALS

Basel, Kirresh

Master's Program in Biomedical Engineering

Supervisor: Assist. Prof. Dr. Bora Büyüksaraç

November 2024, 158 pages

Different studies have been researched in lower body kinematics, gait, and electromyography (EMG) measurements. Understanding these measurements results and conducting thorough study will have a better evaluation of the human body locomotion. Treadmills have been used in the research widely and while they have some drawbacks, they still provide sufficient data. Treadmill limits the space of walking but controls the subject's walking speed. This research provides a thorough study of 25 healthy subjects who walked on the treadmill consisting of speeds starting from 0.8 m/s till 2 m/s. Ten different muscles of both sides of the body on the right and left leg are tested which are Tibialis Anterior, Gastrocnemius Medialis, Soleus, Vastus Medialis, Biceps Femoris. Fifty heel strikes are taken from both feet for the sake of normalized analysis. The tools that were used in this research are Movella, EMG sensors, and a treadmill. Methodically, the data was acquired in a uniform fashion therefore 50 heel strikes were recorded from each leg during the experiment.

Classification processes were carried out to point to the specific speed based on a specific muscle's activity produced by a certain muscle. It is proposedly expected that each EMG activity's pattern at different speeds is classified accurately using machine learning. These speeds which can be considered slow, medium, and fast pace, the muscles mentioned exhibit the most activity in the lower extremities in comparison

with other muscles. The results showed high accuracy while doing the time series method in comparison with the feature-based analysis.

Key Words: EMG, Classification, Machine Learning, Time Series, Feature-based.

ÖZET

YÜZEY EMG SINYALLERININ ZAMAN SERISI VE ÖZELLIK ANALIZI ILE YÜRÜME HIZLARININ SINIFLANDIRILMASI

Basel, Kirresh

Biyomedikal Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Bora Büyüksaraç

Kasım 2024, 158 sayfa

Alt vücut kinematiği, yürüyüş ve elektromiyografi (EMG) ölçümlerinde farklı çalışmalar araştırılmıştır. Bu ölçümlerin sonuçlarının anlaşılması ve kapsamlı bir çalışma yapılması, insan vücudunun hareketinin daha iyi değerlendirilmesini sağlayacaktır. Koşu bantları araştırmalarda yaygın olarak kullanılmaktadır ve bazı dezavantajları olsa da yine de yeterli veri sağlamaktadır. Koşu bandı yürüme alanını sınırlar ancak deneğin yürüme hızını kontrol eder. Bu araştırma, 0,8 m/s'den 2 m/s'ye kadar hızlardan oluşan koşu bandında yürüyen 25 sağlıklı denek üzerinde kapsamlı bir çalışma sunmaktadır. Tibialis Anterior, Gastrocnemius Medialis, Soleus, Vastus Medialis, Biceps Femoris olmak üzere sağ ve sol bacakta vücudun her iki tarafındaki on farklı kas test edilmiştir. Normalize edilmiş analiz için her iki ayaktan elli topuk vuruşu alınmıştır. Bu araştırmada kullanılan araçlar Movella, EMG sensörleri ve bir koşu bandıdır. Metodik olarak, veriler tek tip bir şekilde elde edilmiştir, bu nedenle deney sırasında her bir bacaktan 50 topuk vuruşu kaydedilmiştir.

Sınıflandırma işlemleri, belirli bir kas tarafından üretilen belirli bir kas aktivitesine dayalı olarak belirli bir hıza işaret etmek için gerçekleştirilmiştir. Her bir EMG aktivitesinin farklı hızlardaki örüntüsünün makine öğrenimi kullanılarak doğru

bir şekilde sınıflandırılması önerilmektedir. Yavaş, orta ve hızlı hız olarak kabul edilebilecek bu hızlarda, söz konusu kaslar diğer kaslara kıyasla alt ekstremitelerde en fazla aktiviteyi sergilemektedir. Sonuçlar, özellik tabanlı analize kıyasla zaman serisi yöntemi kullanıldığında yüksek doğruluk göstermiştir.

Anahtar Kelimeler: EMG, Sınıflandırma, Makine Öğrenmesi, Zaman Serisi, Özellik-tabanlı.

This thesis is dedicated to my Palestinian people, who endure through the toughest of times. Your unwavering spirit, resilience in the face of adversity, and enduring hope for a brighter future serve as a constant source of inspiration. May this work contribute in some small way to the pursuit of justice and self-determination for Palestine.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor, Assist. Prof. Bora Büyüksaraç, for his invaluable guidance and support throughout this research. More than just an advisor, his mentorship has profoundly shaped my experience as a master's student, encouraging me to think critically and independently. Beyond his academic guidance, I deeply appreciate his kindness and unwavering support, which fostered a positive and encouraging learning environment. It was a true privilege to learn from and work with someone so dedicated and passionate about his field, and I am incredibly grateful to have met someone like him during this important stage of my academic journey.

I would like to express my heartfelt gratitude to my family, especially my father and mother, for their unwavering love and support throughout my studies. Their encouragement, understanding, and sacrifices have been instrumental in my pursuit of higher education, and I am eternally grateful for their belief in me. I could not have reached this point without them.

I also extend my sincere thanks to my friends, whose companionship and support have enriched this journey. I am particularly grateful to Ellie, Ayman, and Omar, whose contributions to my work and presence in the lab were invaluable. Their willingness to offer help, share ideas, and provide encouragement was essential to my progress. I am also thankful for the support of the many other friends in the lab who created a positive and collaborative environment.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AJC	Ankle Subtalar Joint Complex
AJC	Ankle Subtalar Joint Complex
ANN	Artificial Neural Network
BF	Biceps Femoris
BP	Back Propagation
DF	Dorsiflexion
DT	Decision Tree
EMG	Electromyography
GM	Gastrocnemius Medialis
KNN	K-Nearest Neighbor
LR	Linear Regression
MLLP	Microprocessor-Controlled Lower Limb Prostheses
MUAP	Motor Unit Action Potential
PF	Plantarflexion
PPV	Positive Predictive Value
ROM	Range of Motion
SOL	Soleus
SVM	Support Vector Machine
TA	Tibialis Anterior
TPR	True Predictive Rate
VM	Vastus Medialis
MAV	Mean Absolute Value
RMS	Root Mean Square
TM	Temporal Moment

Chapter 1
Introduction

1.1 Theoretical Framework

Gait is defined as the walking pattern of the *Homo sapiens* conduct from the movement of the body's joints. Walking in gait involves moving the body joints and exerting some force on them from the muscles. The gait analysis is carried out by observing 3D analysis by considering the data of joint angles, joint forces, muscular activity, foot pressure, and energetics. Locomotion is considered very complex and involves the arrangement of the bone, the joint range of motion, and neuromuscular intervention. Gait analysis can be evaluated in different ways. One is observing the person's normal walking and other is a computerized method by using 3D motion analysis as well as considering biomechanical energy (Chambers & Sutherland, 2002).

Gait is a motion that starts with one foot strike and ends with another strike performed by the same foot. Gait consists of two phases of the gait cycle which illustrates the individual's walking. Stance phase or initial phase starts with a foot strike and ends with a toe off. This phase is about 62% of the complete gait cycle and the swing phase which is about 38% begins with toe off and ends with foot strike. The cycle that takes place has different stages and each stage is divided into percentages. Initial contact is defined as 0% and the same following strike of the same leg is defined as 100%. Different phases of the gait cycle which are foot strike, opposite foot's toe off, opposite foot's strike then toe off, foot clearance, tibia vertical, and sequential foot strike. The stance phase has three periods of time. First is the initial double limb support, single limb stance, and pre-swing. Two most obvious events in the double limb support are foot strike and opposite toe off. The known events for single limb stance are the opposite leg's toe off and foot strike. The swing phase is the second phase of the gait cycle, and it is divided into three sections. Initial swing, mid-swing, and terminal swing are the three parts in the swing phase. Initial swing's subsequent events are toe off and foot clearance. Mid-swing part starts with foot clearance and finishes with tibia vertical (Chambers & Sutherland, 2002).

Temporal parameters are an important factor in gait analysis. Temporal factor is the time and distance conducted by the subjects. A temporal parameter which is the velocity which is reported to have different velocities for different age ranges. A 7-year-old child's velocity is 114 cm per second while an adult of 40 years and more has

a velocity 123 cm per second. The step length is measured from taking the foot strike of one foot till the foot strike of the other foot. The stride length is the measurement of one foot strike till the next foot strike of the same foot. For gait analysis there are several factors that should be considered including neurological or gait abnormalities (Chambers & Sutherland, 2002).

Muscle action potential can be observed and deduced from the subject's walking pattern. Although this is possible, it cannot be certain that a specific movement can cause the muscle to be active or non-active. This type of observation is important to know which therapeutic intervention to use to solve the problem. Surface or fine-wire EMG are two measuring equipment to measure the muscle activity's amplitude of the tested subject (Chambers & Sutherland, 2002).

Biomedical signals are an electrical signal that any organ gives which can illustrate a physical variable that is relative to the research topic. The biomedical signal is usually the function of time. Amplitude and frequency are the factors that are observed in each measurement to observe the effect of muscles. There could be noise throughout the measurement which comes from the tissues because of the movement. EMG is being researched to manage and rehabilitate the motor disabilities for patients in clinics. Motor unit action potentials shape and firing rates in EMG signals will provide critical information for diagnosing neuromuscular disorders. During the process of acquiring data from the EMG signal, there are two concerns that affect the accuracy of the signal. Signal-to-noise ratio and distortion of the signal are the two problems that are encountered in EMG acquiring data. The noise is the unwanted signal in the EMG which should be filtered or cancelled. The distortion of the signal means that any frequency element should not be changed to maintain EMG frequency. The surface EMG that is directly placed on the skin acquires all the signals from the muscle fibers combined. There can be a negative or positive value of voltage because the action potential happens randomly. Motor unit action potential is the activity from all the muscle fiber action potentials combined and this activity can be detected by a noninvasive electrode. The electrode processes the signal, and it is amplified. What interests the user is the amplitude of the signal therefore it is rectified, and averaging is applied to show the EMG amplitude (Reaz, Hussain, & Mohd-Yasin, 2006).

Electromyography (EMG) is the neuromuscular activity that is produced from contracting the muscle. EMG is a useful tool commonly in identifying neuromuscular disorders. Neuromuscular disorder is any nerve or muscle dysfunction. Myopathy and neuropathy are two severe neuromuscular diseases and can be classified based on EMG characterization systems (Yousefi & Hamilton-Wright, 2014).

Surface EMG was used in walking research projects. Subjects were investigated and data was obtained while they are walking on treadmill at different speeds. Lower extremity muscles are mainly observed for these kinds of tests. A set of speeds is tested by using the treadmill and a set on ground level. The use of treadmill rather than ground walking is for normalization purposes. The increment in speeds that have been used in research is 0.1 m/s. The temporal parameters have main events which are both the swing and stance. There are several phases in stance which should also be evaluated. Gait speeds can be used as a method to differentiate between the healthy and the rehabilitating subjects in which the spatiotemporal parameters will indicate the difference (Hebenstreit et al., 2015).

The parameters that are collected during the gait and EMG experiment are the joint angles and muscle activity. EMG has been acquired from different muscles in previous studies such as Tibialis anterior, Gastrocnemius Medialis, Biceps Femoris, Soleus and Vastus Medialis. (Moreira, Figueiredo, Fonseca, Vilas-Boas, & Santos, 2021). The sEMG is the method for measuring muscle activity and there are two different sEMGs. One is static and the other is the dynamic EMG in which the dynamic is the mostly used in the walking experiments. A PC interface biometric acquisition system receives the EMG signal and processes them (Kunju, Kumar, Pankaj, Dhawan, & Kumar, 2009).

Recognizing gait phases is a primary step to gain a spatiotemporal characterization of the actively required muscle during walking. Some researchers have looked and collected data in this matter by using machine learning methods on treadmill gait collected data. Previous research has tried to identify certain stages in the gait like foot-floor contact and ground clearance (Morbidoni, Cucchiarelli, Fioretti, & Di Nardo, 2019). In the gait domain, it always has been a problem identifying a moving object from the gait analysis. Clinical diagnostics can be used by solving this problem as well as identifying the subject's speed plays a major role in monitoring his

health, tracking the mentally incapable of performing specific tasks, and in addition it will help improve the assistive devices for the patients with orthopedic problems. Machine learning methods have been effective when it comes to implementing them on gait analysis problems with large data sets. Even though it might be effective the magnitude of the feature set that is fed into the machine learning plays a major role. A big feature set can cause confusion for analyzing other data sets for the long term. In machine learning, there are many algorithms to be used within the scope gait analysis and EMG analysis. Many of these that are used in research which make automated decision making are "Logistic regression (LR), Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN)" (Chakraborty & Chattaraj, 2023). Several factors play a role when it comes to machine learning analysis and observing the result of the classification. Accuracy, precision, recall, f1-score, and ROC curve can be either done in one vs one strategy or one vs rest strategy. In these of gait and EMG analysis, the one vs rest strategy is more frequently used since each class was separated individually. One vs all strategy can be used in this kind of research since it can be used to differentiate multiclass groups (Gao et al., 2021).

Other research has collected data by selecting four different situations which are slow walking, moderate walking, fast walking, and some other activity; some have included sitting and others have conducted an experiment on subjects that go on incline or decline. This research will study a wide range of speeds and conclude the normal speeds of walking subjects based on several percentages within the recorded data. Additionally, this strategy is used to avoid too many complications in analysis purposes in this work and others (Chakraborty & Chattaraj, 2023) (Kim, Lee, & Shin, 2021).

1.2 Statement of the Problem

One of the main aims of speeds and gait analysis research is to make assistive devices to support people with walking problems. Walking is one of the most active motional tasks in the human body. It can be affected though from neurological diseases like spinal cord injury and stroke can be a huge impact. Robotic assistive devices are used to help people with abnormal walking. Different factors play a role in the human locomotion state which are the kinematic, kinetic, and electromyography (EMG).

People that have defects in their lower limbs walk slower than normal subjects at the tested speeds (Moreira, Figueiredo, Fonseca, Vilas-Boas, & Santos, 2021).

1.3 Purpose of the Study

Classification for a set of incrementing walking speeds starting from slow to high walking speeds. This classification will propose a deep learning approach based on sEMG data to better understand and classify human's normal walking speeds from analyzing a diverse range of controlled speeds set on a treadmill. The aim of this study is to classify walking speeds based on the muscle activity of the lower limbs. EMG measurements were carried out for the muscles which are involved in human locomotion. The muscles we have selected for this study are vastus medialis, biceps femoris, gastrocnemius medialis, tibialis anterior, and soleus. The activity of these muscles is measured for 50 gait cycles at 15 different speeds. The speeds which are in respect to meters per second are 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0. These speeds can be considered as slow, moderate, and fast walking speeds which were investigated thoroughly in this research. The steps within this research are as follows: data acquisition, data pre-processing, data modelling, data analysis, classification and interpretation.

1.4 Hypothesis/ Research Question

Research of the study is differentiating and classifying different speeds of a subject from EMG and gait data based on time-series and feature based analysis to establish a firm understanding of EMG data that suits that certain speed.

Hypothesis: different speeds affect the lower extremities muscles showing significant differences in gait and EMG activity and classification tools can detect which muscle activity best suits a certain using speed time-series and feature based analysis.

1.5 Significance of Study

Further study of EMG signals to try and differentiate different gait cycles of subjects' gaits to identify pathological disorders through recognizing that specific speed from that EMG data. This will better shape a diagnosis in clinics for patients with walking disorders.

Chapter 2
Literature Review

2.1 Joint Angles

Joints in brief are the meet point of two bones. The movement and stability come from the joints that bring the bones together. As we age, the bones become more fragile and more sensitive to external factors. Diseases throughout the lifetime affect the joints, making them weaker and they can alter normal walking. Three types of joints are found in the human body. Fibrous, cartilaginous, and synovial are all kinds of joints that are found in the human body making them special in their own ways. The fibrous is immovable while the cartilaginous is semi-movable and the synovial is a joint that moves freely making it the most flexible type of joint. The synovial which is freely movable is the main joint of the body because it allows mobility and withstands weight while maintaining low friction and does not allow water passing; all this occurs between two articulating bone surfaces.

From the musculoskeletal point of view, the disorder that is mostly common is arthritis and is a big case in the population of the world. This disease limits movement and causes physical disability in that specific joint. People with arthritis are limited to daily tasks like walking or climbing stairs. Ageing and poor physical activity are factors that contribute to weaknesses or loss of mobility. Analyzing humans from the joint posture and movement's point of view will help that person in rehabilitation or guided training in case of help with physical activities. To do this, tests are required to see the mobility and musculoskeletal status in detail. Specific parameters are extracted while analyzing that subject's data through sensing technologies. Many of these technologies are complex and require huge set-up time. The advancement in technology has to lead to faster and reliable equipment. These sensors can record real-time data. These sensors can detect and record several data; these sensors can be EMG, ECG, and more (Faisal et al., 2019).

Lower extremities joints play a major role in stabilizing the human body and maintaining a firm stance on the ground. These extremities are the foot, ankle, knee, and hip. Different intense activities require the aid of the extremity joints which comes from the leg and foot synergetic process (Ugbolue et al., 2021). The ankle and the foot consist of 26 separate bones of the foot. There is a long bone connecting the lower

limb and the foot and ankle making a total of thirty-three joints. The ankle subtalar joint complex (AJC) which is the ankle and the subtalar joint forms the major part in these intense activities. The joint within this complex moves in a synergetic fashion in order to have a stable step on uneven surfaces (Yamaguchi, Sasho, Kato, Kuroyanagi, & Banks, 2009).

Talocrural joint which is also called the ankle joint is moved by two joints. The tibia and fibula from the leg connected to talus of the foot. Both the tibia and the fibula of the foot form an inverted U-shaped structure in the region of connection of the talocrural joint. Below this structure comes the talus which is on the foot, which has a wedged form in which its head is almost 90 degrees to the tibia (McKeon & Hoch, 2019). The subtalar joint is a complex structure that is found in the ankle. It promotes the movement of the foot including pronation and supination. Within these two movements there are also dorsiflexion (DF) and plantar flexion (PF) (Angin & Demirbüken, 2020).

The anatomy of the subtalar joint can be divided into two parts, an anterior and posterior part. In the anterior part, the talus head is located and in the posterior part is where the curve of the talus will lie on the curve of the calcaneus. Calcaneus is the largest part in the ankle. The range of motion that is in the subtalar joint is from 25 to 30 in inversion and 5 to 10 in eversion. Some tendons come over the subtalar joint thus balancing the ankle in the stance phase and throughout the gait cycle (Angin & Demirbüken, 2020) (Brockett & Chapman, 2016).

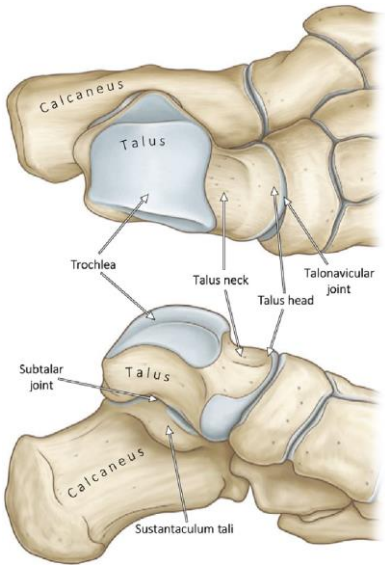

Figure 1. Talocrural joint (Angin & Demirbüken, 2020).

Knee plays an important role in affecting the DF and PF angles at the ankle from its connection with the Achilles tendon. The Achilles tendon either contracts or relaxes depending on the surfaces pressed on. The femoral condyle that is on the knee is transferring load to the tibial plateau. This load is transferred and pressures on the medial and lateral parts of the tibial plateau relieve on the knee structure and if some deformities such as varus and valgus exist then the factors will be changed. The vargus and valgus deformity is when this suspected individual has an internal or external pushing of the knee. Any of these deformities or some other disability is visible then the ankle-subtalar joint is at risk and can cause undesirable changes. The knee assists in the movement of the body from neuromuscular structures in which it allows the lower limbs to keep stability and helps the biomechanics of the lower limbs. The hip has three-dimensional motion including two ligaments that stabilize the hip (Ugbolue et al., 2021) (Gibson & Prieskorn, 2007).

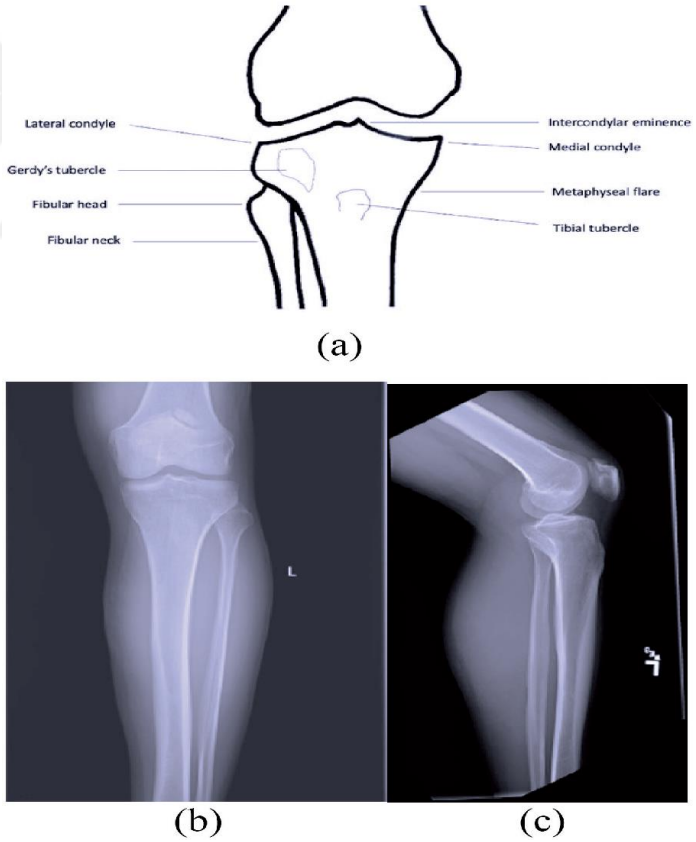

Figure 2. Knee joint (Schmidt, Szatkowski, & Riehl, 2020).

A specific motion or activity that the human body makes will cause that person of interest to make a joint angle motion in which the ideal range is within each person's capacity and strength. This ideal range can be reached within the muscle obtaining

enough length to withstand maximum strength. In literature, it has been provided with the ideal range of joint angles of healthy individuals. The range of motion (ROM) will alter from several factors such as sex, age, body structure, and life routines. In the figure below it is shown the range of motion and the name of that motion of different joints in the body. Humans have these different joints that contain different ranges of motion and the structure of the joint in that area allows this specific motion for some joints and not in others (Schmidt, Szatkowski, & Riehl, 2020).

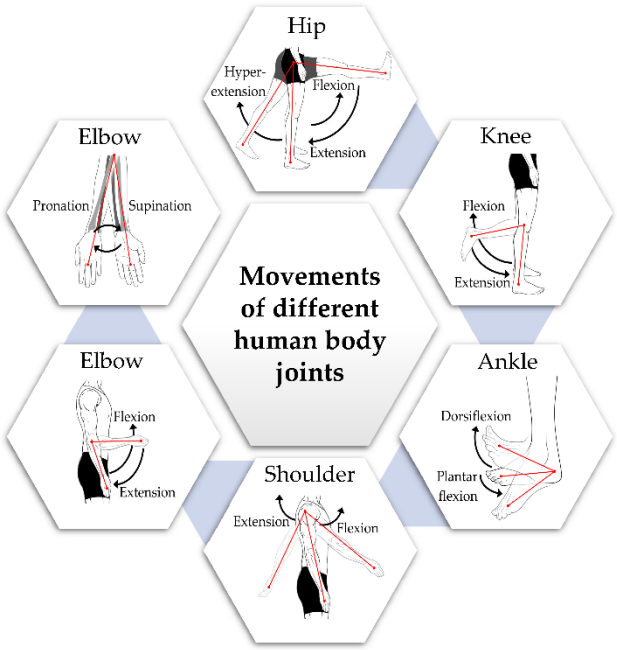

Figure 3. Representation of different movements of human body (Schmidt, Szatkowski, & Riehl, 2020).

When explaining about the motion of the joint its always referred from the joint's center location of that joint of interest. There are five different joint motions which are flexion, extension, adduction, abduction, and rotations. A range of motion value will show a full movement of that specific joint so for example for a knee flexion its value is around 137 degrees for males of ages between 20 and 44 years old. One of the ways to measure this ROM is through specific monitoring systems like 3D IMU sensors that can detect the motion in all planes which will give any motion conducted by the person of interest (Schmidt, Szatkowski, & Riehl, 2020).

2.2 Gait

Gait process is the unique walking style of a certain individual. Gait analysis can be done by either observing the subject's walking or using a 3D analysis that considers

the joint angles, joint forces, and muscular activity, and foot pressure. This analysis will help the physician to design a treatment plan accordingly with the patient's needs. The locomotion is a complicated topic which involves the bones aligned and their interaction, joint range of motion, and neuromuscular activity. Any congenital disability or injury can greatly impact and negatively affect the gait process. An analysis of the individual's gait is the best option to evaluate and design a process to correct the person's gait (Chambers & Sutherland, 2002).

The characteristics of gait cycle are composed of two different stages. The gait cycle is when an individual takes a step from one heel strike of the same foot twice. Stance phase is the first stage of the gait cycle in which "which begins with foot strike and ends with toe-off, usually lasts for about 62% of the cycle; the swing phase, which begins with toe off and ends with foot strike, lasts for the final 38%." In each cycle that takes place there are consequential steps and illustrating that event as part of the whole gait cycle will normalize the gait cycle. The first foot strike is considered 0% and the next heel strike of that same foot is considered 100% of the gait cycle. The different stages in the gait cycle are as follows foot strike, opposite foot toe off, "reversal of fore shear to aft shear, opposite foot strike, toe off, foot clearance, tibia vertical, and successive foot strike". Stance stage is divided into three main parts which are initial double limb support, single limb stance, and pre-swing. This is illustrated in the figure below where it shows the complete gait cycle. Double limb support means that both legs will touch the ground. One foot will conduct the foot strike while the other foot is in toe off stage. A single limb stance would be one toe off leg and one foot strike leg. Reversal of fore to aft shear is also another part of single limb stance in which it has the midstance and terminal stance. The swing phase is divided into initial swing, mid-swing, and terminal swing. The initial swing phase is when the toe is off the ground and foot clearance. The mid swing stage will begin with foot clearance and its final stage is tibia vertical. The terminal stage is the last stage in which the same foot that strike first will strike here again. Two different forces give good analysis results. Temporal parameters and force are both good contributors for gait cycle analysis (Chambers & Sutherland, 2002).

Temporal, which is the time and distance parameter meaning in other words velocity in cm/s. Cadence is the number of steps per second which is another parameter in the temporal spectrum. Step length is the distance between one foot strike and the

other opposing foot's strike. Another important parameter in the temporal spectrum is stride length in which it is the distance between the foot strike and the consequent foot strike when talking about the same foot. Force parameters are important in maintaining balance. Gait is the interchange of losing balance and maintaining balance. Center of mass (COM) of the body will keep changing constantly with the individual's forward movement. When the person moves forward unto the limb carrying weight, COM of the human body will move forward. Gravity and ground reaction forces both modify and affect humans in motion (Chambers & Sutherland, 2002).

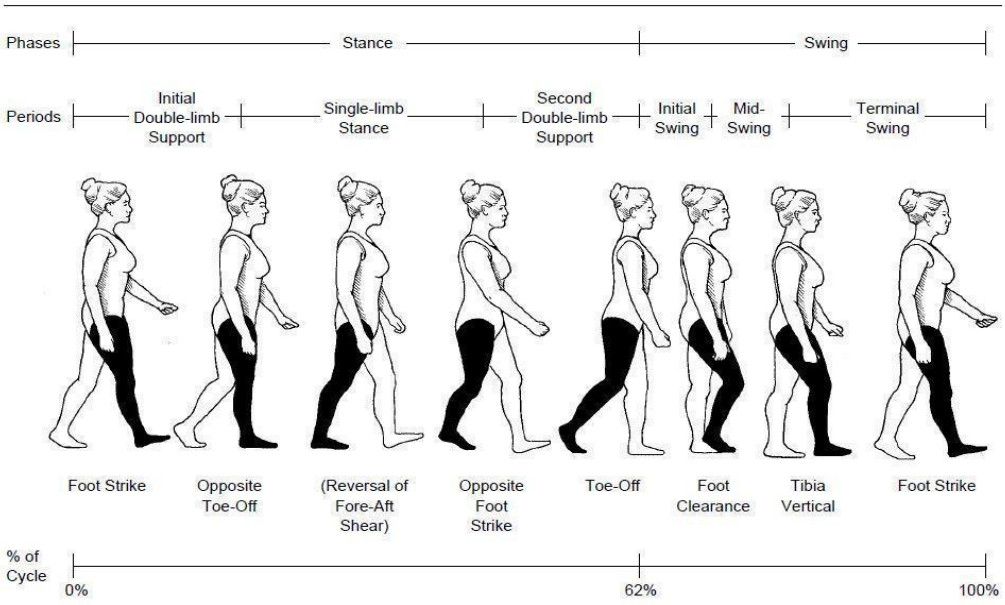

Figure 4. Phases of gait cycle (Chambers & Sutherland, 2002).

Analyzing human locomotion such as pedaling, walking, race walking, and running shows a consistent and cyclic pattern with time. It seems like a simple task to do but in fact it involves a complex function which will need accurate synergy. The lower limbs will need to cooperate and in parallel muscle contraction should occur. While humans walk, the stride-to-stride varies due to the corrections made by the walking system to prevent ordinary movements. The more there is variability in the person's movement the poorer the coordination level is while low variability shows high coordination level. Each human's locomotion and speed on ground can vary which can result in a change of comparative analysis.

2.2.1 Normal walking. While private cars take first place as the most preferrable way of transportation, the second place is walking. The daily activities and functions which the individual does throughout the day will affect his or her walking speed. People that have past strokes or acute diseases which affect their walking are affected by two factors that play a main role and affect gait speed which are the time of stay of the foot and discharge of the foot. There have been several studies that proposed the regular comfortable walking speed. Men in their 20s and 30s have a comfortable walking speed of 139.3 cm/s for 20s men and 145.8 cm/s for 30s men. For women in the 20s its 140.7 cm/s and 30s its 141.5 cm/s (Bohannon, 1997) (Bohannon & Andrews, 2011).

Walking is one of the least skillful and does not require any equipment which makes it the best exercise for the individual to perform daily. A weekly effort of a minimum of 150 minutes of medium intensity of any physical exercise is required. The non active people or low active people can benefit in a healthy way from performing easy or little effort task. Speed intensity is recommended by some research conducted in labs. The pattern and manner of how someone walks will explain a lot about his health. The walking speed of the person is a strong factor in data analysis which can predict mortality in a clear fashion (Murtagh, Mair, Aguiar, Tudor-Locke, & Murphy, 2020) (MejiaCruz et al., 2021).

Normal and easy walking pace can be around 5-8 km/hr and will achieve the recommended physical activity standards. Although this seems like the ideal daily activity to do, not many meet this standard. One way to motivate the public is to walk in groups. It will attract people who are not interested in walking and promote this required walking style. The group walking will also further improve physical activities in public (Hanson & Jones, 2015).

2.2.2 Treadmill walking. Treadmill walking provides consistent and efficient recording of the gait. Treadmill is well supported for measuring long-duration trials which will provide sufficient and comparative data for analysis. Restricted control over the speed can also be achieved while using the treadmill for acquiring data. Walking pathologies and disorders are widely researched on treadmills (Padulo et al., 2023).

Sensorimotor conditions differ when performing overground and treadmill walking. In depth gait analysis can be performed while the subject is conducting the over ground walking path, but it will hinder processing. It will require the participant to perform many turns, and unusual turns therefore will add complexity to the data. Treadmills eases the data acquired for the researcher and helps the participants to reduce walking complications. Small labs can perform treadmill gait testing in which many continuous gait cycles can be recorded. Precise control over speed is one of the important factors here in which it is set in a treadmill and is the core for the studies that analyze gait patterns which are speed dependent. Adaptation period differs from person to person. In general, the adaptation period of a healthy individual can range from a few seconds to several minutes. Several researchers have looked at into the ideal adaptation of walking speed when walking on a treadmill. One research showed that acclimatization happens in the first 10 seconds, but an ideal state of adaptation reached at 10 minutes. Some other work showed that 30 seconds is needed for an ideal state of adaptation, but it will require 10 minutes to reach the plateau. Three different studies have proven that between 4-6 minutes is the adaptation period. Although this is the time it takes for adaptation, in older adults this was not the case. It took even more than 15 minutes, and it was not of a huge adaptation significance (Meyer et al., 2019).

Several purposes can be shown when walking on a treadmill comparable to walking overground. Naturally, over-ground walking is more reflective on the normal human walking, but treadmills' use has increased throughout the recent years since home treadmills are more frequently used. More importantly, treadmills are widely used in clinics for rehabilitation purposes and measuring gait performance. Gait disorder such as cerebral palsy or Parkinson's disease can be further aided and researched on treadmill to be used as a rehabilitation method. When testing, stationary individuals will give more options for a therapist to intervene and help if needed. All treadmill-based studies mimic overground kinematics and gait to assess easier and provide accurate results (Semaan et al., 2021).

2.3 EMG

In the biomedical field, electromyography or EMG is widely used for muscle activity signal detection. Biomedical signals are electrical signals that are detected

from the muscle or specific tissue through skin contact when attaching the required sensor. The EMG signal is a function of time and can be explained in either amplitude, frequency, or phases. During contraction, the muscle emits some electrical signal, neuromuscular activity, which can be detected from an EMG sensor. Therefore, this EMG signal can be considered as a complex signal since the nervous system manages the muscle's activity. Also, the anatomical and physiological factors affect the EMG signal indicating that different individuals emit different EMG signals depending on their muscle structure and function. Data analysis and EMG acquiring will help in biomedical fields and clinical diagnosis. The motor unit action potential (MUAPs) gives a lot of meaning to the EMG data; this can be deduced from its shape and firing pattern (Reaz, Hussain, & Mohd-Yasin, 2006).

EMG signals can be measured through different muscles in the human body. Different road paths can affect EMG signals diversely. The human's body contraction and muscle movement are what the EMG receives as a signal. Surface EMG is what was used in research to detect the motor unit action potential. Surface EMG is used as a non-invasive technique in different scopes of study. It can be used to study different pathologies like back pain and neurological diseases. These tests depend on several factors like statistical results and in addition one of the most important factors is electrode placement. Signal filtering methods are very useful when measuring EMG signals. Noise, signal instability, and motion artifact are to be expected during measurements, but all can be Withdrawn during filtering. EMG steps are as follows: data acquisition, data pre-processing, data modelling, data analysis, and interpretation. In this study, the lower limbs are mainly studied (Claudiane Arakaki Fukuchi, 2019).

Gait and EMG were researched in many ways. Research showed the effect of decreasing speed from fast to slow speed and in contrary from slow to fast speed. Walking speed is shown to be a main affecting factor in kinematic and kinetic changes as well as the biomechanical variables like joint kinematics, muscle activity, and spatiotemporal gait parameters (Claudiane Arakaki Fukuchi, 2019). A biomedical signal is a function of time and has different outputs such as amplitude, frequency, and phase. These different outputs that were extracted from gait and EMG analysis resulted in different analysis for different concepts (M. B. I. Reaz, 2006). Treadmill was preferred in the works since it gives a good movement within a certain space. There were results that overground walking and treadmill walking showed significant

difference in the step distance and step time (Sam Chesebrough,2019). Individuals have various muscle activation characteristics, and it was tested using support vector machine learning. Myoelectrical activity is tested from lower extremity muscles like vastus medialis, biceps femoris, gastrocnemius medialis, tibialis anterior, and soleus. Classification based on the gait and EMG can be used to differentiate different subjects. The classification process involves training and testing sets. Studies were performed on individuals to test if they have their own signature. Muscle activation differences should stay the same for some time. Another sign that this is their own signature is that this action that was done shouldn't be the same for any 2 people doing that specific task. Support vector machine approach was used to differentiate these muscle activation signatures. The length a gait cycle takes is a crucial spatiotemporal parameter. Understanding this cycle's parameters will create a differentiation between healthy and pathological defects in the human body (François Hug,2019). Gait phases mainly are affected by the speed of which the person is advancing. Musculoskeletal or neurological diseases can majorly affect gait cycles. The spatiotemporal parameters can help the researchers differentiate between different levels of pathological gait. When observing the basic temporal parameters, it is about time of swing and stance phases. These spatiotemporal parameters along with EMG signals are implemented for classification (Felix Hebenstreit, 2015).

EMG signals can be measured through different muscles in the human body. Different road paths can affect EMG signals diversely. The human's body contraction and muscle movement are what the EMG receives as a signal. Surface EMG is the used tool in research to detect the motor unit action potential. Surface EMG is used as a non-invasive technique in different scopes of study. It can be used to study different pathologies like back pain and neurological diseases. These tests depend on several factors like statistical results and in addition one of the most important factors is electrode placement. Signal filtering methods are very useful when measuring EMG signals. Noise, signal instability, and motion artifact are to be expected during measurements, but all can be Withdrawn during filtering. EMG steps are as follows: data acquisition, data pre-processing, data modelling, data analysis, and interpretation the lower limbs of interest.

2.4 Rational for Muscle Selection

2.4.1 Tibialis anterior. Tibialis anterior is one of the most researched lower extremities muscles in the human body when researching gait and EMG. Using sEMG, many acquired data was previously researched on the TA and especially in different gait speeds. It has been observed that throughout the increasing speeds, there was an increase in amplitude in the TA. All the lower leg muscles show critical and important effort while walking and especially in the stance phase. TA has two main key functions while walking. It helps lift your foot upwards, which is also called a dorsiflexion movement. This dorsiflexion action happens during the swing phase of the stride which makes sure the toes leave the ground and the foot lands precisely. The second key function is controlling the foot as it lands. As the foot touches the ground, the TA muscle will aid in controlling the plantarflexion movement which will absorb the shock of the force applied on the ground. Mid-stance and late stance show differences in TA activity while investigating different speeds (Péter et al., 2019) (Maharaj et al., 2019).

2.4.2 Gastrocnemius medialis. The ankle push-off or the ankle plantar flexion in the late stance phase and the GM plays an important role in that action. The ankle plantar flexion performs the push off during walking. (Kibushi, 2023). The muscle head which is called femur is what helps in flexing the knee. It also relates with the posture of the foot especially in walking, running, and jumping. When walking, the intervention of the quadriceps and gastrocnemius takes place. The two phases in the gait that need assistance of the GM is the first phase of foot support and swing phase (Bordoni & Varacallo, 2023). The gastrocnemius muscle quickly and powerfully contracts while walking which helps push off the ground and create a forward momentum for each step. This step will generate some speed and force in the ankle. During the human walking, as speeds increased it has been proven that it will require more work from the plantar flexors (Farris & Sawicki, 2012).

2.4.3 Soleus. Like the GM, the soleus also is a plantar flexor muscle that helps in push off. It is part of the triceps surae complex in which they are the medial gastrocnemius, lateral gastrocnemius, and soleus. The soleus is a muscle that is within the gastrocnemius muscle. In a walk without running, it is the soleus muscle that intervenes most in the plantar flexion of the foot. Soleus muscle is responsible for the contraction of the fibers in the calves, and it is about 62% of the effort needed on higher gait speed. The monoarticular soleus which is the most extensive muscle of the plantar flexors. It forms the most mechanical power at the ankle joint while walking (Kharazi et al., 2023). The soleus also has a role in maintaining posture to maintain standing position. The SOL is known as "first gear muscle" which means it is the most primary and effective muscle when plantarflexion action takes place during a bent knee position. During plantarflexion, soleus is the primary muscle for that task because of its fatigue-resistant slow-twitch fibers (Vandervoort & McComas, 1983) (Olewnik et al., 2020).

2.4.4 Vastus medialis. The quadriceps femoris consists of several muscles which are rectus femoris, vastus lateralis, vastus intermedius, and vastus medialis. All these muscles help in extending the knee joint. Vastus medialis is activated in between the angle of 0 and 60 degrees. The extension of the knee, which is mainly part of the gait cycle, will take place with the help of the quadriceps muscles. When the knee is extended and some resistance is applied on it, it's possible to visualize that the quadriceps muscles contract. The femoral nerve branches out to give support to all quadriceps muscles. The knee joint is stabilized from contracting the quadriceps (Chang et al., 2015) (Jacob, 2008).

2.4.5 Biceps femoris. The short head and long head of the biceps femoris joint together as they go down the thigh. The muscle of the biceps femoris makes the flexion at the knee and extension at hip when they both contract. When the thigh is flexed, the biceps femoris help in tilting the pelvis backwards (Bakkum & Cramer, 2013). The biceps femoris, which is a double-jointed muscle has two roles in the human joints. It helps in extending the hip and flexing the knee which makes it a hard muscle to train in resistance workouts. It is also responsible for 80% hamstring strain injuries (Stevens et al., 2022).

2.5 Classification (Deep Learning and Machine Learning)

Deep learning (DL) and artificial intelligence are starting to become widely more popular now. The deep learning was made from artificial neural network (ANN). Deep learning has wide area of use. It is applied in healthcare, recognizing voice and images, and cybersecurity. Building a DL model is a complex and tricky task to do since it is based on a problem in real life. Real life problems are usually unpredictable and very different, which makes the DL model hard to build. DL can seem like a limited topic sometimes due to the few information about its core type of work. Deep network can be divided into several classes. The most common are the supervised and the unsupervised learning. Supervised learning is the discriminative type in which it can differentiate between several things while the unsupervised type is when the deep learning does not learn the data fed to it. In 2006, deep learning was first discovered by Hinton et.al and artificial neural network is the basis of how it works. Deep learning showed success when the training set is trained in a well manner, then it will provide a fine result with high accuracy.

The regular neural network process works as follows. This network consists of neurons, connected points, and each one will have an effect to produce the result. It consists of the following mathematical formula, in which there is an input, weight, bias, summation function from all the nodes, activation function, and the output that would be weighed and resulted.

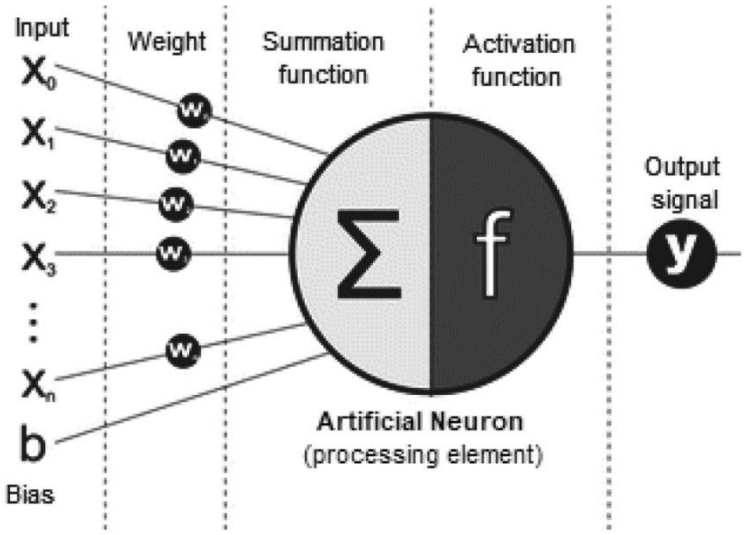

Figure 5. Neural Network illustration (Sarker, 2021).

To understand better deep learning, it is required to learn more about the categories of it. the supervised, unsupervised, and hybrid which is both supervised and unsupervised; these categories will help the user apply the most appropriate data that is needed according to his or her needs. Artificial intelligence (AI) or machine learning (ML) or deep learning are different concepts used that refer to a system that works intelligently. DL is part of machine learning and machine learning is a part under AI. Neural network in DL means that the data will go through several stages of processing to give an output therefore giving the required model.

There are several forms of data in which the DL learns from such as sequential data, image or 2D data, and tabular Data. Sequential data is a type of data in which is in sequence like audio or videos sections and time-series data such as an EMG recording. Image or 2D data are data points that are ordered or set in a rectangular array. Tabular data is a data that is organized in columns or rows. Every column in the tabular data should contain a naming for that field and a certain data type. The deep learning stages are acquiring the data and understanding it, preprocessing it, and finally a model is built, trained, validated and interpreted. The famous machine learning techniques that are commonly used are support vector machine (SVM), decision tree (DT), and K-nearest neighbor (KNN). The DL models are formed through using neural network types (Sarker, 2021).

Machine learning which is used in this research is used broadly in science, statistics, and engineering. It consists of several classes which are supervised learning (SL), unsupervised, semi-supervised, and reinforcement learning. The unsupervised machine learning extracts conclusions without prior labeling. The supervised machine learning will try to evaluate the inputs given and a targeted element. The supervised method is split into two sections which are classification and regression. The regression takes sequential values, but the classification-based system will need class labels (Soofi & Awan, 2017).

Machine learning's core is learning by itself, which involves enough statistical models and finding the similarities and differences between two variables to expect an outcome. The supervised method is a way that it teaches the system based on the fed data or training data. The regression analysis is in the supervised method which learns in a way that it learns the connection between dependent and independent variables.

Response variable is the outcome of the regression analysis, and the predictors are what is fed in the system therefore the system will find a relationship between them. Classification is a type of supervised learning in which it makes different classes and the data that is fed is separated and set to specific class. The classes here are predefined to set the classification and make it ready for analysis. Unsupervised learning is the process of training without prior knowledge of the output. It will create its own groups using the data set that is fed. The two algorithms used here are clustering and association analysis. Semi-supervised learning is another method which brings both supervised and unsupervised to increase accuracy. Reinforcement Learning (RL) uses machines and software programs to figure the needed behavior to reach the required outcome. RL depends on trial and error which means over time it becomes a better training model.

There are several types of classifications such as binary, multi-class, multi label, and imbalanced classification. SL algorithms classify different classes by using the concept of AI. The training is on the data fed and the outcome data is categorized according to its class. Regularly, the data that is fed into this algorithm is split into training and test set to receive the accuracy result. In the figure below are the SL classification algorithms. Zero R is a method in which it depends on target data and neglects other factors such as predictors. One R will make one rule for every predictor therefore it will classify it specifically to that predictor and but with low accuracy.

Other important prediction models are the decision tree and support vector machine. The decision tree works based on nodes. The nodes work in a way that when it passes from the root node to the child node then leads to the result which is the class label. Support vector machine (SVM) is one of the most used techniques. SVM works well when it comes to classifying different groups. The way it works is by setting up a high hyperplane that will act as border dividing two classes. This separating line is set in a way that it is far away from the classes in a n-dimensional space. Error is minimized and the supporting vectors is the data that is near to the decision surface (Alnuaimi & Albaldawi, 2024).

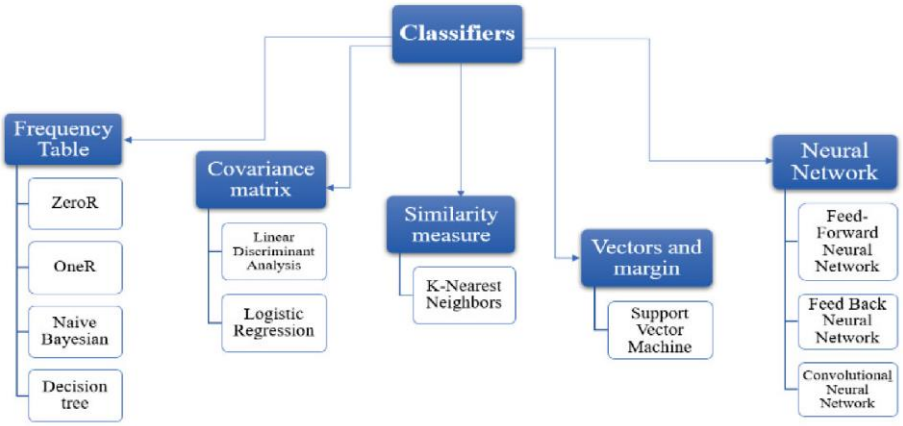

Figure 6. Classifier groups (Alnuaimi & Albaldawi, 2024).

Some studies have applied machine learning when walking on ground level. Classifications of stance or swing phases and predicting when the individual's foot touches the floor when walking normally (Morbidoni, Cucchiarelli, Fioretti, & Di Nardo, 2019). It has been a topic for a while to recognize the gait of a specific subject. The gait of the person can be used to identify that specific person since every individual has his or her own unique gait. To recognize a gait, it can be done in two methods. It can be either a model-based approach or model free approach. To build the first stage it is required to consider the body dimensions, location, and motion of the human body. The model free approach does not require a specific subject's data, and it can be done by feeding it 2D images. Some studies show a skeleton-based approach to identify gait. Although sEMG is difficult because of their complicated shape and uneven patterns, neural networks were used to understand and learn complex systems (Kim, Lee, & Shin, 2021). Machine learning tools that are mostly used by researchers are Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Deep Neural Networks (DNN) (Chakraborty & Chattaraj, 2023). Recognizing changes in gait cycles and analyzing it will help detect earlier the risk of having an disordered gait. Artificial intelligence machines such as SVM can detect normal gait (Begg, Palaniswami, & Owen, 2005).

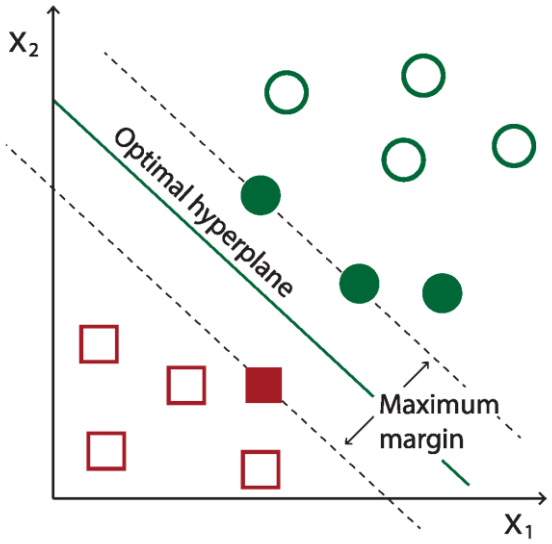

Figure 7. Illustration of SVM machine model concept and how it divides data. (Toledo-Pérez et al., 2019).

2.6 Different Speeds and Classification

Many research works have investigated the differences in speeds but in a narrow range which is currently underexplored, highlighting the need for more thorough analysis to observe speed-dependent changes. Muscle activation patterns can be learned and revealed in the gait of an individual. Every individual's movement and differences in EMG patterns can be differentiated. Machine learning methods can be used to make this differentiation. Support vector machine is one of the most used machine learning algorithms to classify and recognize the unique patterns of a certain individual. Myoelectrical activity in walking studies is mainly taken in the lower extremities. Most lower extremities muscles that are used are vastus medialis, rectus femoris, gastrocnemius medialis, tibialis anterior, soleus, and biceps femoris. These muscles can help give a signature for a specific individual. Recently, many studies have worked on the individual's movement signature from different biomechanical features (Hug et al., 2019).

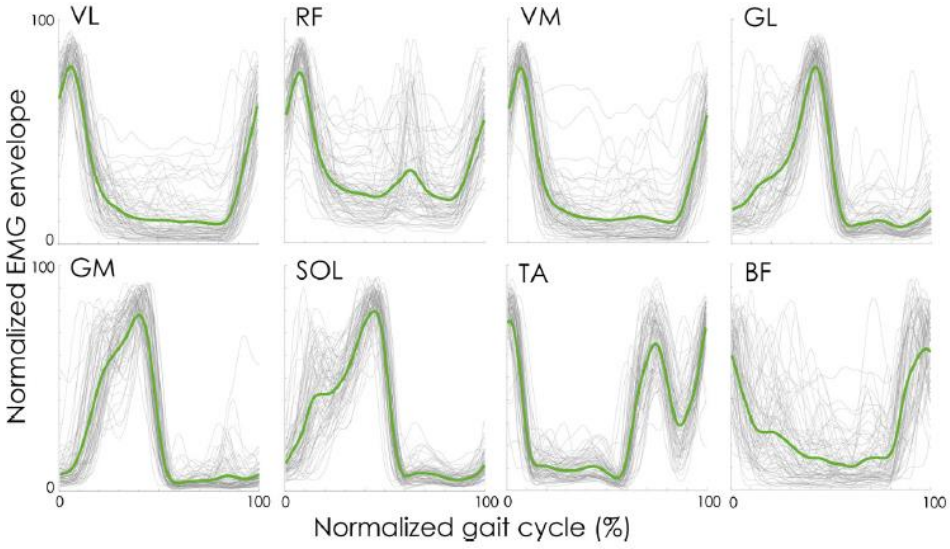

Figure 8. Different muscle activities during gait cycle (Hug et al., 2019).

Some other research worked on recognizing different gaits while walking on 5 different road conditions. Lower limb EMG signals can help identify different gait conditions. Easier rehabilitation protocols were formed for upper limbs. Although the lower limbs contain 54 muscles which add to the complexity of lower limb movements and pattern recognition. The knee flexion is performed with the help of biceps femoris and the gastrocnemius muscle. The ankle plantar and dorsal flexion is contracted by the help of the tibialis anterior, gastrocnemius, and soleus. A recognition method is feeding the muscles' signals into back propagation (BP) neural network to recognize the gait. Another research has used deep learning methods to identify and classify gait phases (Wang et al., 2021).

A deep learning approach for sEMG-based classification of the different phases in gait cycle. Classifying the stance and swing phases in addition to predicting the foot-floor-contact signals during an over-ground state. The classification process of this method was done over an eight shaped walk. This study used artificial neural network (ANN) method to classify the gait events. The presented method is to predict foot-floor-contact signal from EMG signals (Morbidoni et al., 2019d).

A gait signature recognition was also possible by using recurrent neural network. This type of model can detect individual differences in human gait. A broad range of speed was studied from very low speed to high running speed. In addition to this, the 3D kinematic and kinetic data was also extracted. In all speeds the gait

remained subject-specific showing great identification accuracy. slow speeds and high speeds affect in their own ways on the gait. Slow speeds may change the individual's coordination and momentum. It will need some alterations from the biomechanical point of view to stabilize the gait. Faster speeds need higher sensorimotor control and coordination. Support vector machine was utilized to recognize individual's gait (Winner et al., 2024).

Some studies researched the EMG and gait detection using a small set of speeds. Acquiring EMG signals from only two muscles which are Gastrocnemius and soleus. The subject walks at two speeds (fast and slow) which are selected by him. Root mean square amplitude levels and median frequency of EMG showed a remarkable difference. It involves the cooperative action of 28 muscles that control the jointed trunk and limbs to apply the correct force, counter gravity, and dive the body forward with minimum energy usage. Even a single motion like flexion and extension involves the activity of a group of muscles. The methodical part of this research involved the subject walking for three sets of data. All data was analyzed using RMS filters and median frequency filters (Kunju, 2016).

Machine learning methods were also used in a synergy study to classify different gait events. Muscle synergy features can be a potential and accurate parameter when classifying different gait events. Eight individuals participated in such study and four muscles in the lower extremities were measured while walking using EMG sensors. Several machine learning algorithms were used to classify the gait such as Decision tree (DT), k-nearest neighbors (KNN), support vector machine (SVM), and neural network (NN). Muscle synergy features showed better accuracy when compared to classifying other features in EMG. Identifying different gait phases is critical for accurate control of the lower limb assistive devices. The muscle synergy classification shows how the central nervous system synchronizes the activity of different muscle groups. SVM algorithm is mainly used in this study to classify gait phases. EMG sensors were attached to four different muscles which are tibialis anterior (TA), soleus (SOL), gastrocnemius lateralis (GL), and rectus femoris (RF). Treadmill walking was performed with a built-in force plate. A different adaptation method to the treadmill was applied here which is increasing the speed from a slow pace and decreasing from high speed to slow speed. A preferred speed by the subjects was selected which was 0.84 ±0.30 m/s. This preferred speed was conducted under a

one-minute walking duration. Data preparation includes some processing on the EMG data. Fourth-order Butterworth band-pass filter and cutoff frequency of 20-350Hz. The classification was applied on the data using a 10-fold cross-validation and dividing the training by using 90% of the data and 10% for the test set (Park et al., 2023).

Other research showed limited number of speeds and did not represent the true speed effect on gait. For example, five different speeds were investigated, and some other research has categorized three different states in which the person conducts: normal, slower than normal, faster than normal were the groups that identified the gait speed. The classification into three different speeds might be a negative result as these group's speeds might overlap. 285 steps were acquired from different subjects to analyze the difference between phase duration and speed. This research has focused on the gait phases and speed, but others have also defined a classification procedure for both gait and EMG.

In one research, the gait sub phase durations were studied in which 12 different walking speeds were observed. Walking speed from 0.6 to 1.7 m/s in healthy subjects were tested to separate different gait phases. Spatiotemporal gait parameters are considered the core calculation when it comes to gait performance. Musculoskeletal and neurological diseases can negatively affect gait performance. The spatiotemporal gait parameters are crucial for classifying and differentiating between healthy and non-healthy individuals. This will show the stages of neurological disorders which lead to gait disruptions. The gait speed can help in classification method to distinguish between healthy and rehabilitating subjects based on spatiotemporal parameters (Hebenstreit et al., 2015).

2.7 Rational for Features Selection

Signal pattern recognition is mainly dependent on feature selection that is mostly relative to EMG signals. Due to biological variability in different individuals even the same application of features on the data set will show variations in the features' output. Topological data analysis is one of the methods used to analyze the structure and shape of data. This feature grouping and analysis will allow similar features to be utilized allowing a better understanding and accurate classification. Mapper is a topology-based method that identifies relevant sub-groups unlike other classical clustering and dimensionality reduction methods such as principal

component analysis which it won't be able to identify. When it comes to the threshold in feature selection, each subject has a unique threshold that makes it difficult to unify one. In addition, the threshold that is set for each subject will change over time so it's not a feasible approach. Kamavuako made an optimum threshold that can be suitable for all subjects and datasets to balance between classification accuracy and generalizing the result.

A new strategy came up to be used to solve this problem by bringing up a simple topological chart. This new strategy will lead to feature selection in a principled way and keeping away non-useful features. The Mapper, a data analysis method, will classify the sub-groups and functional groups of the 58 advanced EMG features. Three different factors are in which the features' groups are separated, and they are maximum class separability, robustness, and complexity. The features based on the mapper method can give almost the same classification accuracy in comparison when using all features. The process of how it started is first preprocessing the EMG signals then recognizing the patterns. The components of pattern recognition are data segmentation, feature extraction, classification and controller. After analyzing the signal and are transformed into features, they are projected to a lower dimension to remove complexity and classify easier. These 58 state-of-the art features are represented in both the time and frequency domain. The method presented here will organize features in a clear chart which will help in developing a system for controlling artificial limbs (Phinyomark et al., 2017b).

The EMG features are classified into several groups: Time domain (TD), Frequency Domain (FD), and time-frequency domain. EMG features in other words can be represented in both the linear and nonlinear analysis (Phinyomark et al., 2013). Usually, the time frequency domain features need more processing and dropping their dimensionality. Features in the time domain are easier to process, implement, and classify. There are many time domain features such as integrated EMG (IEMG), mean absolute value, simple square integral, variance of EMG, absolute value of the temporal moment, root mean square, v-order, log detector, Willison amplitude, and average amplitude change. Integrated EMG shows the firing points of the EMG signals. Mean absolute value is one of the most used EMG signal analysis features and is like IEMG. Simple square integral is the energy of the EMG signal. RMS relates to

constant change in amplitude in accordance with the force when the muscle contracts and avoiding fatigue contraction (Phinyomark et al., 2012).

Table 1

Definitions And Functional Significance Of EMG Signal Features

Features	Explanation
MAV (Mean Absolute Value)	It is related to the magnitude of the EMG signal and provides a measure of overall muscle activity. Higher MAV values indicate stronger muscle contractions. (Jingwei et al., 2019).
RMS (Root Mean Square)	It is related to the power of the EMG signal and reflects the level of muscle activity. RMS is more sensitive than MAV to changes in signal amplitude. (Jingwei et al., 2019).
VAR (Variance of EMG)	Indicates the variability or fluctuations in the EMG signal. Higher VAR values suggest greater variability, often associated with changes in muscle activity patterns (Jingwei et al., 2019).
SSI (Simple Square Integral)	This feature is directly related to the energy of the EMG signal and serves as an indicator of its intensity. (Jingwei et al., 2019).
EMAV (Enhanced Mean Absolute Value)	EMAV is a modified version of MAV, designed to provide a more robust measure of a signal's magnitude by emphasizing its most informative portions. EMAV assigns higher weight (0.75) to samples in the middle region of the signal (20%-80%) and lower weight (0.50) to samples at the beginning and end, where less valuable information is presumed to reside (Jingwei et al., 2019).

Table 1 (cont.d)

EWL (Enhanced Wavelength)	EWL is an adaptation of the WL feature, designed to provide a more accurate representation of the waveform's complexity by emphasizing its middle portion. Like EMAV, EWL assigns greater weight to the middle of the signal while WL is calculated as the cumulative length of the waveform, potentially reflecting changes in motor unit firing rates (Jingwei et al., 2019).
MMAV (Modified Mean Absolute Value)	MMAV is an extension of the MAV that improves the estimation of a signal's magnitude by prioritizing its central portion. Using a weighted window function, MMAV assigns a weight of 1 to the middle 50% of the signal and 0.5 to the remaining samples, focusing on the section most representative of the signal's main characteristics, like EMAV (Jingwei et al., 2019).
TM (Temporal Moment)	Represents the distribution of the EMG signal's energy over time. Higher-order moments, like the 3rd order, are sensitive to the shape of the signal's waveform. Changes in TM may reflect alterations in motor unit recruitment patterns or firing rates (Phinyomark et al., 2017).
SD (Standard Deviation):	Reflects the overall variability or dispersion of the EMG signal's amplitude (Phinyomark et al., 2017).
ME (Average Energy)	Directly related to the intensity of muscle contraction. A higher average energy indicates a stronger contraction (Phinyomark et al., 2017).
VO (V-Order)	Sensitive to larger deviations in EMG signal amplitude, particularly with higher values of 'v' capturing aspects of the signal's magnitude and variability (Phinyomark et al., 2017).

Table 1 (cont.d)

AR (Auto Regressive Model)	The coefficient in this feature captures the temporal dependencies within the EMG signal, representing it as a linear combination of its past values. Changes in these coefficients may reflect alterations in motor unit recruitment or firing rate (Phinyomark et al., 2017).
COV (Coefficient of Variation)	It is a measure of the relative variability of a signal (Phinyomark et al., 2017).
LDAMV (Log Difference Absolute Mean Value)	Represents the average change in signal amplitude between consecutive samples. That LDAMV is sensitive to amplitude changes, on a compressed scale, potentially useful for detecting subtle changes in muscle activity. The use of differencing indicates a focus on capturing dynamic changes in the signal rather than its absolute magnitude (Phinyomark et al., 2017).

2.8 Prosthetics and Lower Limbs

Recent research has shown the potential of electromyography (EMG) therefore microprocessor-controlled lower limb prostheses (MLLPs) are developing to be designed to be integrated in the EMG sensors. This will give a more responsive effect on prosthetics and develop their ability to achieve a fluid motion. The MLLPs are a type of prosthetic devices that can replicate the human's biological movement. Real-time responsive control is signaled by the sensors connected on the individual. The use of these devices is not controlled by the user, but the device acts on how the sensors recorded that information. When users find that the device is lacking control, they start to decline these kinds of devices which decreases their use. Poor mobility is one of the main factors that makes the individual abandon the prosthesis.

Prosthetics are trending in restoring the human locomotion. Through the neural network, the user can directly interact with the device. Neuro controllers can interpret information from the central or peripheral nervous system. EMG sensors can be used

to detect and predict motion intentions from the neural signals in which the muscles serve as a final amplifier of motor signals. Volitional and rhythmical are two types of neuro control commands within the human motion biomechanics. Rhythmic locomotion is the repetitive motion of the limbs like walking or running. This does not involve conscious interference of the subject but is the involvement of specific neural networks in which the sensory-motor reflexes are activated. Volitional movements on the other hand will require motion planning and motor control such as knee flexing (Cimolato et al., 2022).

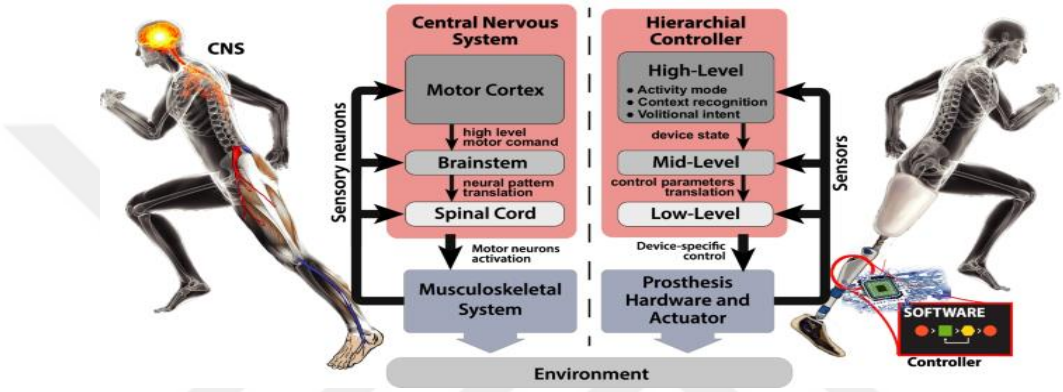

Figure 9. Human limb vs prosthetic (Cimolato et al., 2022).

The extracted features' classification methods will anticipate movement intention while the patient uses prosthetics. This classification method simultaneously works while the individual is moving. The decoding process of the EMG is as follows: pre-processing, feature extraction, classifying muscle activities, and post-processing. Once the classification step is conducted, which will be using the features, the intended movement of the participant will enable to move the prosthesis. The classification will be a phase-dependent one in which certain aspects within the gait cycle implement different stages within the gait cycle (Ahkami et al., 2023).

Chapter 3
Materials and Methods

3.1 Research Design

The research design of this thesis focuses on classifying different speeds using a treadmill. There is several sensor equipment used which are EMG and Movella. The programs in which these two devices work acquire the data. The type of data acquired is converted for specific use. The aim of all this is to enhance and present a machine learning model that will better identify muscle activity patterns with a specific speed. In addition, this will provide a new rehabilitation protocol to be used in clinics. A short video of every subject's walking pattern was taken to ensure accurate walking.

3.2 Participants/Working Group

25 subjects have been taken for this test. The tests are treadmill walking with a wide range of speeds. The subjects' ages are 24±5 and weight of 72.3 ± 12 and height of 175.5 ± 10. Mostly males were taken for this project due to the physiological and physical attributes. Almost all subjects that were taken for this experiment are active individuals. The subjects are healthy which are capable of walking on treadmill on all speeds smoothly and correctly after undergoing an adaptation period.

Exclusion criteria

- Chronic heart problems
- Lower limb injuries
- Abnormal gait
- Any type of medication that might hinder muscle movement
- Under 60 kg
- Above 85 kg
- Neurological disorders

3.3 Data Collection

3.3.1 Equipment.

3.3.1.1 EMG. EMG is the device used to record muscle signals that are produced from the human body. Specifically, surface EMG (Figure 10) is used in this protocol to prevent from a side effects of the invasive EMG which can be uncomfortable for the subject. The EMG device (Trigno, Delsys, United States) is a wireless biofeedback system that can record at a frequency of around 2000 Hz. These types of sensors are used for their noise-resistant electrode structure. These EMG sensors have built-in inertial measurement unit (IMU) and a triaxial accelerometer. The available channel of this device is 16 channels in which 16 different muscles can be attached at once. The wireless transmission reach of the sensor is around 20 m. The bandwidth is selectable between 10-850 Hz and 20-450 Hz. EMG sensors contain lithium polymer battery, and its working hours are between 4 and 8 hours depending on the usage. A trigger model device of Delsys was also used along with the EMG device. This trigger synchronization device is also used to synchronize between the EMG and Movella (Xsens) devices (Reeves et al., 2020) (Halabi et al., 2019).

Figure 10. Trigno wireless EMG sensors and measurement system (Choi & Lee, 2015).

3.3.1.2 Movella (Xsens) MVN Awinda. Xsens MVN Biomech (Amsterdam, Netherlands) are inertial unit high motion capture sensors which record different parameters. Three-dimensional acceleration, angular velocity, and displacement data are all different parameters which are recorded by the Xsens MVN device. It records all these parameters in real time while the subject is conducting the motion. This technology can acquire accurate data for biomechanical analysis. The Xsens device consists of 17 inertial sensors which are placed on different places on the body in which it records 22 different joints. The sampling frequency of this device is 60 Hz. The sensors are placed according to the guidelines of the manufacturer which is placing them on different joints in the body. They are placed on the feet, thighs, wrists, forearms, upper arms, chest, lumbar spine, head, and below the knees. Calibration of N or T pose is performed after placing all sensors to ascertain accurate data collection throughout the experiment (Yao et al., 2024).

3.3.1.3 Treadmill. The treadmill brand used here is a Ultima Dc-1500 as shown in figure 11. It has a walking area of 126 cm * 43 cm. The maximum load that it can withhold is 110 kg. The speed ranges from 0.8 till 14 km/hr. It has a shock absorber system that can withhold the body's pressure while walking and especially while running.

Figure 11. Ultima Dc-1500 Treadmill model.

3.3.2 Data collection procedure.

3.3.2.1 EMG. SENIAM.org website is first observed to know the location of the muscles to attach the EMG sensors. Prior to attaching the sEMG electrodes, the skin is cleaned with alcohol and shaved for better data acquisition. Ten sensors were attached on the skin which were on both legs, specifically on the targeted muscles; they are tibialis anterior, gastrocnemius medialis, soleus, vastus medialis, and biceps femoris. The image below illustrates the attachment of the sensors on the lower extremities. The subject will contract the muscle to show the muscle area of interest. The area of the sensor contact is cleaned before attachment. To ensure better quality signal and less noise, several stretch bands were attached around the EMG sensors to further tighten the attaching point; there was not much pressure applied to ensure that muscle would contract freely. Finally, Delsys (EMG) acquisition program is used to record data.

Figure 12. A. Movella MVN sensors and EMG sensors are shown here. B. Lower extremity muscles are shown here, which are Vastus Medialis and Tibialis Anterior. C. Posterior extremity muscles which are Biceps Femoris, Gastrocnemius Medialis, and Soleus.

3.3.2.2 Movella (Xsens) MVN Awinda. Prior to data collection, the subject under test should wear the Awinda vest for the purpose of attaching sensors more firmly and proper sensor placement. The tight and firm to the body vest was worn by the subject to ensure stability of the sensors while moving. There are 15 different Xsens sensors attached on the human body on different joints. The Xsens MVN manual was used as a guide to ensure correct sensor placement. Different measurements of the body were taken including: full height, shoulder height, elbow span, wrist span, hip height, hip width, knee height, ankle height, foot length, and sole thickness. The overall sole thickness is in the range of 2-3 cm based on different shoe brands. A measurement confirmation was done to make sure of the sole thickness. All subjects were measured twice when it comes to full height which was with shoe and barefoot. All subjects wore running shoes for comfort during the experiment.

Prior to starting to record data, the subject will perform calibration which requires the individual to stand in the N- pose position. This position is putting hands on the thigh and keeping head straight. A calibration period will start in which the subject walks in which it takes about half a minute.

Figure 13. Subject conducting walking experiments on the treadmill.

3.4 Experiment's Protocol

Following the calibration of the Xsens sensors in accordance with the body joints, the subject is required to ambulate on the treadmill. Subjects undergo an adaptation period of 5 minutes. He is required to look forward while keeping attention to the walking environment. The speeds are pre-selected at km/hr which are 2.9, 3.2, 3.6, 4, 4.3, 4.7, 5, 5.4, 5.8, 6.1, 6.5, 6.8, and 7.2. Speed increments of 0.3 or 0.4 interchangeably were applied between each speed. Subjects will walk each at an average of one minute and a half or at least 50 heel strikes. There is a resting step of 5 minutes every 3 speeds to not reach fatigue. The whole experiment's duration is around 45 minutes.

3.5 Data Analysis

Many steps were taken to ensure accurate data analysis to get the most optimal accuracy results for classification. EMG preprocessing steps were conducted for more reliability to ensure good quality signal and reduce noise. One of the most important pre-processing tools is to remove power lines' electrical interference. A 50 Hz notch filter is applied to remove this kind of noise which disturbs EMG signal. Bandpass filtering was performed by setting a range of 20-500 Hz. A 4^{th} order Butterworth filter works with this type of data and the range set will focus on the wanted signal. Winsorization then takes place to fine-tune the signal and deals with extreme values. An interquartile method is also used to deal with outliers and prevent skewness of data.

Xsens data was used to segment the gaits by using defined heel strikes. EMG data along with the Xsens data were synchronized to further collect parameters and features from the data. The following features used for classification are as follows:

3.5.1 Features formulas.

EMG signal vector x[n], where n is the index of the data sample, and n is the total number of samples.

- **Mean Absolute Value (MAV)**

$$MAV = \frac{1}{N}\sum_{n=1}^{N}|x[n]|$$

- **Root Mean Square (RMS)**

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{n} x[n]^2}$$

- **Temporal Moment (TM)**

$$TM_k = \frac{1}{N}\sum_{n=1}^{N} n^k \times x[n]$$

Temporal moments TM_k quantify the signal's temporal distribution for a given order k. For k=1, it is the first moment.

- **Variance (VAR)**

$$VAR = \frac{1}{N}\sum_{n=1}^{N} (x[n] - \mu)^2$$

μ is the mean of $x[n]$

- **Standard Deviation (SD)**

$$SD = \sqrt{VAR}$$

- **Average Energy (ME)**

$$ME = \frac{1}{N}\sum_{n=1}^{N} x[n]^2$$

- **V-Order Statistic (VO)**

$$VO(V) = \left(\frac{1}{N}\sum_{n=1}^{N} |x[n]|^v\right)^{\frac{1}{v}}$$

- **Enhanced Mean Absolute Value (EMAV)**

$$EMAV = \frac{1}{N}\sum_{n=1}^{N} \omega[n] \times |x[n]|$$

$\omega[n]$ is the weight function.

- **Simple Square Integral (SSI)**

$$SSI = \sum_{n=1}^{N} x[n]^2$$

- **Autoregressive Model Coefficients (AR)**

$$x[n] = \sum_{k=1}^{P} a_k \times x[n-k] + e[n]$$

a_k are AR coefficients, and $e[n]$ is noise.

- **Coefficient of Variation (COV)**

$$COV = \frac{SD}{\mu}$$

- **Enhanced Wavelength (EWL)**

$$EWL = \sum_{n=1}^{N-1} \omega[n] \times |x[n+1] - x[n]|$$

$\omega[n]$ is the weight function

- **Log Difference Absolute Mean Value (LDAMV)**

$$LDAMV = log\left(\frac{1}{N}\sum_{n=2}^{N} |x[n] - x[n-1]|\right)$$

- **Modified Mean Absolute Value (MMAV)**

$$MMAV = \left(\frac{1}{L}\right) \sum_{i}^{1'} w_i |x_i|$$

The weight of the function is defined as follows:

$$w_i = \{1, if\ 0.25L \leq i \leq 0.75L\ \ 0.5, otherwise$$

- **AAC (Average Amplitude Change)**

$$AAC = \left(\frac{1}{(N-1)}\right) * \Sigma_i^{=1N-1} |x_i^{+1} - x_i|$$

- **WL (Waveform Length)**

$$WL = \Sigma_i^{N-1} |x_i{}^{+1} - x_i|$$

- **MFL (Maximum Fractal Length)**

$$L(k) = \left(\frac{1}{k}\right) * \Sigma_m{}^k L_m(k)$$

- **KURT (Kurtosis)**

$$KURT = \frac{\left[\left(\frac{1}{N}\right) * \Sigma_i{}^{=1N} (x_i - \mu)^4\right]}{\left[\left(\frac{1}{N}\right) * \Sigma_i{}^{=1N} (x_i - \mu)^2\right]^2}$$

- **LD (Log Detector)**

$$LD = exp\ exp\left(\left(\frac{1}{N}\right) * \Sigma_i{}^N\ ln\ ln\ (|x_i|)\right)$$

- **SKEW (Skewness)**

$$SKEW = \frac{\left[\left(\frac{1}{N}\right) * \Sigma_i{}^{=1N} (x_i - \mu)^3\right]}{\left[\left(\frac{1}{N}\right) * \Sigma_i{}^{=1N} (x_i - \mu)^2\right]^3 /^2}$$

- **DVARV (Difference Variance Value)**

$$DVARV = \left(\frac{1}{(N-1)}\right) * \Sigma_i{}^{=1N-1} (\Delta x_i - \Delta\mu)^2$$

where:

- $\Delta x_i = x_{i+1} - x_i$
- $\Delta\mu = (1 / (N-1)) * \Sigma_{i=1}{}^{N-1} \Delta x_i$

LDASDV (Log Difference Absolute Standard Deviation Value)

$$LDASDV = exp\ exp\left(\left(\frac{1}{N-1}\right) * \Sigma_i{}^{N-1}\ ln\ ln\ (|x_i{}^{+1} - x_i|)\right)$$

IEMG (Integrated EMG)

$$IEMG = \Sigma_i^N |x_i|$$

- **LCOV (Log Coefficient of Variation)**

$$LCOV = ln\ ln \left(\frac{\sigma}{|\mu|}\right)$$

where

`σ` is the standard deviation of the EMG signal: $\sigma = sqrt((1/N) * \Sigma_{i=1}^{N} (x_i - \mu)^2)$

`μ` is the mean of the EMG signal.

- **LTKEO (Log Teager Kaiser Energy Operator)**

$$LTKEO = \left(\frac{1}{N}\right) * \Sigma_i^{N-1}\ ln\ ln\ (x_i^2 - x_i^{-1} * x_i^{+1})$$

- **FZC (New Zero Crossing)**

$$FZC = \Sigma_i^{=1N-1} 1[(sgn(x_i * x_i^{+1}) < 0) \wedge (|x_i - x_{i+1}| \geq threshold)]$$

where:

`sgn(x)` is the sign function:

`threshold` is a predefined value.

- **VAR (Variance)**

$$VAR = \left(\frac{1}{N}\right) * \Sigma_i^N (x_i - \mu)^2$$

3.5.2 Spearman correlation. Spearman correlation was used to find the most redundant features to exclude them. 13 features of 27 were redundant therefore they were not used for classification.

3.5.3 Classification of muscle activity patterns. Classification was conducted in the MATLAB program in which several methods were investigated to find the most efficient and accurate approach in which several models were evaluated. Mainly two methods were classified using Classification Learner which are time-series and feature based analysis. The training set used for this classification is 80% and the testing dataset is 20%. All models were tested to find the most efficient and accurate performance.

3.5.3.1 Time series analysis.

3.5.3.1.1 Whole data acquisition of right and left leg separately. The whole data was taken after down-sampling to 100 frames. The data of five muscles on the right leg and five muscles on the left leg were fed to the classification learner to validate and test the data. The left leg's data was taken as well to classify but each leg was separated into two different arrays.

3.5.3.1.2 Skipping speeds acquisition. Each leg was taken as a separate data to classify. Two methods were applied here to classify the data. The first method was skipping speeds that are in the odd set which is named here subset 1. The second method is skipping speeds that are in an even set which is named subset 2. The effect of this is to observe the change in the model's accuracy when the speeds have been altered. When the difference between successive speeds has increased by double this will show a more accurate and specific classification.

3.5.3.1.3 Skipping frames acquisition. The same protocol was followed here by taking each leg as two separate arrays. The skipping frames' method was applied by implementing different approaches. The frames were skipped once by dividing the total frames by 2, 4, 5, 10, or 20 and the remaining used for classification.

3.5.3.1.4 Gait segmentation. This method is dividing the frames of the data into 60 and 40 frames. These frames illustrate the gait cycles. This method will test which accuracy was higher either the stance or swing phase.

3.5.3.2 Feature based analysis.

3.5.3.2.1 Independent analysis of right and left leg motion data. Fourteen features are used to analyze the data on the right leg's data and on the left leg's data. Each leg is done independently which is similar to the method done in the time-series' method. These features will work row wise meaning taking the gait cycle in that row and specifically for that muscle.

3.5.3.2.2 Skipping speeds acquisition. The features along with speed skipping are applied here. The speeds will either start from the first one and skip one by one which is also named here subset 1 or from the second speed and skip one by one which is also named here subset 2. Both methods will have features applied to them. The reasons why this method was also applied here are according to what was mentioned previously.

3.5.3.3 Performance metrics for classification.

3.5.3.3.1 Accuracy. The accuracy of the classification is measured by an equation which is the positive predictive value (PPV). High result of this indicates that the algorithm worked very accurately to identify the values given to it.

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

3.2.1.1.1 Precision. Precision can be calculated by the number of positive encounters that are correctly classified by the model. The equation used in precision for the algorithms in this research is true positive rate (TPR). False negative rate (FNR) which is calculated as the accuracy ratio of positive samples that were classified incorrectly.

Chapter 4

Results

Different confusion matrices presented to show the accuracies of the speeds and how well they were classified. Results are for both the time-series and feature based analysis. Whole frames of right leg data, whole frames of left leg data, speed selection: alternating between even and odd patterns, skipping frames by leaving 5,10,20,25, and 50 frames. To ensure experimental validity, both legs were treated equally. Feature based analysis was treated equally and additionally a spearman correlation was tested to remove redundant features.

4.1 Spearman Correlation for Feature Selection

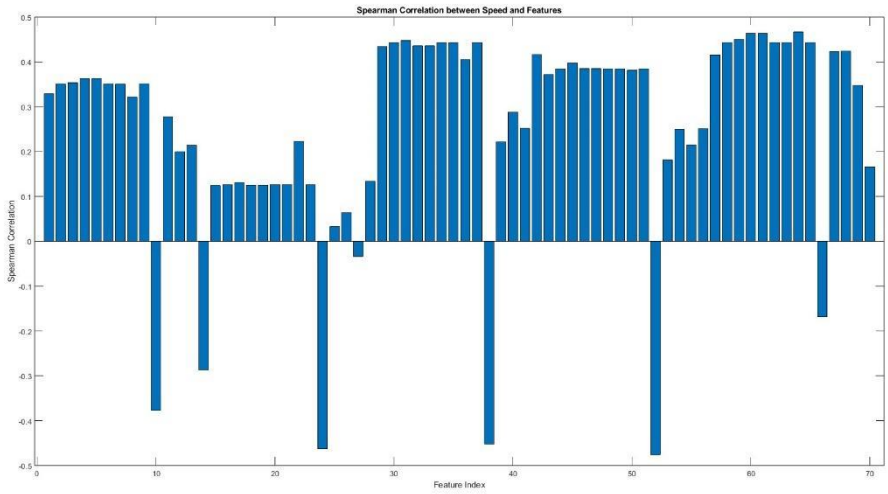

Figure 14. Right leg Spearman correlation.

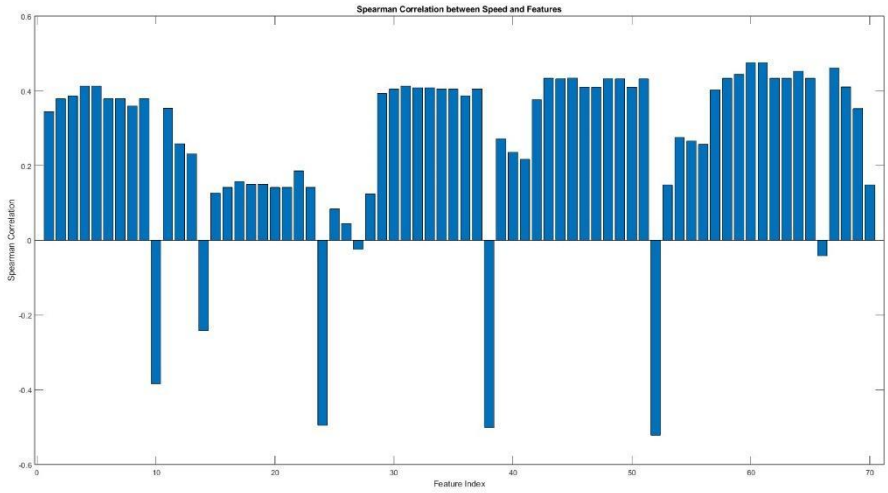

Figure 15. Left leg Spearman correlation.

4.2 Time Series Manipulations

4.2.1 Independent analysis of right and left leg motion data.

4.2.1.1 Right leg analysis.

4.2.1.1.1 Complete time-series data. These confusion matrices illustrate taking the complete time series data and classifying it. The test confusion matrix had a higher percentage than the training confusion matrix (Figure 16 and Figure 17).

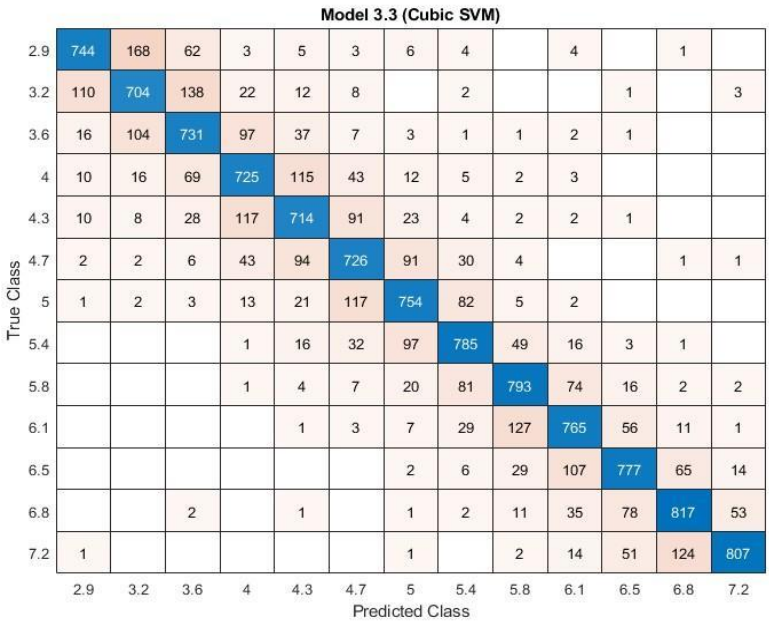

Figure 16. Training confusion matrix for right leg using complete time-series.

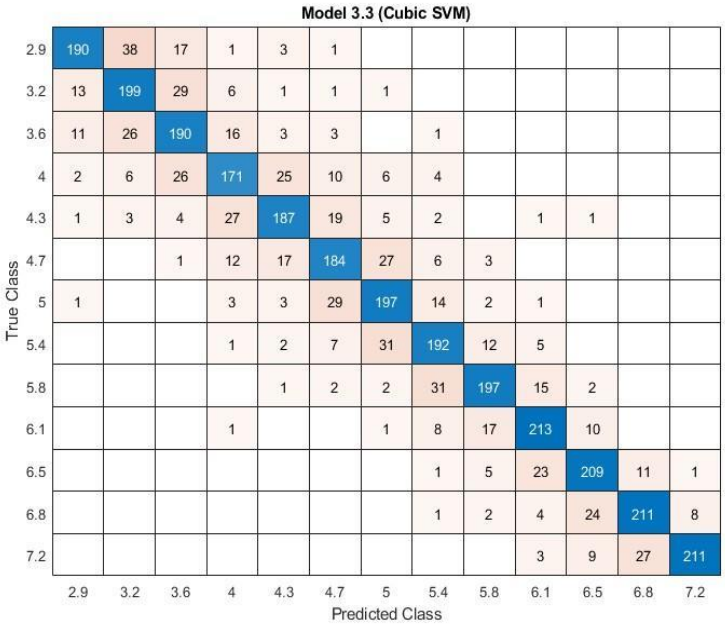

Figure 17. Test confusion matrix for right leg using complete time-series.

The statistical results show the accuracy and precision of the classification. It shows very high precision in detecting the speeds. The TPR plot shows high sensitivity in confirming the detection of speeds.

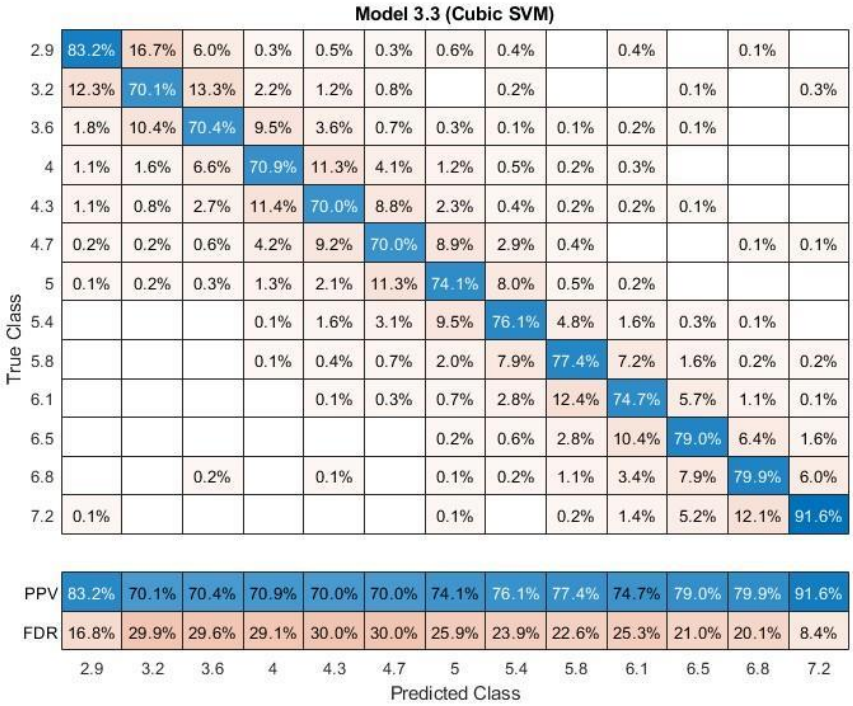

Figure 18. PPV-FDR training confusion matrix for right leg using complete time-series.

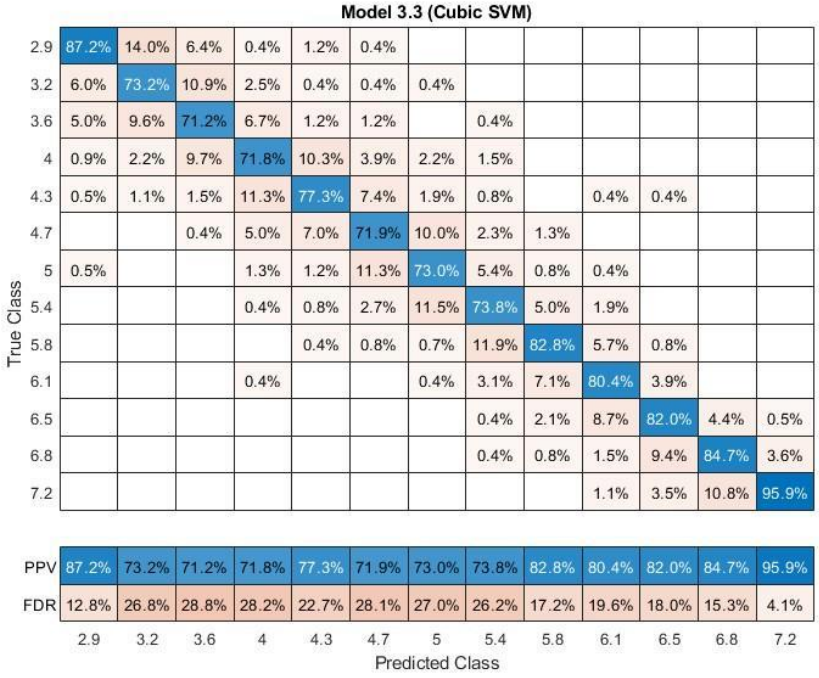

Figure 19. PPV-FDR test confusion matrix for right leg using complete time-series.

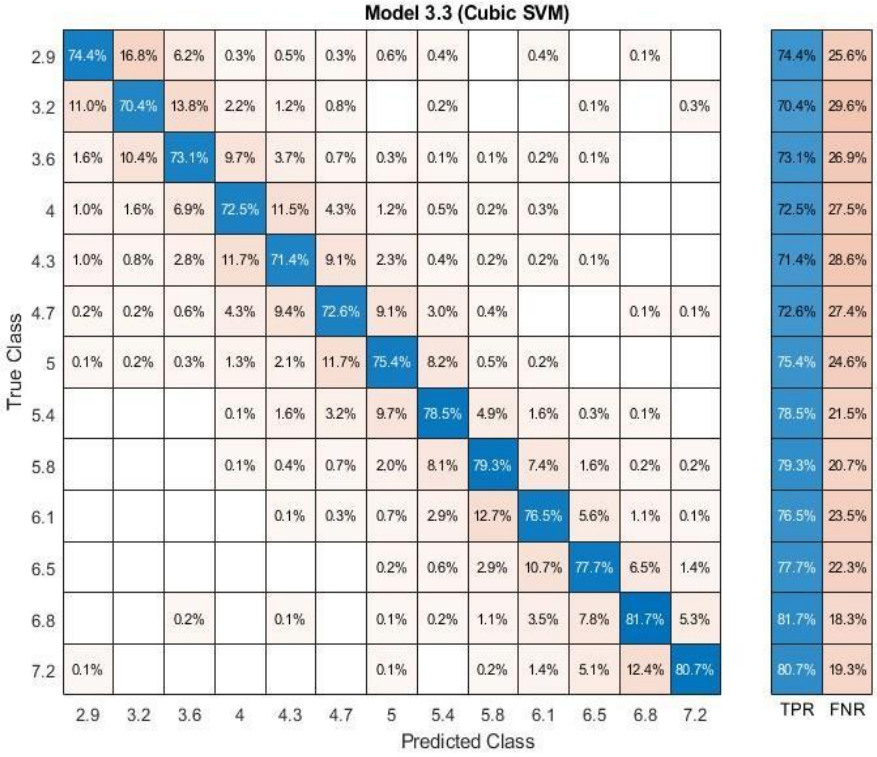

Figure 20. TPR-FNR training confusion matrix for right leg using complete time-series.

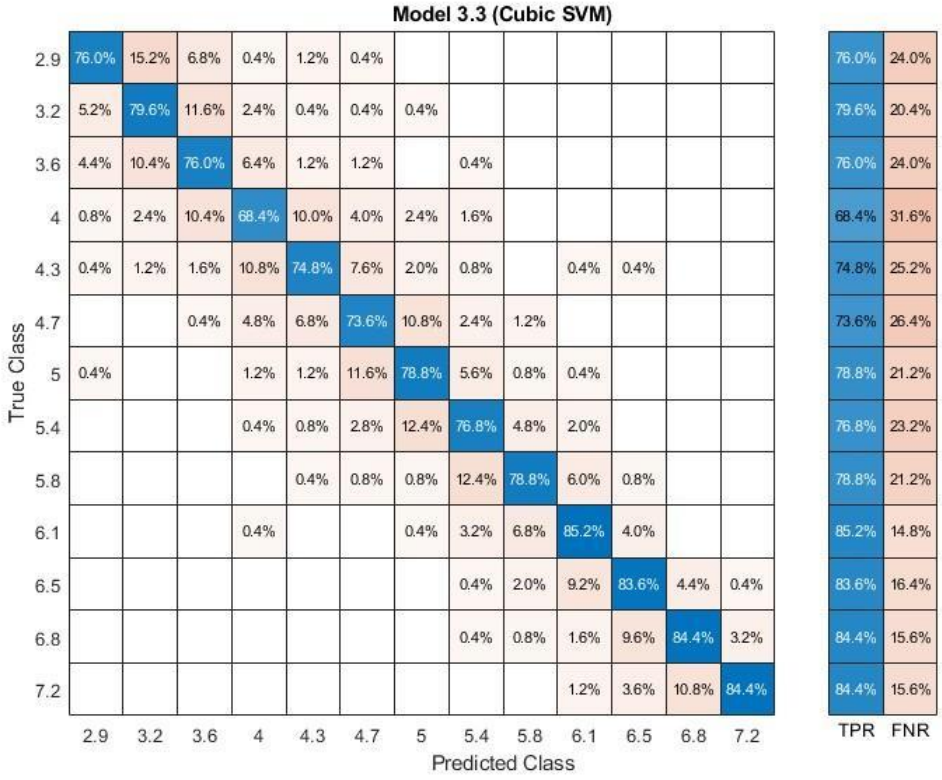

Figure 21. TPR-FNR test confusion matrix for right leg using complete time-series.

Highest accuracy is in the cubic SVM model. It has very high accuracy in the training and even increased in the test set.

Table 2

Models' Accuracy For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	75.70769	78.49231
3.2	SVM	71.68462	74.55385
3.5	SVM	70.17692	73.07692
4.6	KNN	68.54615	70.83077
5.3	Neural Network	67.20769	68.49231
4.4	KNN	66.93846	69.69231
4.2	KNN	66.73846	69.35385
4.1	KNN	65.80769	68.67692
4.5	KNN	62.03846	64.64615
5.2	Neural Network	51.92308	56.49231
3.4	SVM	46.3	48.4
4.3	KNN	46.09231	52.83077
3.1	SVM	45.66923	46.36923

Table 2 (cont.)

5.4	Neural Network	44.10769	43.38462
5.5	Neural Network	42.59231	47.50769
5.1	Neural Network	42.45385	44.03077
3.6	SVM	35.28462	36.98462
1	Tree	28.24615	29.66154
2.1	Tree	28.24615	29.66154
2.2	Tree	22.48462	22.61538
2.3	Tree	17.94615	17.44615

4.2.1.1.2 Skipped speeds odd pattern (subset 1). The skipped speeds method in the figures below are even higher classified speed-specific in comparison to when performing the classification method on the percentile data.

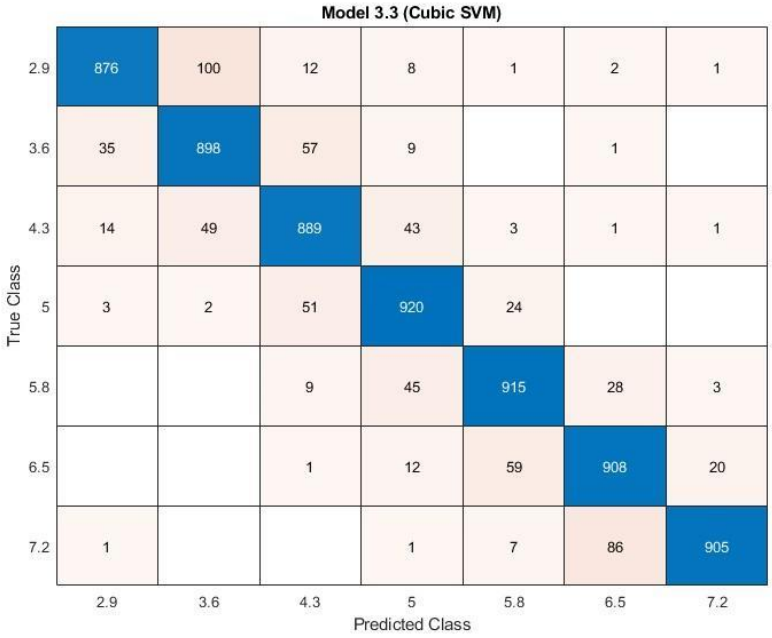

Figure 22. Training confusion matrix of subset 1 for right leg using complete time-series.

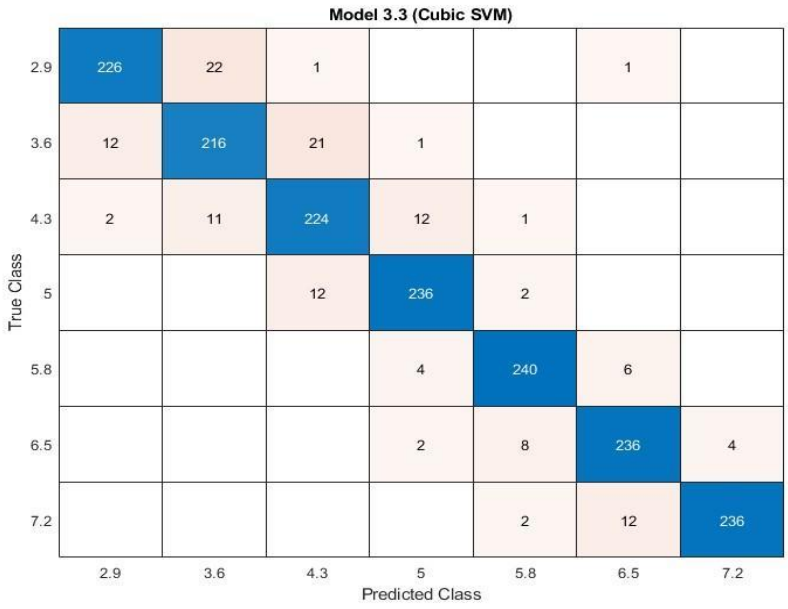

Figure 23. Test confusion matrix of subset 1 for right leg using complete time-series.

The two statistical methods in the subset 1 skipped speeds method show very high precision and sensitivity. They are both with high percentages that can confirm the results of the confusion matrices.

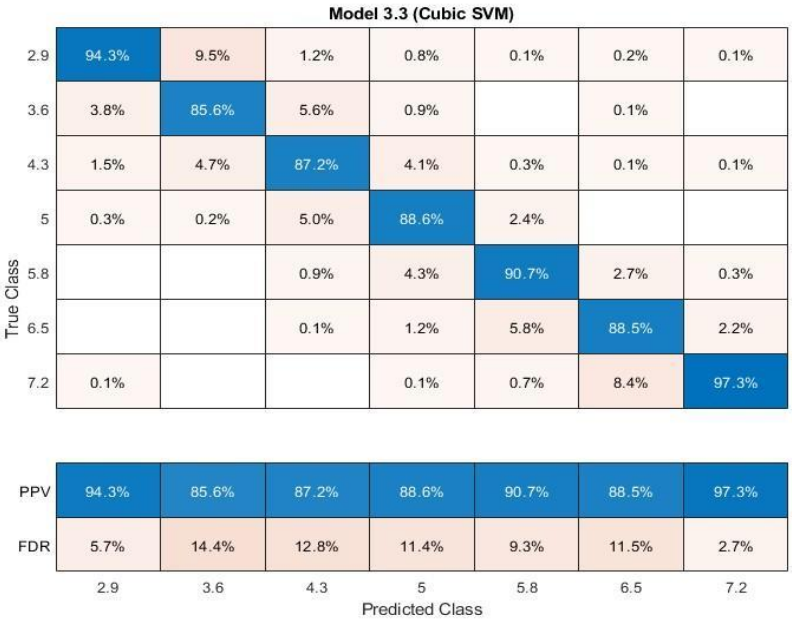

Figure 24. PPV-FDR training confusion matrix of subset 1 for right leg using complete time-series.

Figure 25. PPV-FDR test confusion matrix of subset 1 for right leg using complete time-series.

Figure 26. TPR-FNR training confusion matrix of subset 1 for right leg using complete time-series.

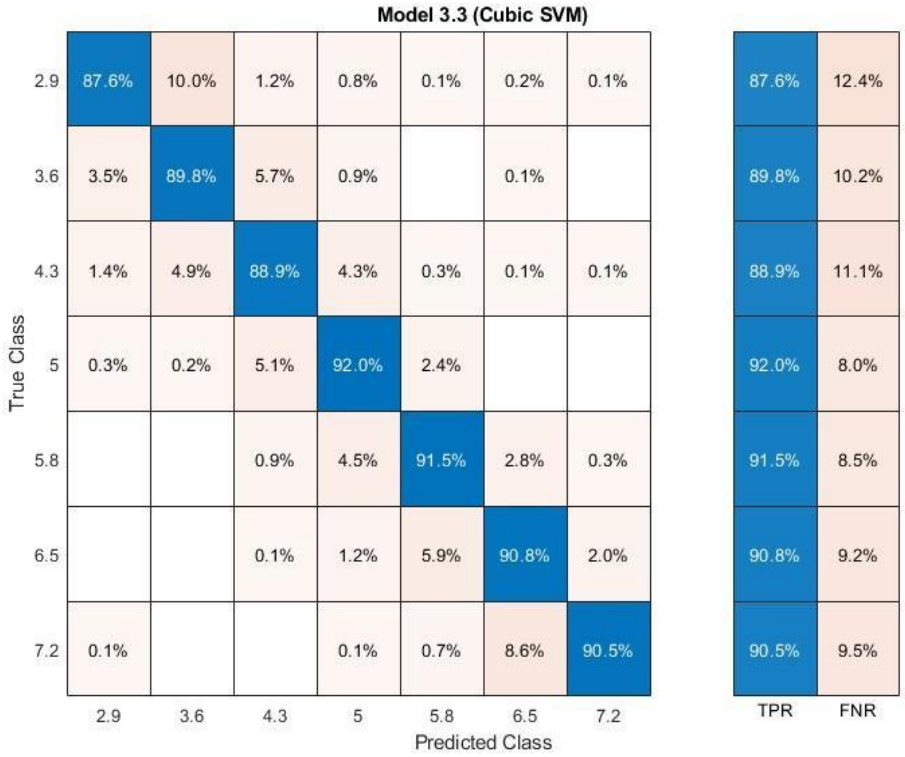

Figure 27. TPR-FNR test confusion matrix of subset 1 for right leg using complete time-series.

The cubic SVM model demonstrated the highest accuracy, performing exceptionally well on the training set and further improving on the test set.

Table 3

Models' Accuracy Of Subset 1 For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	90.15714	92.22857
3.5	SVM	87.71429	88.68571
3.2	SVM	87.5	89.42857
4.6	KNN	86.18571	87.77143
5.3	Neural Network	85.9	87.82857
4.1	KNN	85	87.31429
4.4	KNN	84.94286	86.57143
4.2	KNN	84.52857	86.74286
4.5	KNN	80.68571	83.14286
5.2	Neural Network	80.24286	81.37143
5.4	Neural Network	69.74286	66.62857
5.5	Neural Network	69.64286	67.82857

Table 3 (cont.)

5. 1	Neural Network	67.75714	67.48571
3.1	SVM	66.92857	68.74286
3.4	SVM	61.68571	62.45714
3.6	SVM	56.01429	58.22857
4.3	KNN	53.92857	63.2
2.1	Tree	49.8	51.71429
2.2	Tree	40.61429	42.11429
2.3	Tree	33.07143	33.14286

4.2.1.1.3 Skipped speeds even pattern (subset 2). In the following figures, the confusion matrices illustrate the classification results on the training and test sets for skipped speeds in even pattern, revealing differences in their performance when applied to the complete time series data.

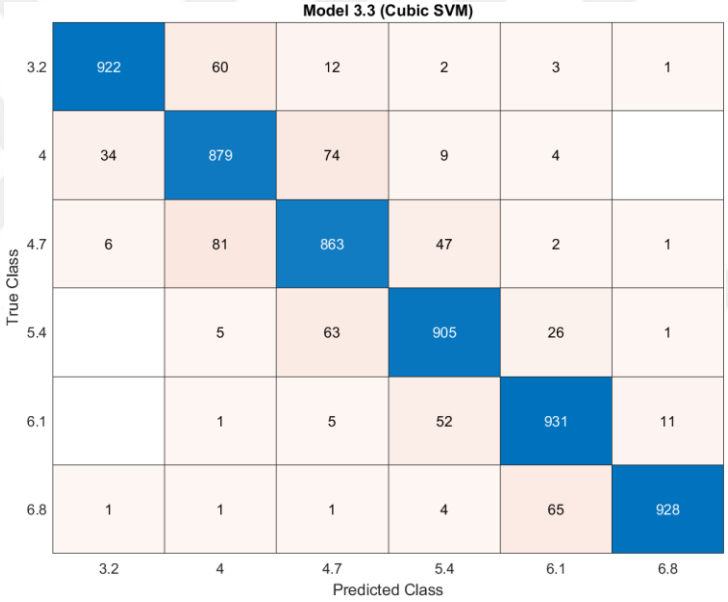

Figure 28. Training confusion matrix of subset 2 for right leg using complete time-series.

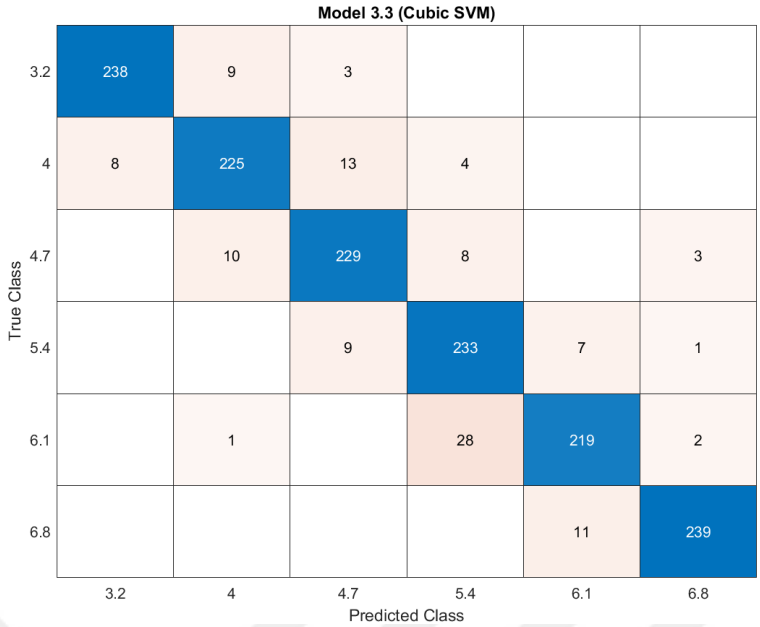

Figure 29. Test confusion matrix of subset 2 for right leg using complete time-series.

Analysis of the confusion matrices reveals high precision in classifying speeds. Furthermore, the TPR plot confirms the high sensitivity of the classification method in accurately detecting speeds as well in the even pattern.

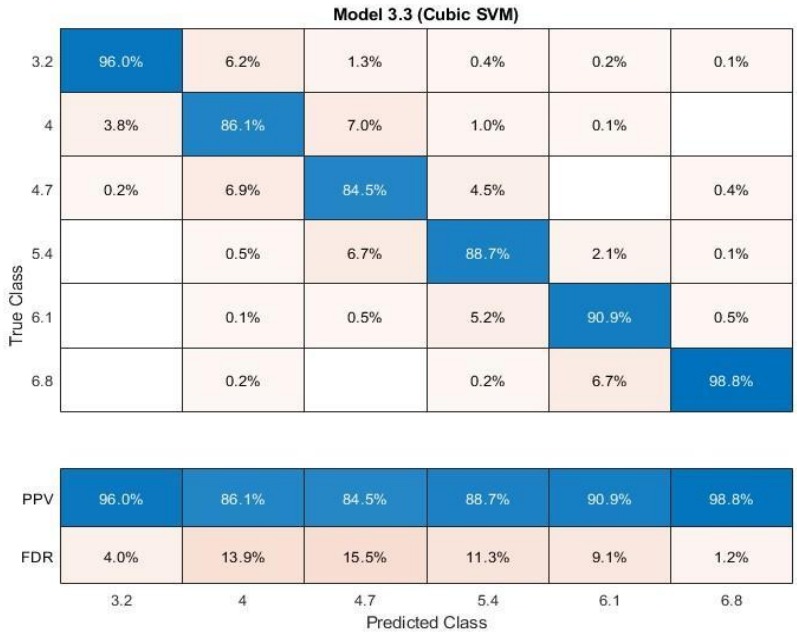

Figure 30. PPV-FDR training confusion matrix of subset 2 for right leg using complete time-series.

Figure 31. PPV-FDR test confusion matrix of subset 2 for right leg using complete time-series.

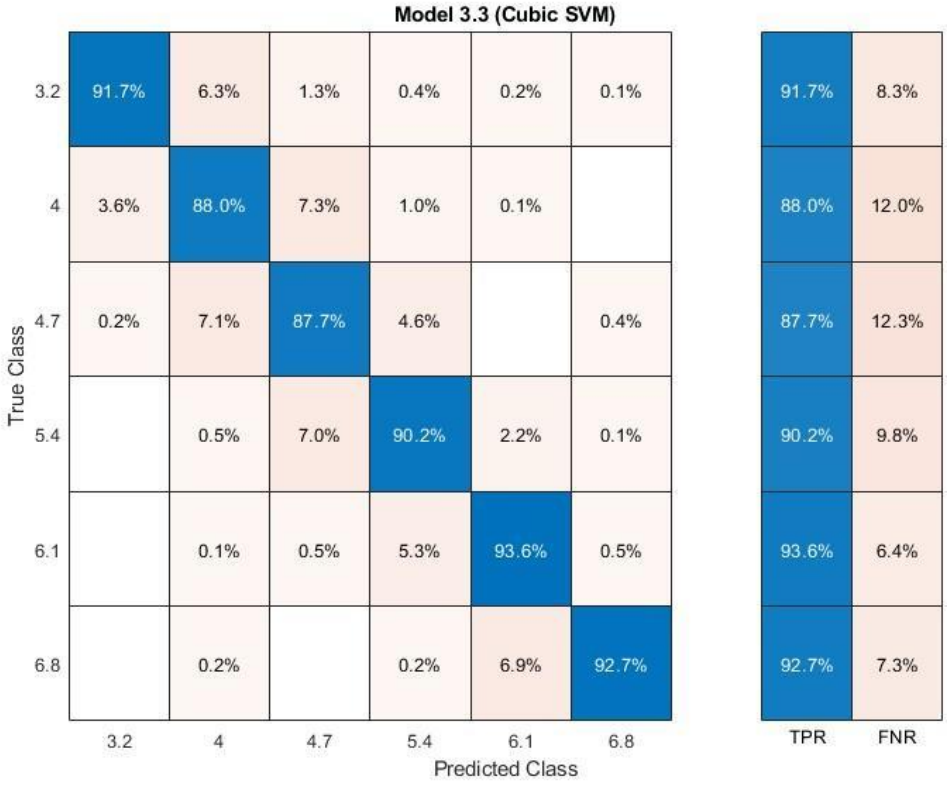

Figure 32. TPR FNR training confusion matrix of subset 2 for right leg using complete time-series.

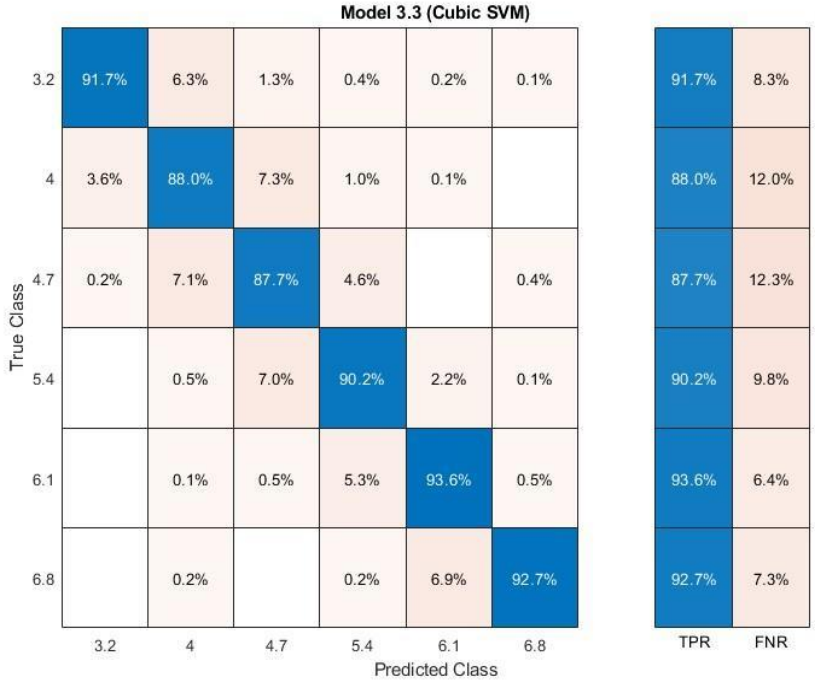

Figure 33. TPR FNR test confusion matrix of subset 2 for right leg using complete time-series.

The cubic SVM model comes up to be the most accurate, demonstrating a significant increase in performance in comparison with the training and test set. This improvement suggests excellent generalization capability, indicates that the model performs well on unseen data.

Table 4

Accuracy Results Of Subset 2 For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	90.65	92
3.2	SVM	89.03333	89.8
3.5	SVM	88.4	90.86667
5.3	Neural Network	87.2	89.2
4.6	KNN	85.51667	88.66667
4.1	KNN	85.03333	86.33333
4.4	KNN	84.91667	85.8
4.2	KNN	84.31667	86.93333
5.2	Neural Network	82.71667	85.93333
4.5	KNN	80.95	83.13333
5.5	Neural Network	73.33333	76
5.4	Neural Network	73.23333	74.2

Table 4 (cont.)

5.1	Neural Network	70.88333	74.6
3.4	SVM	67.3	71.4
3.1	SVM	66.16667	67.2
4.3	KNN	55.56667	65
3.6	SVM	54.58333	58
2.1	Tree	49.86667	50.93333
2.2	Tree	41.1	42
2.3	Tree	34.45	35

4.2.1.1.4 Remaining frames 50. Analysis of the skipped frames confusion matrices shows a drop in accuracy. even though it dropped to half the frames, but accuracy has been affected. This indicates that skipping frames leads to a loss of critical information, resulting in poorer classification performance compared to the methods using skipped speeds or the full time series data.

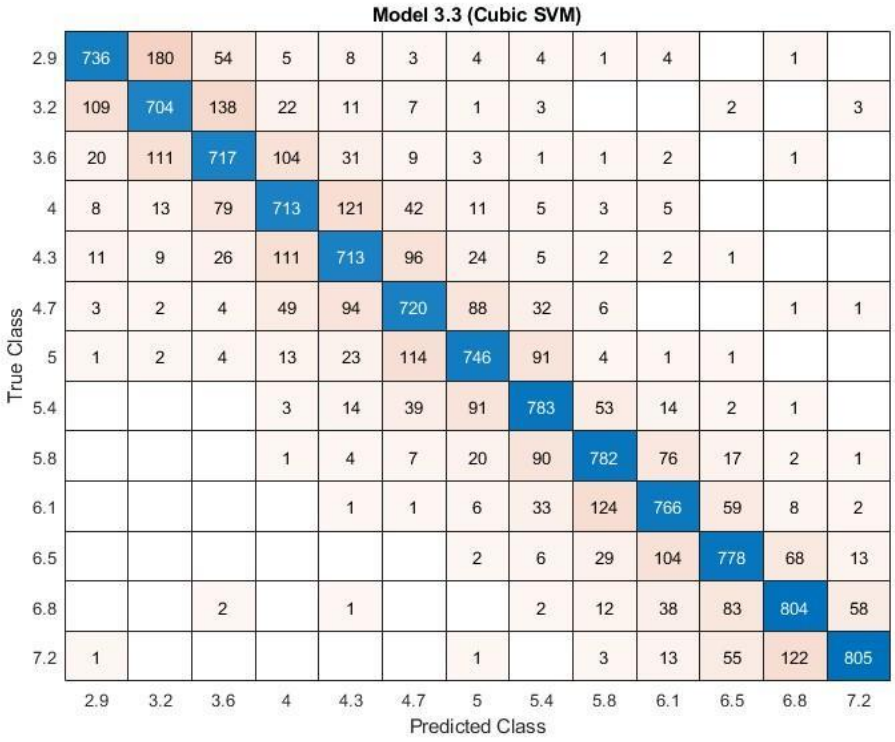

Figure 34. Training confusion matrix of 50 remaining frames for right leg using complete time series.

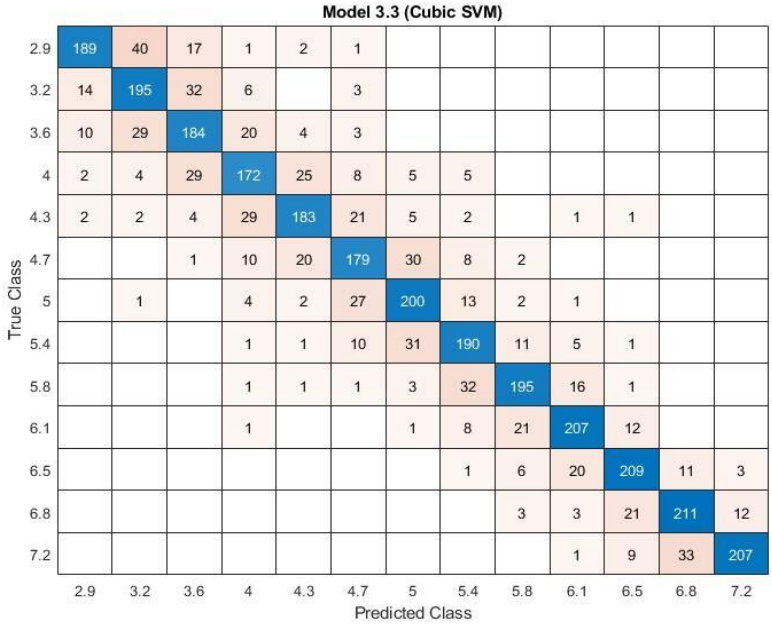

Figure 35. Test confusion matrix of 50 remaining frames for right leg using complete time series.

The high precision indicates a low rate of false positives, meaning the method is reliable in identifying true speed events although frames were dropped. Furthermore, the high sensitivity observed in the TPR plot confirms that the method is very effective at correctly classifying actual speed occurrences.

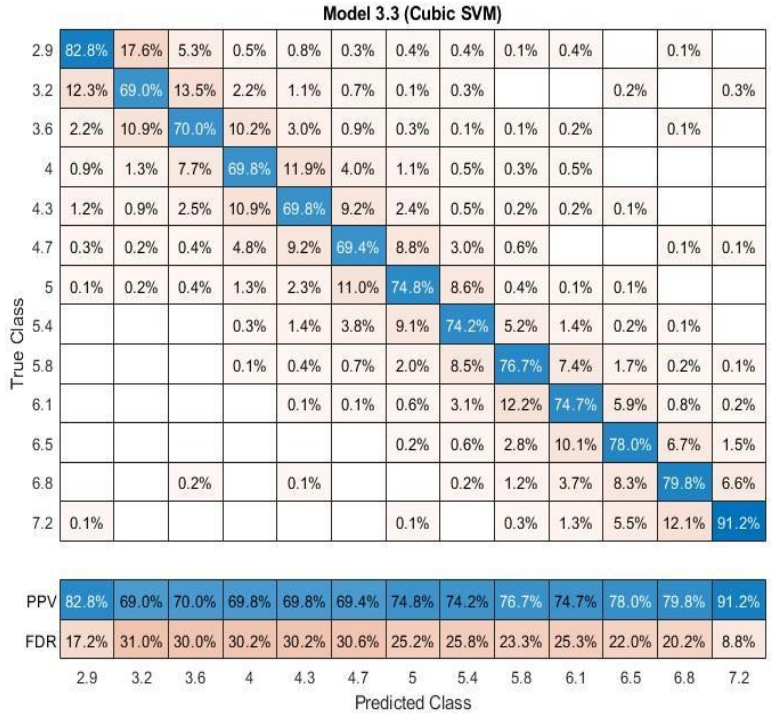

Figure 36. PPV-FDR training confusion matrix of 50 remaining frames for right leg using complete time series.

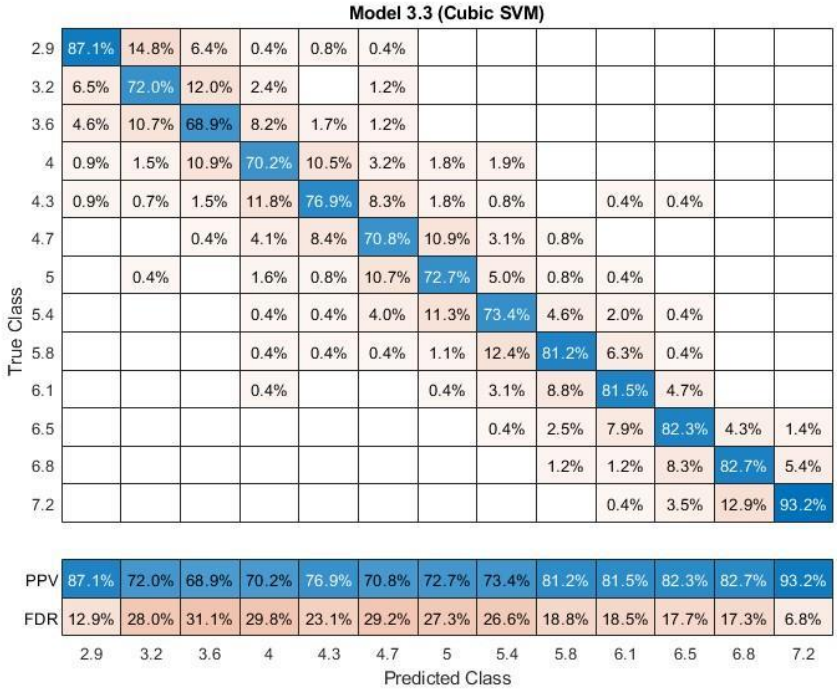

Figure 37. PPV-FDR test confusion matrix of 50 remaining frames for right leg using complete time series.

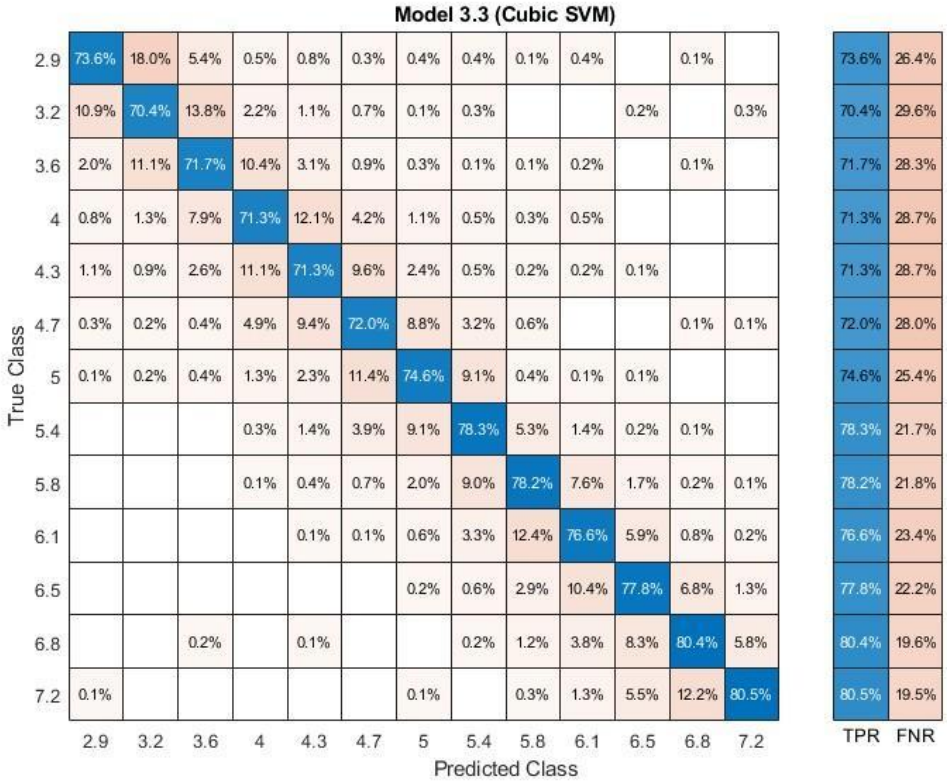

Figure 38. TPR-FNR training confusion matrix of 50 remaining frames for right leg using complete time series.

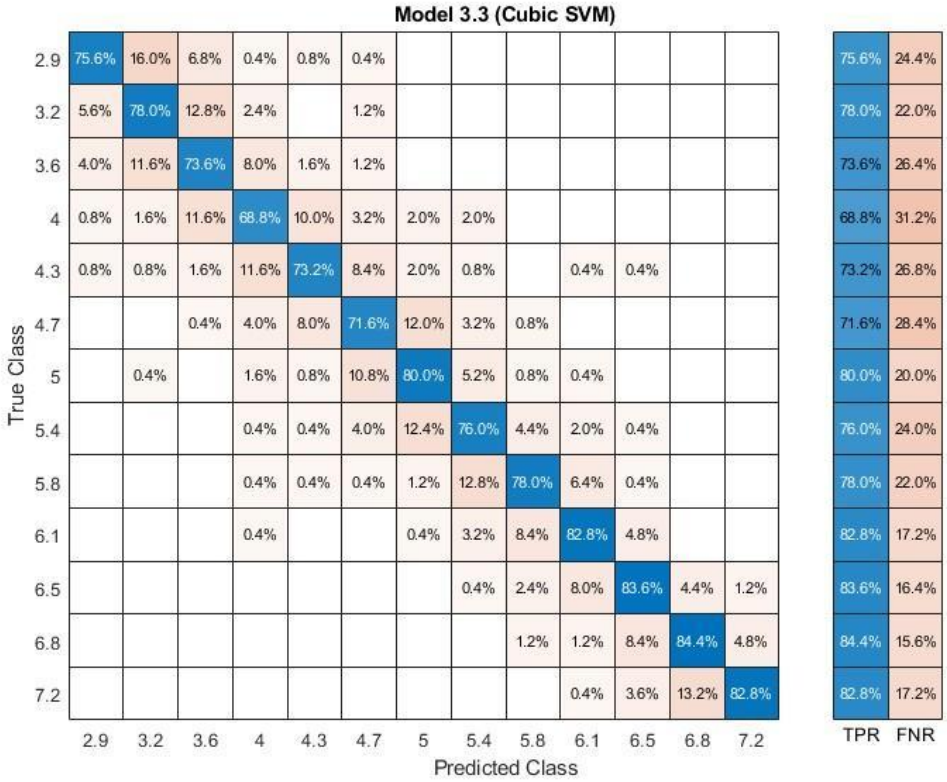

Figure 39. TPR- FNR test confusion matrix of 50 remaining frames for right leg using complete time series.

Among all models evaluated, the cubic SVM model stood out with the highest accuracy. Its accuracy increased from the training set to the test set. This improvement on unseen data is particularly impressive, as it demonstrates the model's ability to generalize beyond the training examples and make accurate predictions on new data.

Table 5

Models 'Accuracy Of 50 Remaining Frames For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	75.13077	77.56923
3.2	SVM	71.16154	73.84615
3.5	SVM	68.95385	71.35385
4.6	KNN	67.96923	69.53846
4.4	KNN	66.50769	69.2
4.2	KNN	66.36154	68.55385
5.3	Neural Network	65.97692	69.07692

Table 5 (cont.)

4.1	KNN	65.44615	68.12308
4.5	KNN	61.53077	63.90769
5.2	Neural Network	51.94615	55.90769
3.4	SVM	46.85385	48.86154
5.4	Neural Network	46.85385	48.95385
4.3	KNN	46.00769	52.30769
3.1	SVM	45.36154	45.26154
5.1	Neural Network	44.85385	45.56923
5.5	Neural Network	44.47692	47.44615
3.6	SVM	34.87692	36.55385
2.1	Tree	27.89231	29.10769
2.2	Tree	21.41538	22.98462
2.3	Tree	16.93846	17.75385

4.2.1.1.5 Remaining frames 25. Among all models evaluated, the cubic SVM model stood out with the highest accuracy. The 25 remaining frames methos show a very good accuracy even though many frames were reduced.

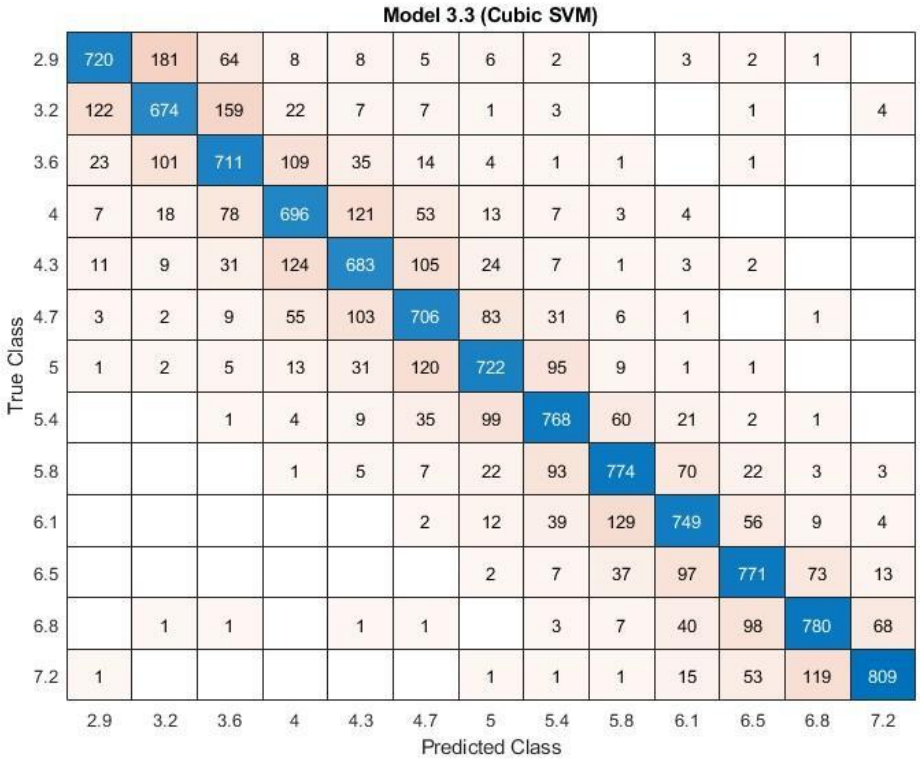

Figure 40. Training confusion matrix of 25 remaining frames for right leg using complete time-series.

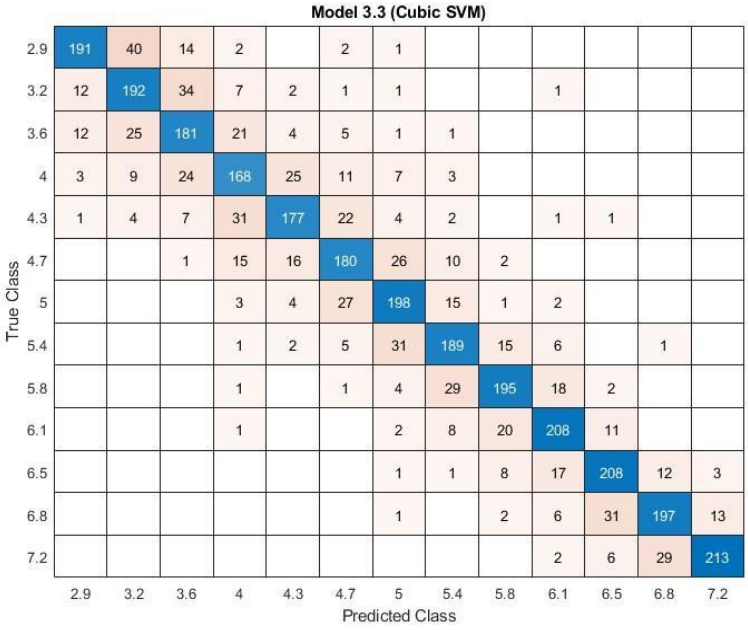

Figure 41. Test confusion matrix of 25 remaining frames for right leg using complete time-series.

The confusion matrices provide a detailed view of the classification performance, revealing high accuracy and precision in speed detection. This high precision signifies that the method excels at correctly identifying true speed events.

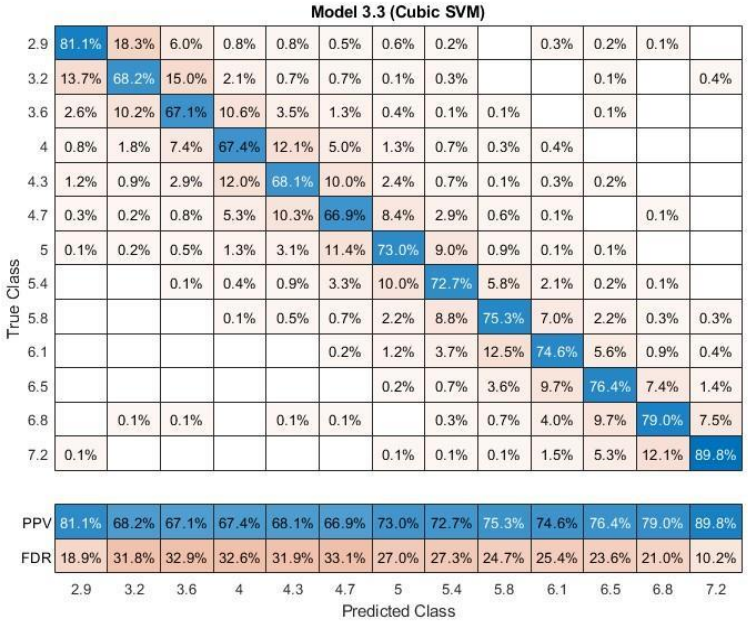

Figure 42. PPV-FDR training confusion matrix of 25 remaining frames for right leg using complete time-series.

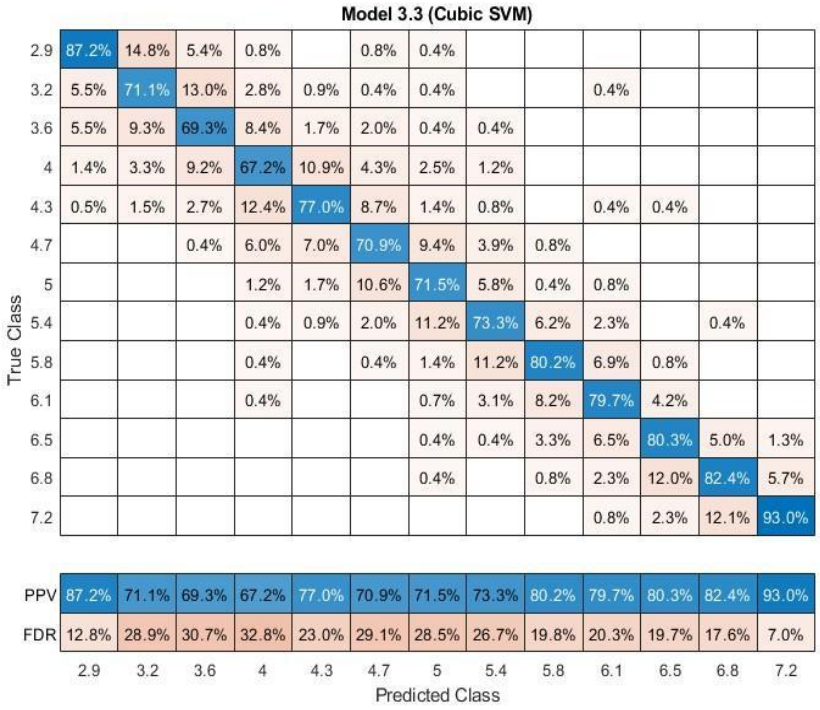

Figure 43. PPV-FDR test confusion matrix of 25 remaining frames for right leg using complete time-series.

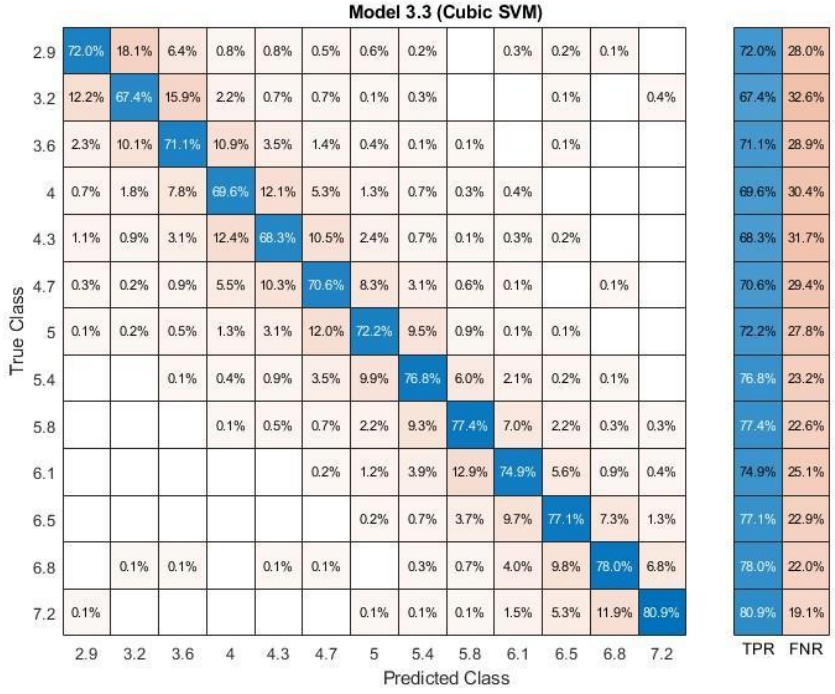

Figure 44. TPR-FNR training confusion matrix of 25 remaining frames for right leg using complete time-series.

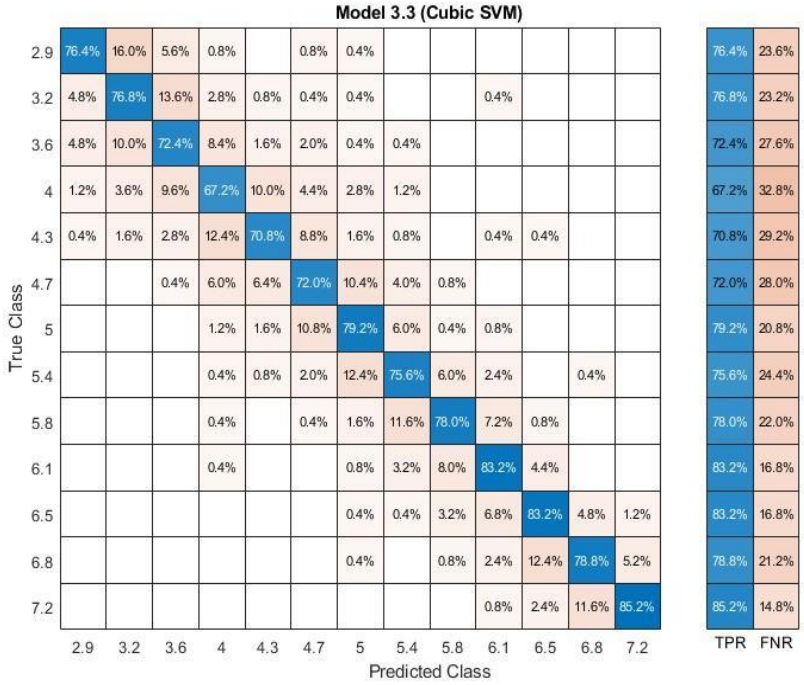

Figure 45. TPR-FNR test confusion matrix of 25 remaining frames for right leg using complete time-series.

The cubic SVM model led the field in terms of accuracy, surpassing all other models evaluated. Its performance was particularly noteworthy on the test set, where it exceeded the already high accuracy observed during training. This result underscores the model's exceptional ability to generalize from the training data to unseen examples, suggesting that it has captured the underlying relationships in the data effectively and can be relied upon to make accurate predictions in diverse situations

Table 6

Models 'Accuracy Of 25 Remaining Frames For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	73.56154	76.83077
3.2	SVM	69.76154	72.8
3.5	SVM	68	71.32308
4.6	KNN	66.64615	68.52308
4.4	KNN	65.33077	66.86154
5.3	Neural Network	64.93846	66.95385
4.2	KNN	64.67692	66.67692

Table 6 (cont.)

4.1	KNN	63.36154	65.38462
4.5	KNN	60.9	62.46154
5.2	Neural Network	55.73077	57.96923
3.4	SVM	46.17692	47.93846
4.3	KNN	45.83077	52.89231
5.4	Neural Network	45.16154	48.27692
5.5	Neural Network	45.16154	45.56923
5.1	Neural Network	44.62308	46.89231
3.1	SVM	43.13846	44.64615
3.6	SVM	34.67692	36.06154
2.1	Tree	27.39231	29.01538
2.2	Tree	21.34615	22.83077
2.3	Tree	16.77692	17.78462

4.2.1.1.6 Remaining frames 20. The confusion matrices demonstrate a significant difference in accuracy between the skipped frames method and the methods using skipped speeds or the full time series data. The lower accuracy observed when skipping frames underscores the importance of retaining complete temporal information for optimal classification. While data reduction techniques can be useful, this finding highlights the need to carefully consider their impact on classification performance.

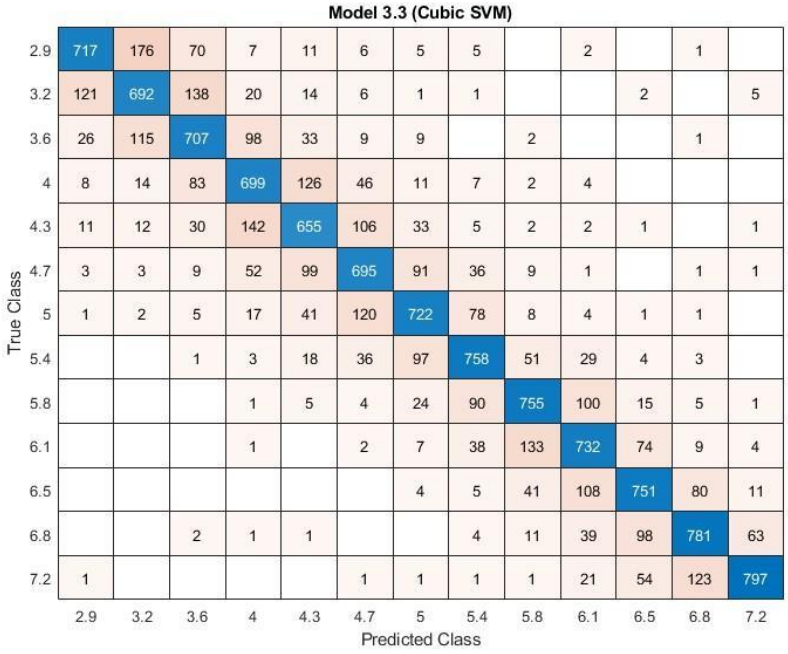

Figure 46. Training confusion matrix of 20 remaining frames for right leg using complete time-series.

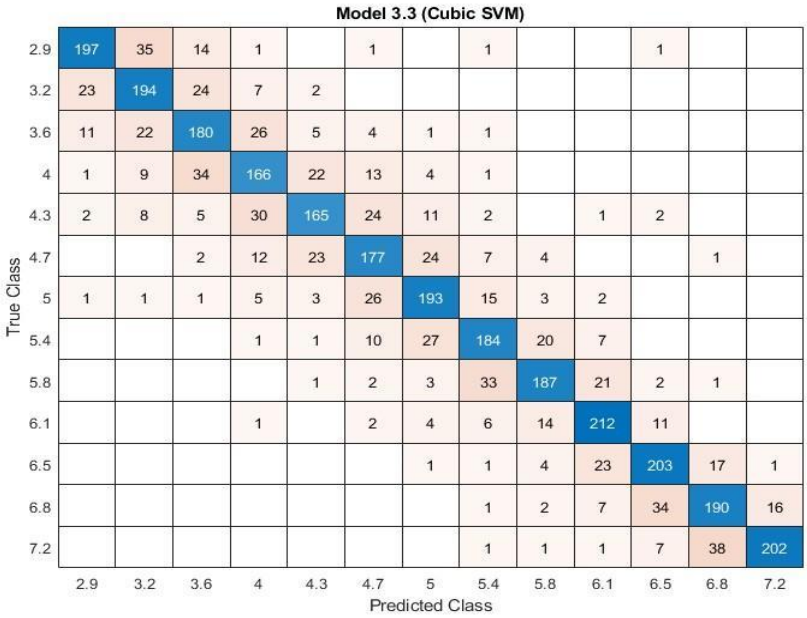

Figure 47. Test confusion matrix of 20 remaining frames for right leg using complete time-series.

The confusion matrices illustrate the impressive performance of the classification method, achieving both high accuracy and exceptional precision in speed detection. This high precision translates to a very low rate of false alarms.

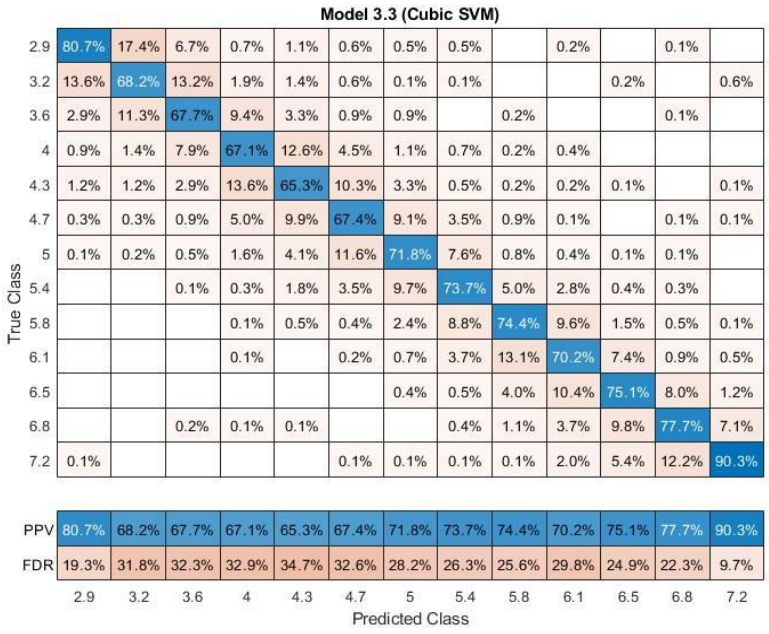

Figure 48. PPV-FDR training confusion matrix of 20 remaining frames for right leg using complete time-series.

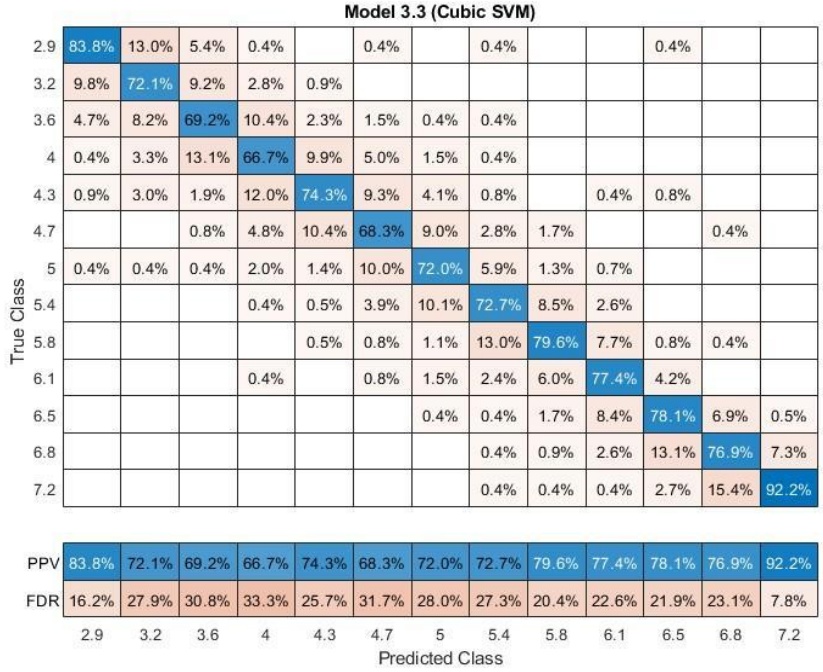

Figure 49. PPV-FDR test confusion matrix of 20 remaining frames for right leg using complete time-series.

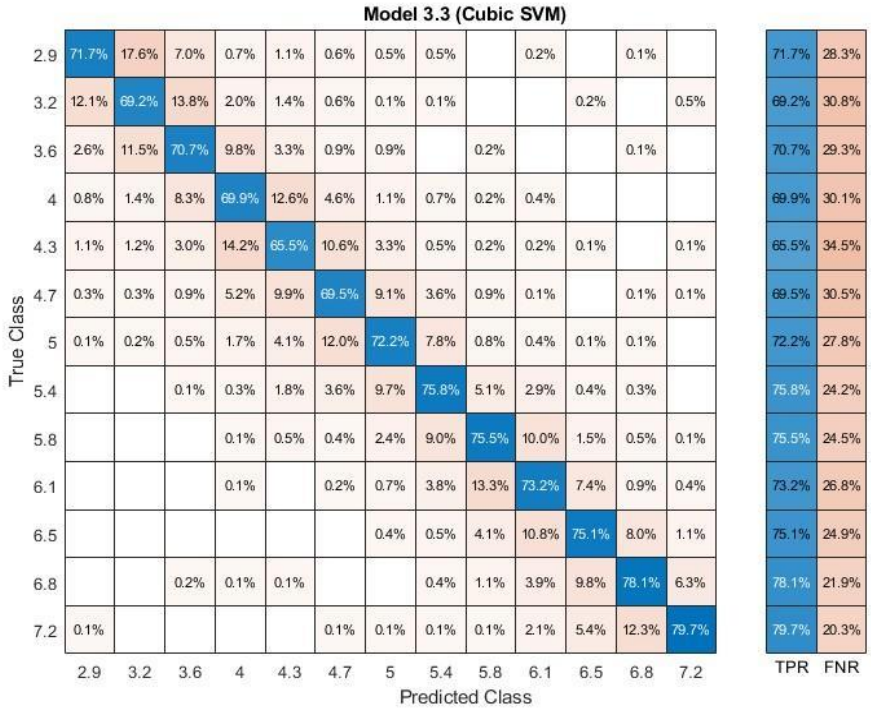

Figure 50. TPR-FNR training confusion matrix of 20 remaining frames for right leg using complete time-series.

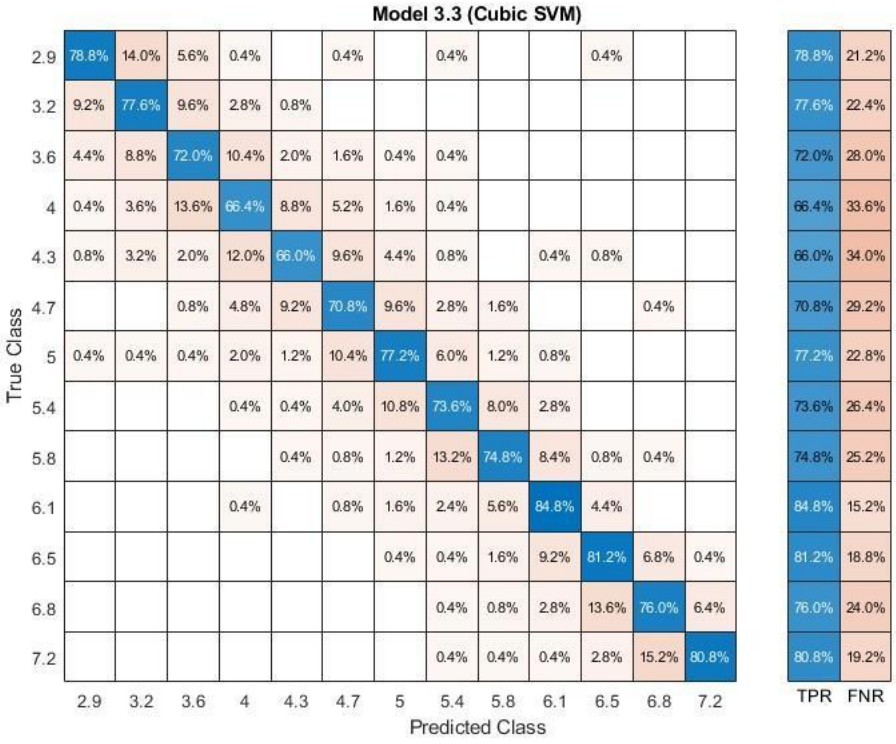

Figure 51. TPR-FNR test confusion matrix of 20 remaining frames for right leg using complete time-series.

The cubic SVM model yielded the highest accuracy, notably improving on the test set, indicating strong predictive ability on unseen data.

Table 7

Models' Accuracy Of 20 Remaining Frames For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	72.77692	75.38462
3.2	SVM	69.03077	71.26154
3.5	SVM	66.37692	69.2
4.6	KNN	65.78462	67.56923
4.4	KNN	64.01538	66.49231
4.2	KNN	63.93846	65.63077
5.3	Neural Network	62.52308	64.73846
4.1	KNN	62.43077	63.87692
4.5	KNN	59.71538	61.2
5.2	Neural Network	55.8	56
3.4	SVM	46.32308	49.44615

Table 7 (cont.)

4.3	KNN	45.65385	52.95385
5.4	Neural Network	44.27692	46.36923
5.1	Neural Network	44	44
5.5	Neural Network	43.2	46.21538
3.1	SVM	41.23077	41.10769
3.6	SVM	33.61538	35.01538
2.1	Tree	26.4	27.81538
2.2	Tree	22.07692	21.04615
2.3	Tree	17.57692	17.6

4.2.1.1.7 Remaining frames 10. A clear difference in accuracy emerged when comparing the skipped frames method with the skipped speeds and full time series approaches. The skipped frames method exhibited lower accuracy, indicating that the removal of frames disrupts the temporal patterns and dependencies within the data, hindering the classifier's ability to learn effectively. This observation underscores the importance of preserving the integrity of the time series data for optimal classification performance.

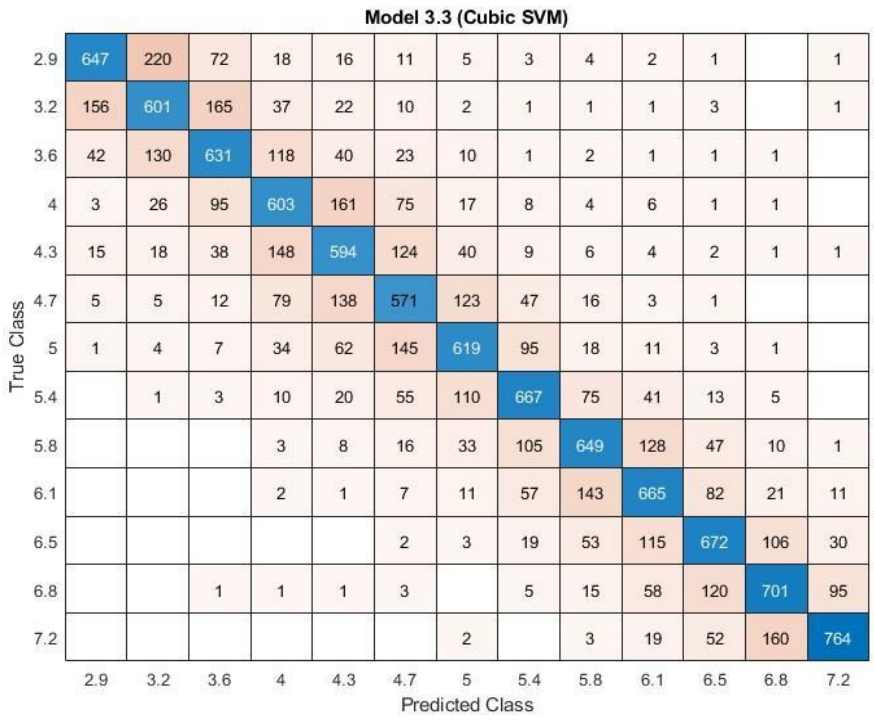

Figure 52. Training confusion matrix of 10 remaining frames for right leg using complete time-series.

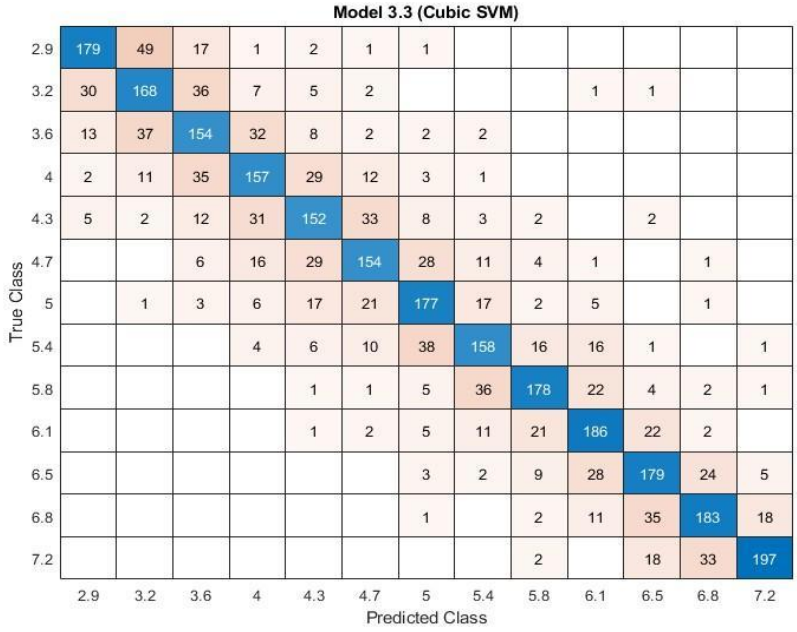

Figure 53. Test confusion matrix of 10 remaining frames for right leg using complete time-series.

The confusion matrices provide a detailed view of the classification performance, revealing high overall accuracy and exceptionally high precision in speed detection.

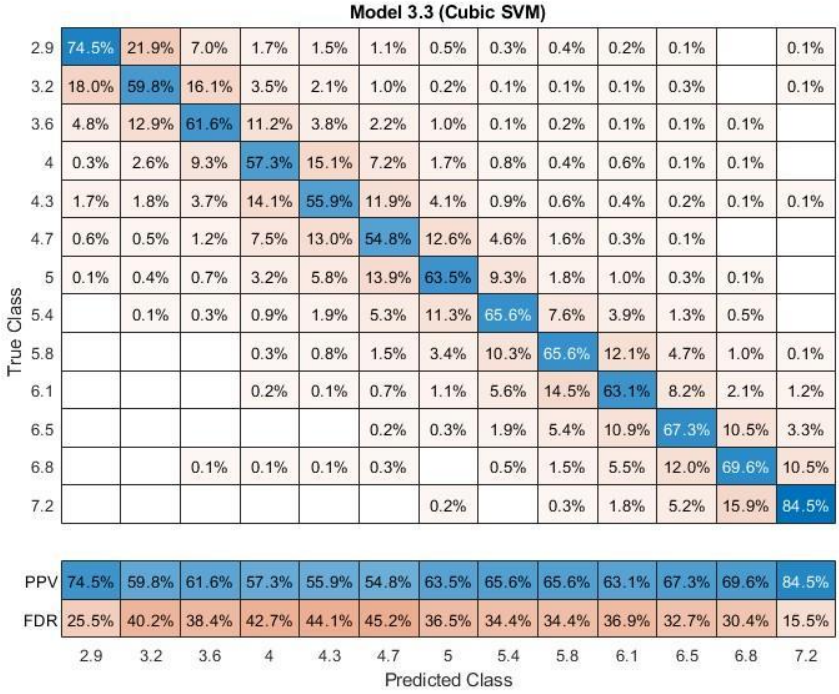

Figure 54. PPV-FDR training confusion matrix of 10 remaining frames for right leg using complete time-series.

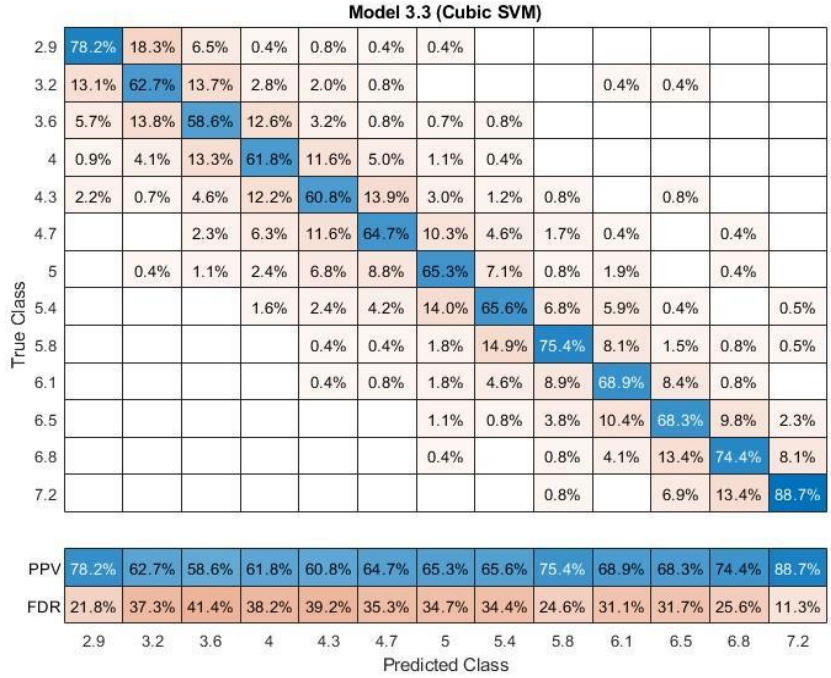

Figure 55. PPV-FDR test confusion matrix of 10 remaining frames for right leg using complete time-series.

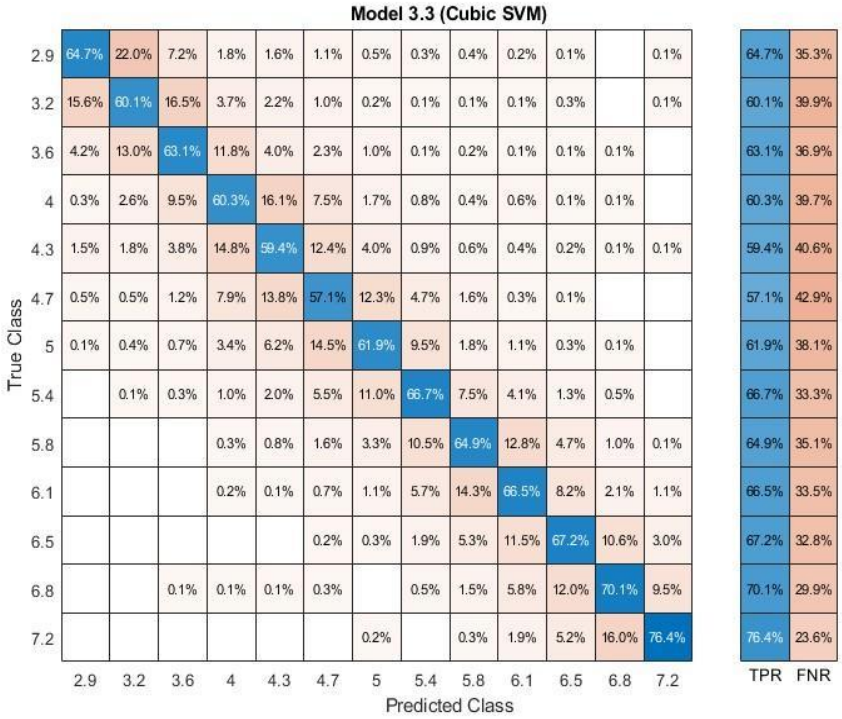

Figure 56. TPR-FNR training confusion matrix of 10 remaining frames for right leg using complete time-series.

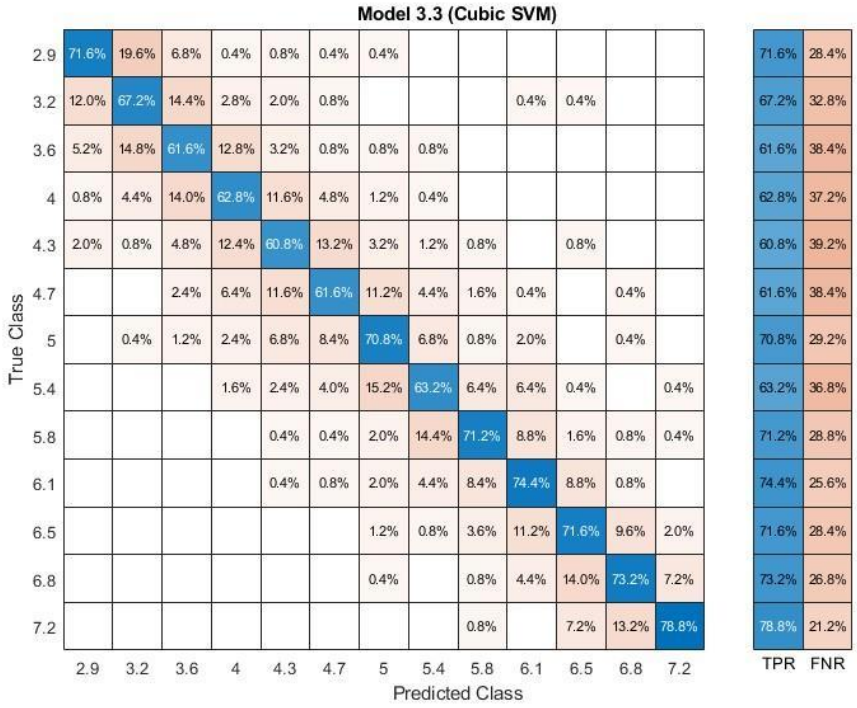

Figure 57. TPR-FNR test confusion matrix of 10 remaining frames for right leg using complete time-series.

Despite being trained on only 10 remaining frames, the cubic SVM model achieved the highest accuracy, notably exceeding its training performance on the test set.

Table 8

Models' Accuracy Of 10 Remaining Frames For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	64.49231	68.36923
3.2	SVM	60.01538	63.35385
4.6	KNN	59.15385	60.58462
3.5	SVM	58.12308	61.63077
4.4	KNN	57.38462	58.67692
4.2	KNN	57.36154	58.83077
4.1	KNN	55.21538	56.4
5.3	Neural Network	55.10769	57.84615
4.5	KNN	53.79231	55.81538
5.2	Neural Network	48.58462	49.07692
3.4	SVM	45.78462	49.10769
4.3	KNN	43.2	49.04615

Table 8 (cont.)

5.5	Neural Network	38.63077	41.63077
5.1	Neural Network	38.28462	38.76923
5.4	Neural Network	38.25385	40.52308
3.1	SVM	33.21538	33.75385
3.6	SVM	29.68462	30.09231
2.1	Tree	25.22308	25.01538
2.2	Tree	19.87692	20.58462
2.3	Tree	16.16154	17.07692

4.2.1.1.8 Remaining frames 5. These confusion matrices illustrate the classification results achieved using only the 5 remaining frames from the complete time series data. Surprisingly, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

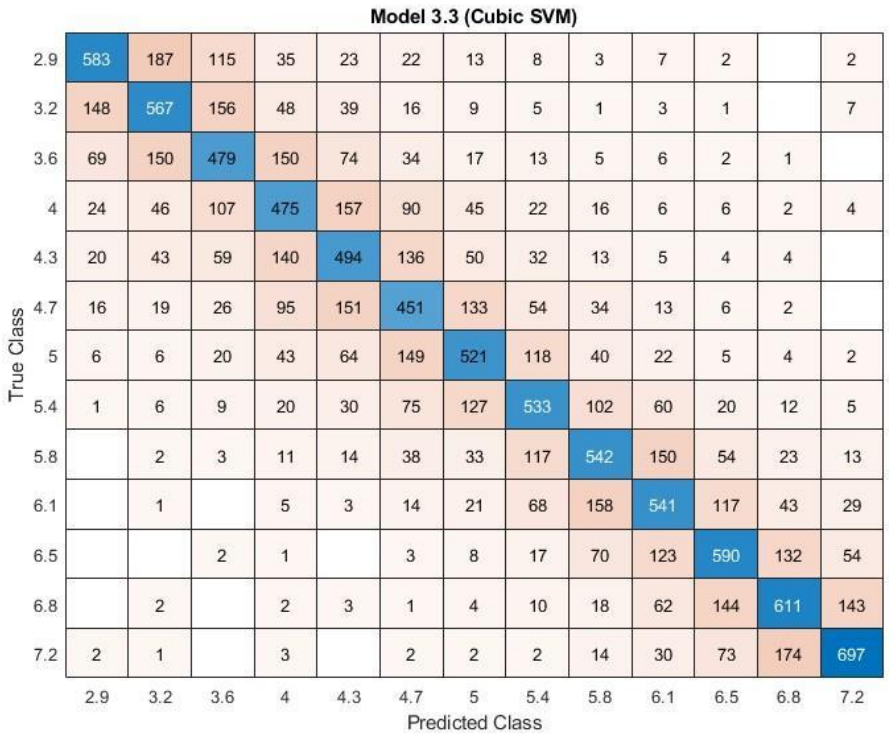

Figure 58. Training confusion matrix of 5 remaining frames for right leg using complete time-series.

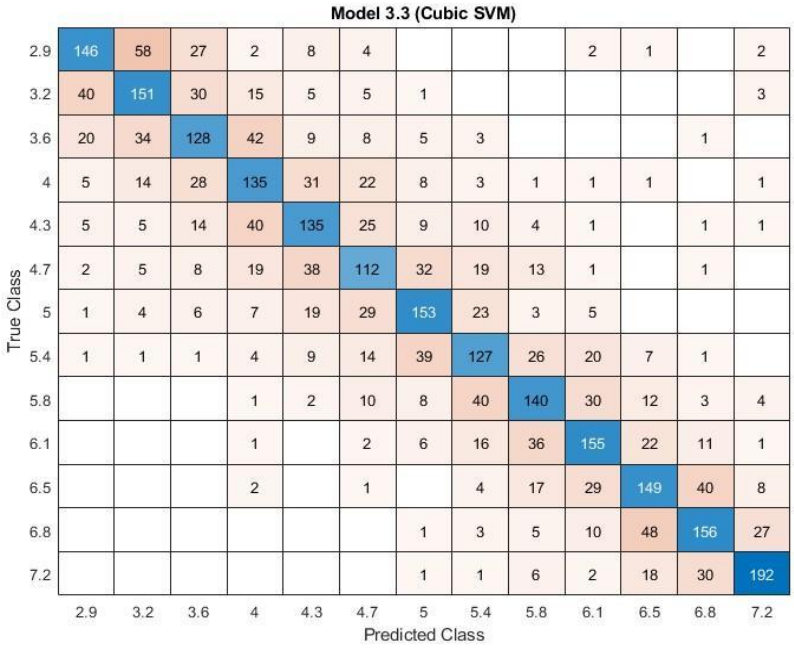

Figure 59. Test confusion matrix of 5 remaining frames for right leg using complete time-series.

The confusion matrices demonstrate high accuracy and exceptional precision in speed detection, minimizing false positives. The high sensitivity in the TPR plot confirms the method's effectiveness in capturing true speed occurrences.

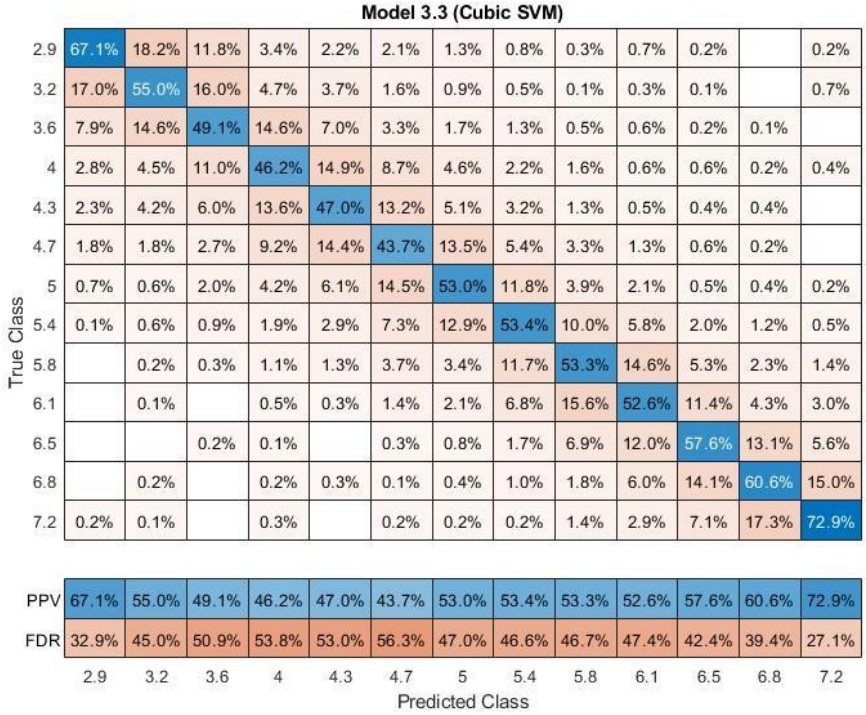

Figure 60. PPV-FDR training confusion matrix of 5 remaining frames for right leg using complete time-series.

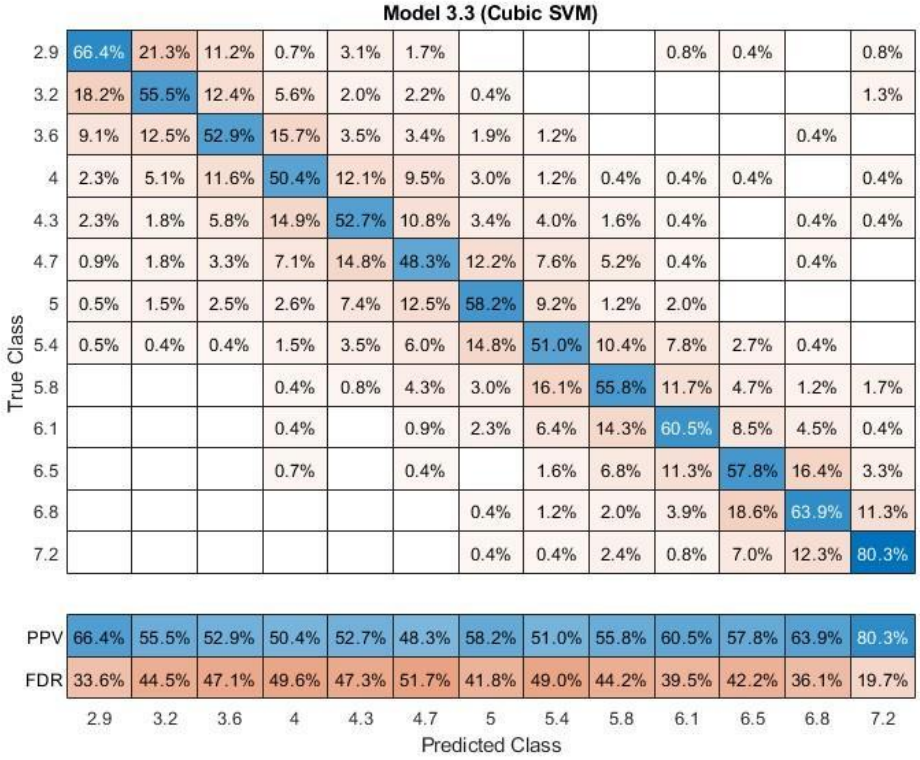

Figure 61. PPV-FDR test confusion matrix of 5 remaining frames for right leg using complete time-series.

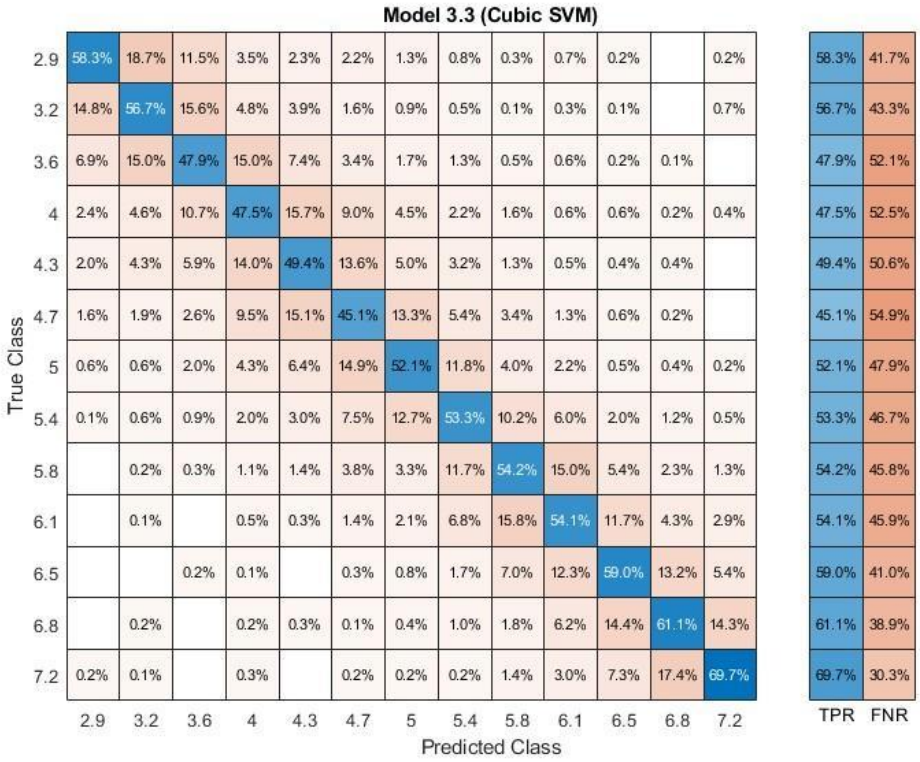

Figure 62. TPR-FNR training confusion matrix of 5 remaining frames for right leg using complete time-series.

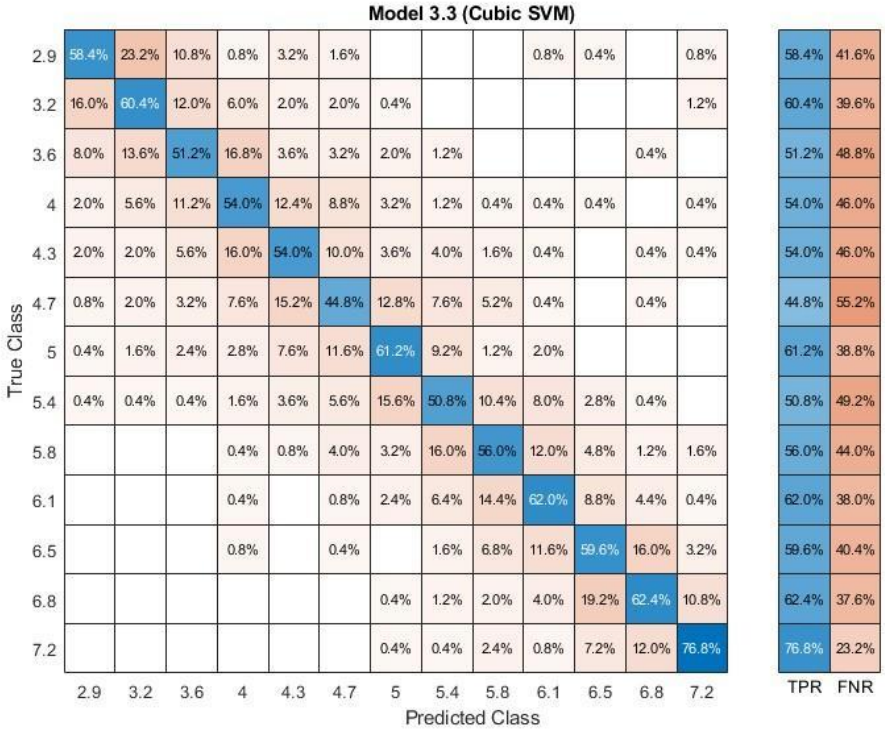

Figure 63. TPR-FNR test confusion matrix of 5 remaining frames for right leg using complete time-series.

Using just 5 remaining frames, the cubic SVM model delivered the highest accuracy, demonstrating its efficiency and generalization ability.

Table 9

Models' Accuracy Of 5 Remaining Frames For Right Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	54.49231	57.81538
4.6	KNN	52.16923	55.13846
4.2	KNN	50.65385	52.70769
4.4	KNN	50.36154	53.10769
5.3	Neural Network	48.86923	51.10769
4.1	KNN	48.43846	50
4.5	KNN	48.11538	49.56923
3.5	SVM	47.83077	50.52308
3.2	SVM	47.64615	49.38462
3.4	SVM	44.19231	46.27692
4.3	KNN	40.14615	43.35385

Table 9 (cont.)

5.2	Neural Network	39.49231	42.24615
5.5	Neural Network	32.13846	32.76923
5.4	Neural Network	31.79231	33.01538
5.1	Neural Network	30.48462	31.04615
3.1	SVM	25.85385	26.58462
3.6	SVM	25.16923	25.56923
2.1	Tree	23.23846	24.21538
2.2	Tree	18.13077	17.63077
2.3	Tree	15.06154	14.83077

4.2.1.1.9 Gait segmentation-first phase (stance phase). The gait segmentation method for the stance phase achieves high accuracy, comparable to that of traditional feature calculation techniques. This accurate segmentation is achieved using an SVM model, demonstrating its effectiveness in identifying the stance phase.

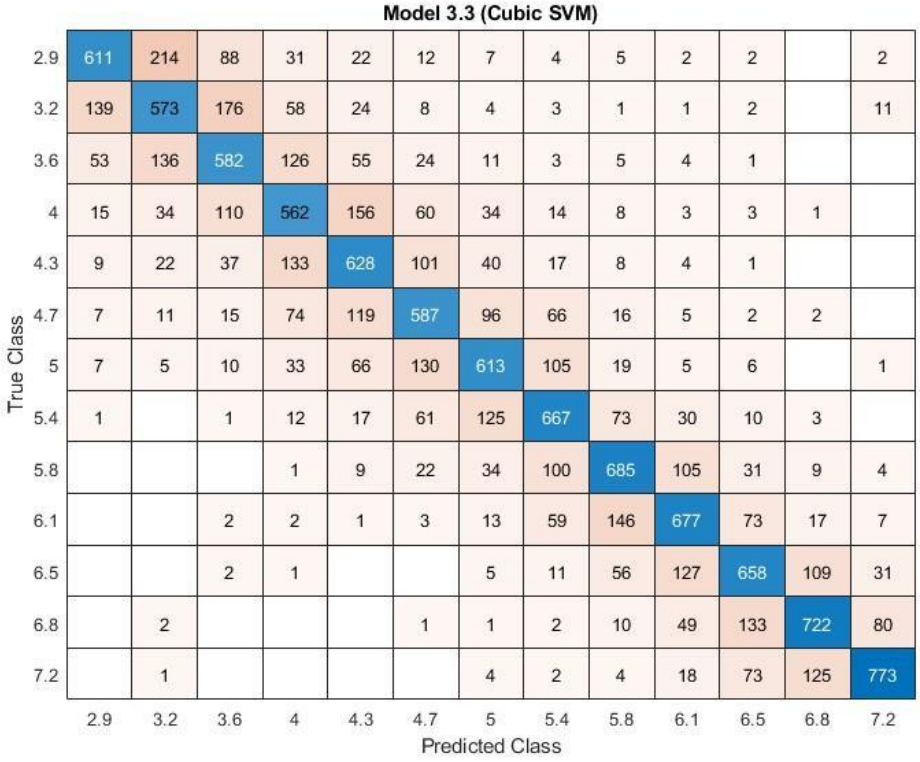

Figure 64. Training confusion matrix of gait segmentation (first phase) for right leg using time-series.

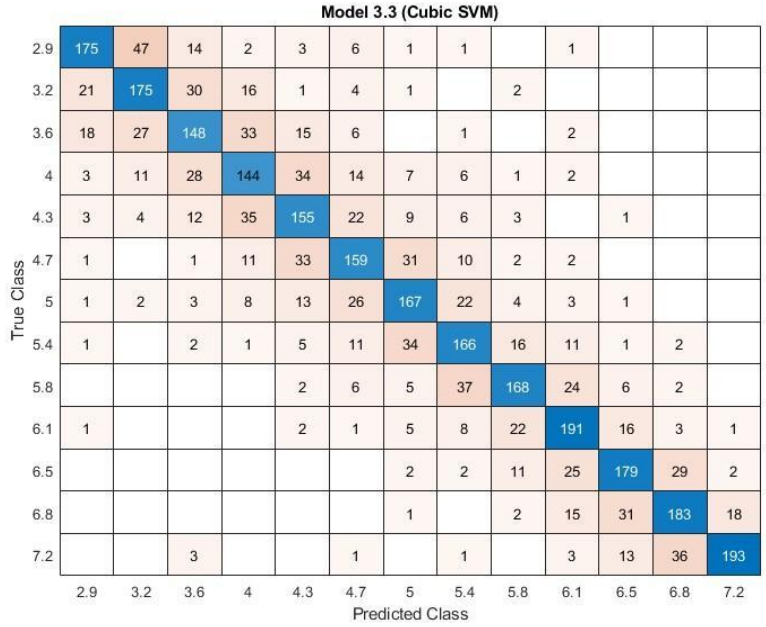

Figure 65. Test confusion matrix of gait segmentation (first phase) for right leg using time-series.

These confusion matrices demonstrate the high accuracy and precision of the classification method in identifying different gait phases. The high precision indicates a low rate of misclassifying gait phases, ensuring accurate identification.

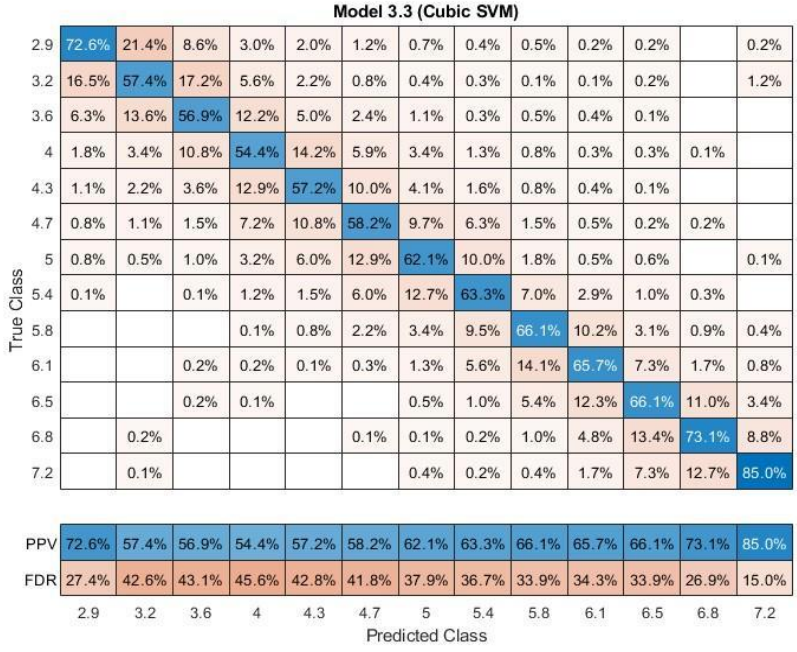

Figure 66. PPV-FDR training confusion matrix of gait segmentation (first phase) for right leg using time-series.

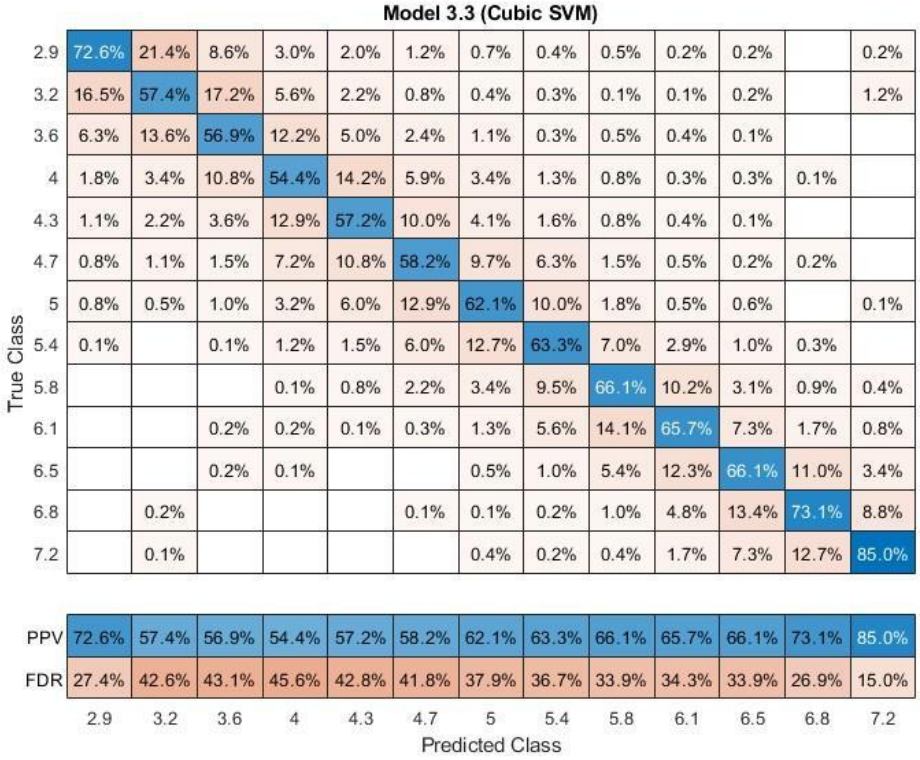

Figure 67. PPV-FDR test confusion matrix of gait segmentation (first phase) for right leg using time-series.

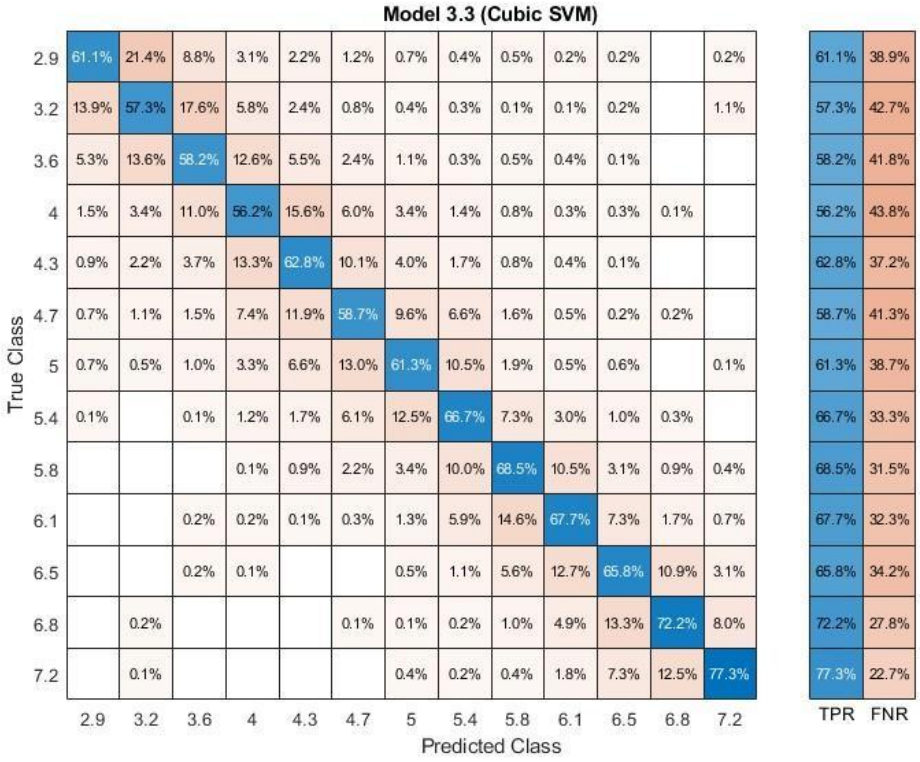

Figure 68. TPR-FNR training confusion matrix of gait segmentation (first phase) for right leg using time-series.

Figure 69. TPR-FNR test confusion matrix of gait segmentation (first phase) for right leg using time-series.

In gait analysis, the cubic SVM model not only demonstrated high accuracy on the training data but also generalized exceptionally well to the test set, exhibiting even higher accuracy.

Table 10

Models' Accuracy Of Gait Segmentation (First Phase) For Right Leg Using Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	64.13846	67.78462
4.6	KNN	60.53077	63.07692
4.2	KNN	58.87692	61.90769
3.2	SVM	58.85385	62.89231
4.4	KNN	58.73846	60.67692
4.1	KNN	58.01538	59.56923
3.5	SVM	57.94615	62.12308
4.5	KNN	55.56923	58.55385
5.3	Neural Network	54.14615	56.70769
3.4	SVM	42.04615	43.2

Table 10 (cont.)

4.3	KNN	41.79231	46.64615
5. 2	Neural Network	40.99231	44.83077
5.4	Neural Network	35.45385	39.10769
5.5	Neural Network	34.71538	35.50769
5.1	Neural Network	33.85385	36.03077
3.1	SVM	33.20769	34.73846
3.6	SVM	28.17692	30.03077
2.1	Tree	24.40769	24.67692
2.2	Tree	18.92308	19.07692
2.3	Tree	15.84615	16.06154

4.2.1.1.10 Gait segmentation-second phase (swing phase). Focusing on the swing phase of gait, these confusion matrices show the results of classifying the complete time series data. Interestingly, the test confusion matrix exhibited a higher percentage of accurate classifications compared to the training confusion matrix.

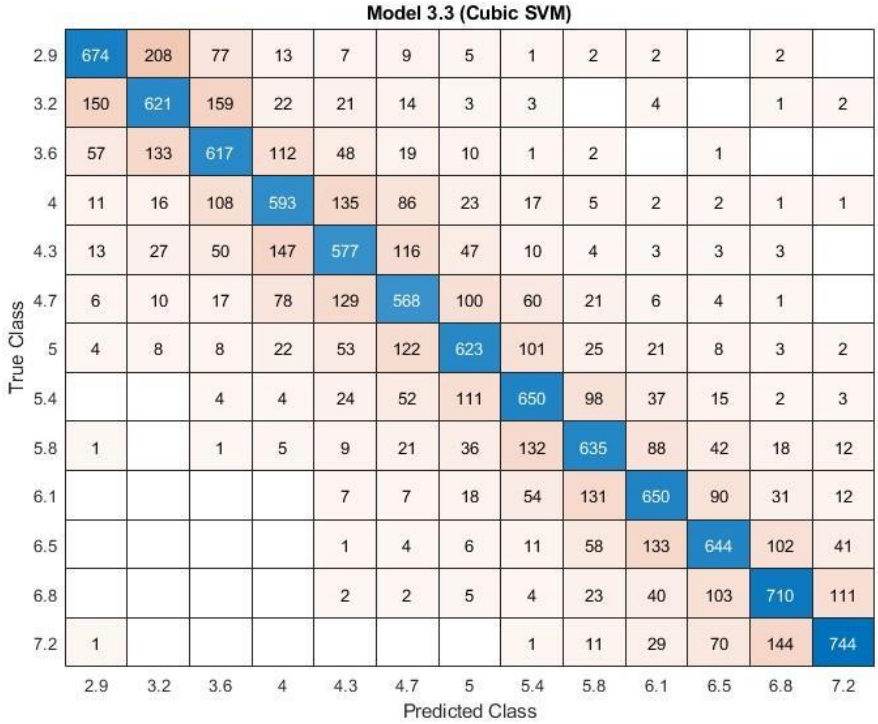

Figure 70. Training confusion matrix of gait segmentation (second phase) for right leg using time-series.

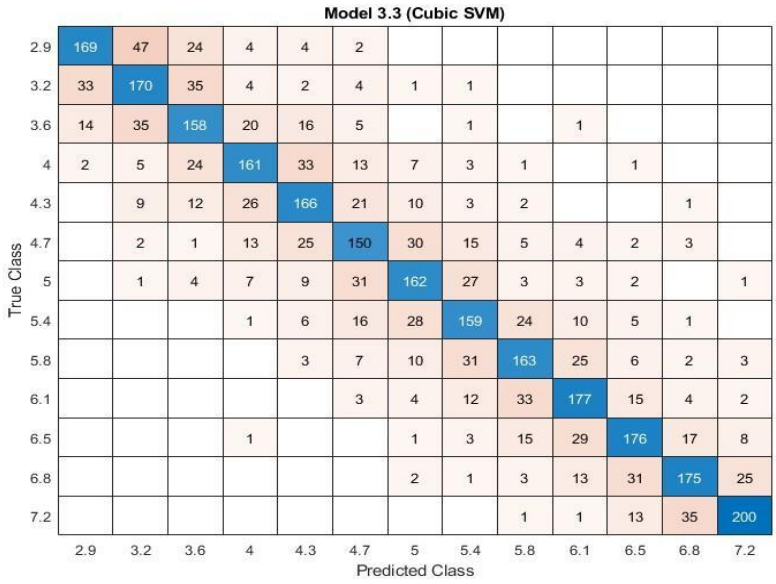

Figure 71. Test confusion matrix of gait segmentation (second phase) for right leg using time-series.

These confusion matrices demonstrate the high accuracy and precision of the classification method in identifying the swing phase of gait. The high precision indicates a low rate of misclassifying the swing phase, ensuring its accurate identification. The TPR plot further supports this by showing high sensitivity in correctly confirming the presence of the swing phase.

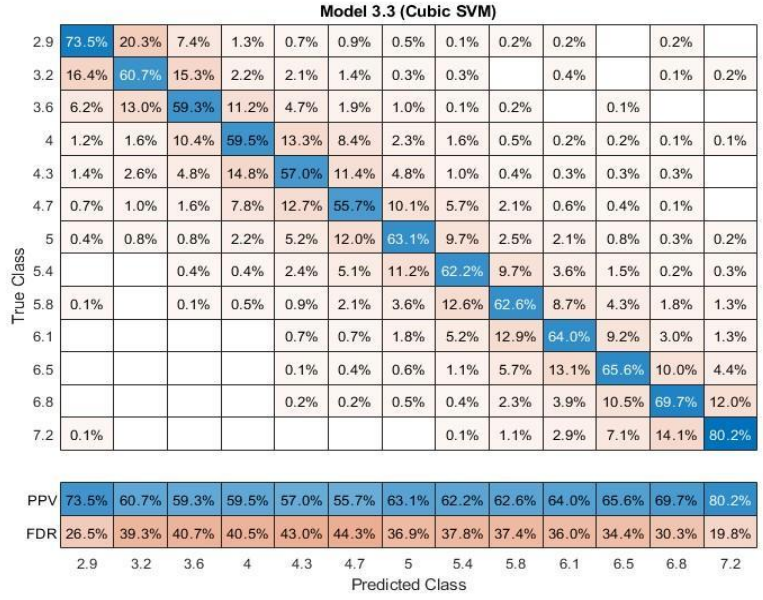

Figure 72. PPV-FDR training confusion matrix of gait segmentation (second phase) for right leg using time-series.

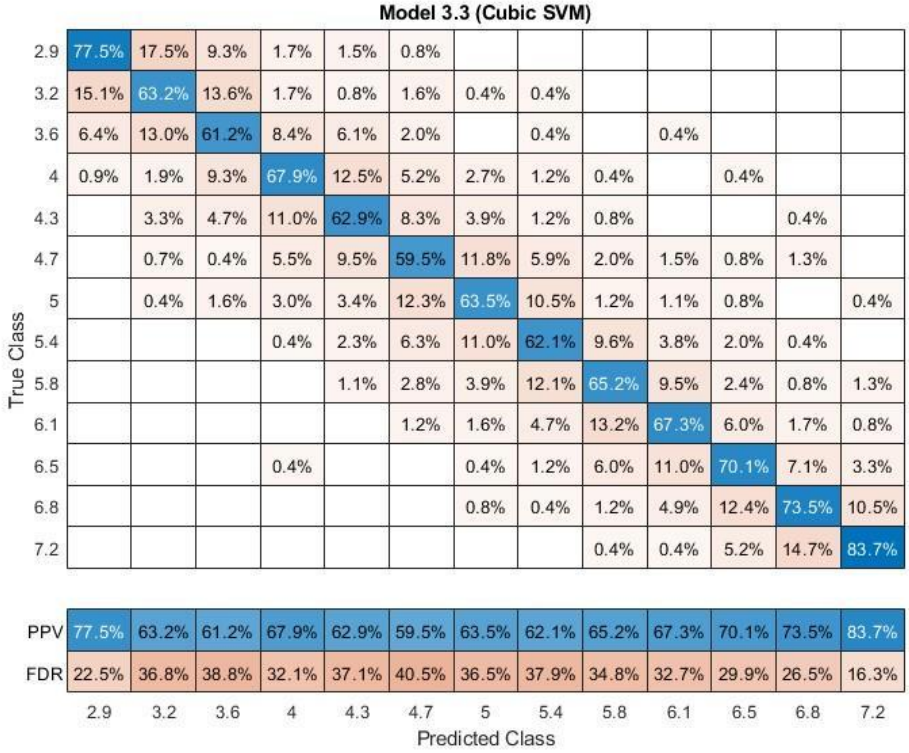

Figure 73. PPV-FDR test confusion matrix of gait segmentation (second phase) for right leg using time-series.

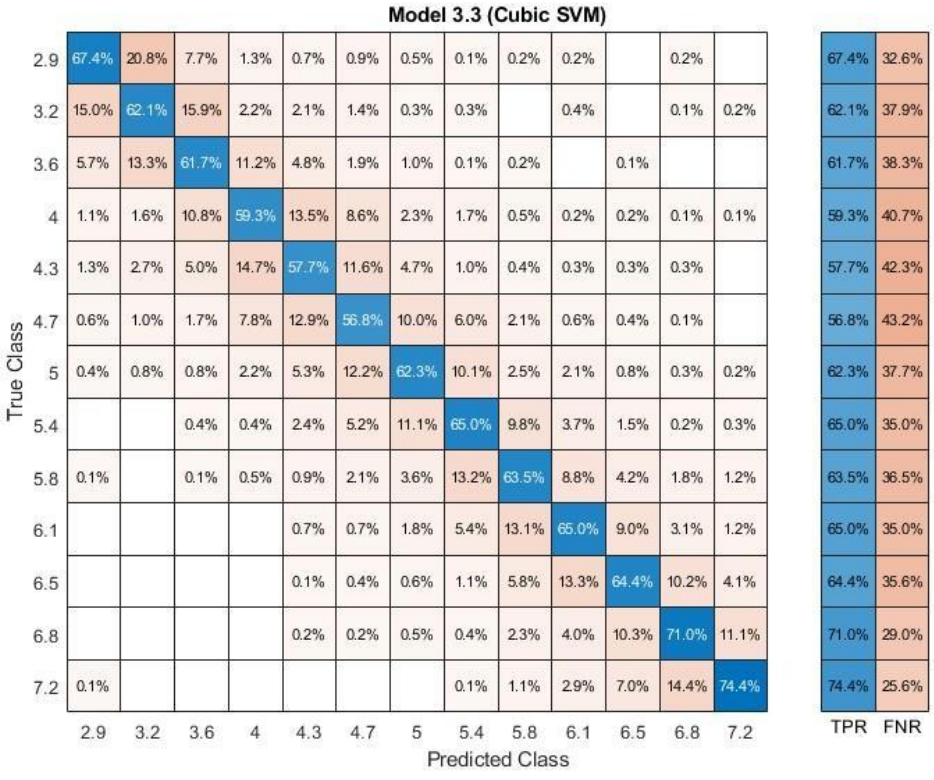

Figure 74. TPR-FNR training confusion matrix of gait segmentation (second phase) for right leg using time-series.

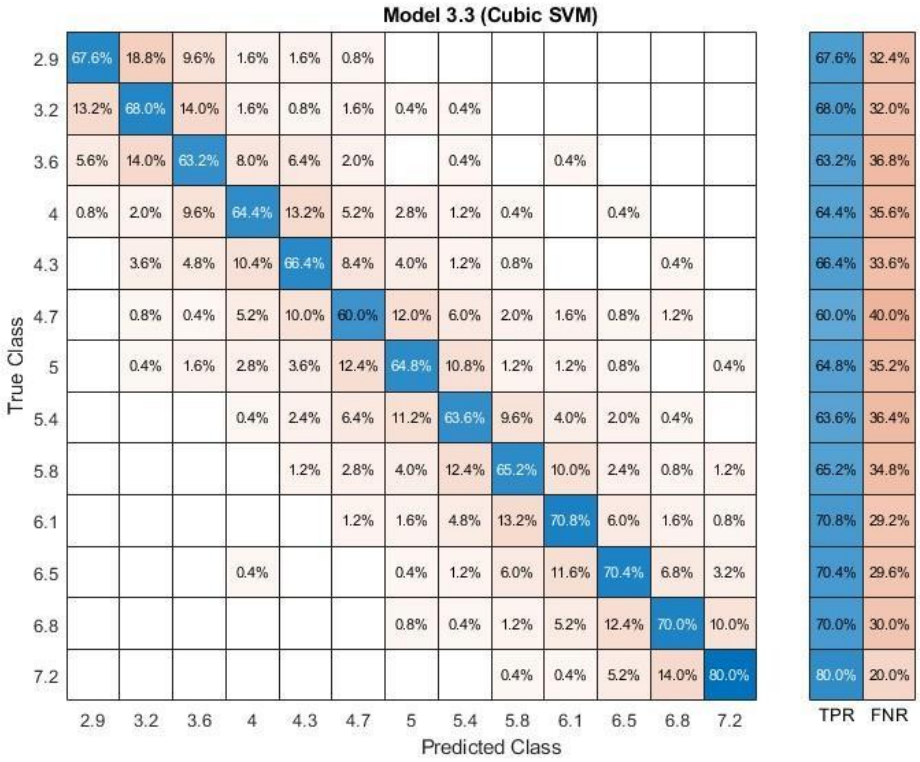

Figure 75. TPR-FNR test confusion matrix of gait segmentation (second phase) for right leg using time-series.

The cubic SVM model demonstrated exceptional performance in swing phase classification, not only achieving high accuracy on the training data but also generalizing well to the test set with even higher accuracy.

Table 11

Models' Accuracy Of Gait Segmentation (Second Phase) For Right Leg Using Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	63.89231	67.26154
4.6	KNN	63.04615	65.93846
4.2	KNN	60.98462	63.6
4.4	KNN	59.82308	63.2
4.1	KNN	59.53846	60.98462
3.2	SVM	58.63077	62.27692
3.5	SVM	57.29231	60.15385

Table 11 (cont.)

4.5	KNN	57.1	59.35385
5.3	Neural Network	55.23846	56.30769
3.4	SVM	47.44615	51.78462
4.3	KNN	45.44615	49.81538
5.2	Neural Network	44.30769	45.32308
5.4	Neural Network	37.14615	38.55385
5.5	Neural Network	36.21538	39.26154
5.1	Neural Network	35.75385	37.90769
3.1	SVM	35.54615	34.83077
3.6	SVM	30.64615	31.16923
2.1	Tree	26.63077	26.46154
2.2	Tree	21.6	20.64615
2.3	Tree	17.3	17.29231

4.2.1.2 Left leg analysis.

4.2.1.2.1 Time series. These confusion matrices illustrate the classification of the complete time series data specifically for left leg gait analysis. Notably, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

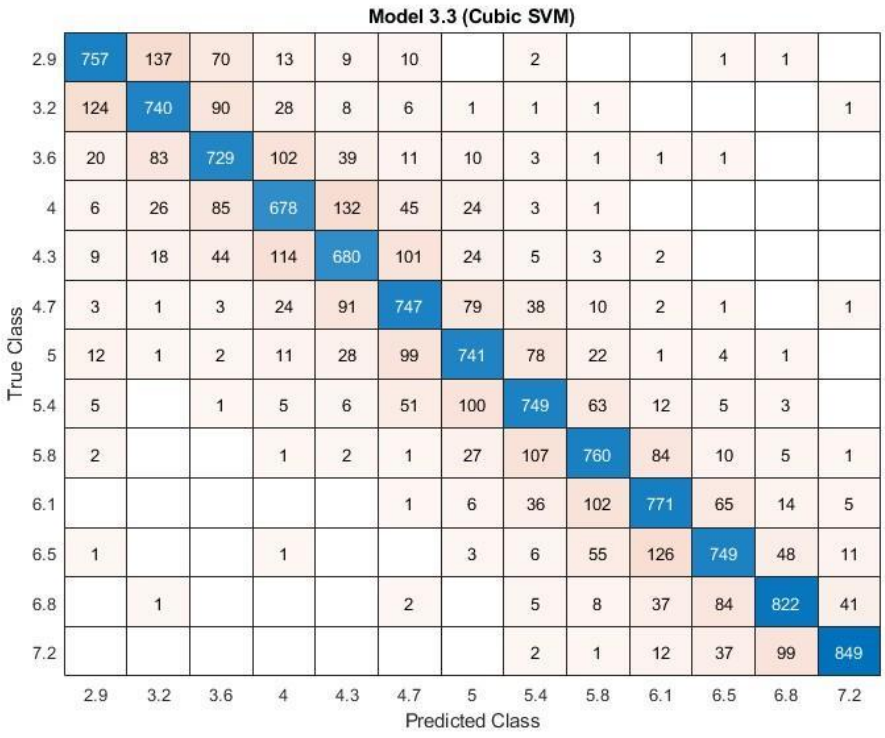

Figure 76. Training confusion matrix for left leg using complete time-series.

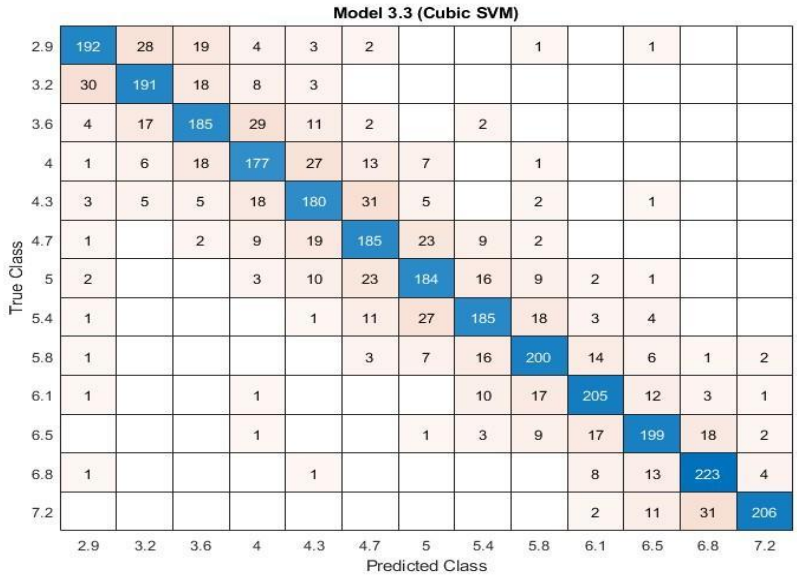

Figure 77. Test confusion matrix for left leg using complete time-series.

These confusion matrices demonstrate the high accuracy and precision of the classification method in detecting speeds specifically for the left leg. The high precision indicates a low rate of misclassifying left leg speeds, ensuring their accurate identification. The TPR plot further supports this by showing high sensitivity in correctly confirming the detection of left leg speed events.

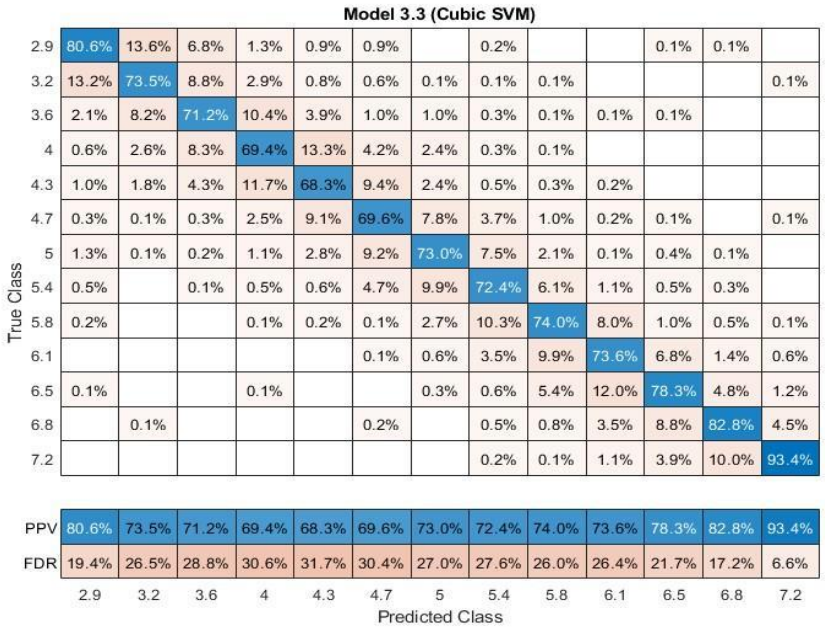

Figure 78. PPV-FDR training confusion matrix for left leg using complete time-series.

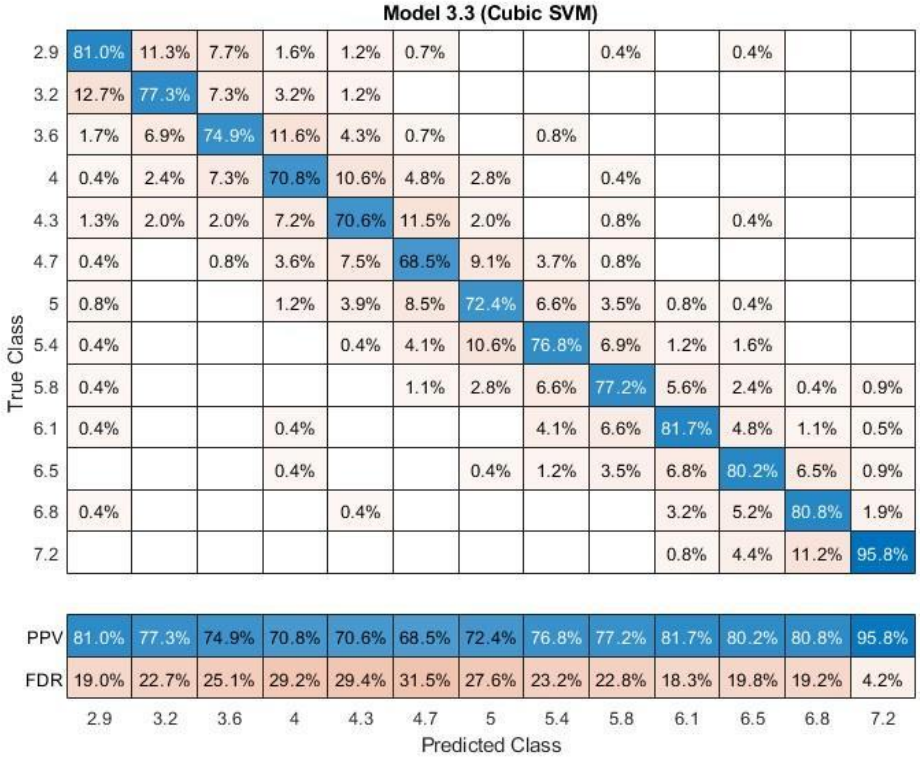

Figure 79. PPV-FDR test confusion matrix for left leg using complete time-series.

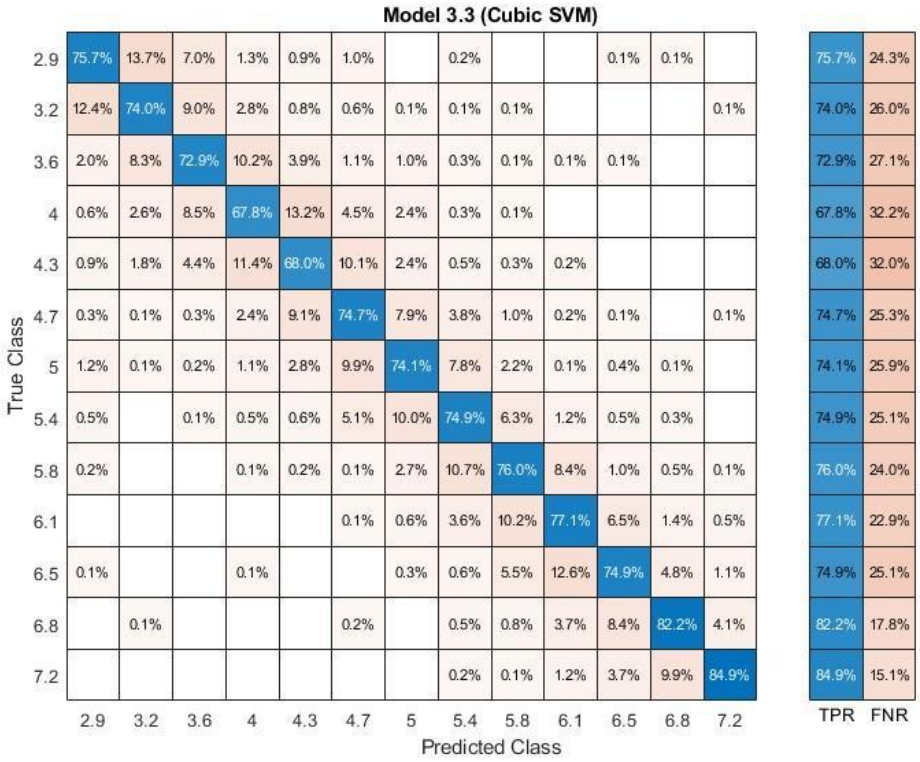

Figure 80. TPR-FNR training confusion matrix for left leg using complete time-series.

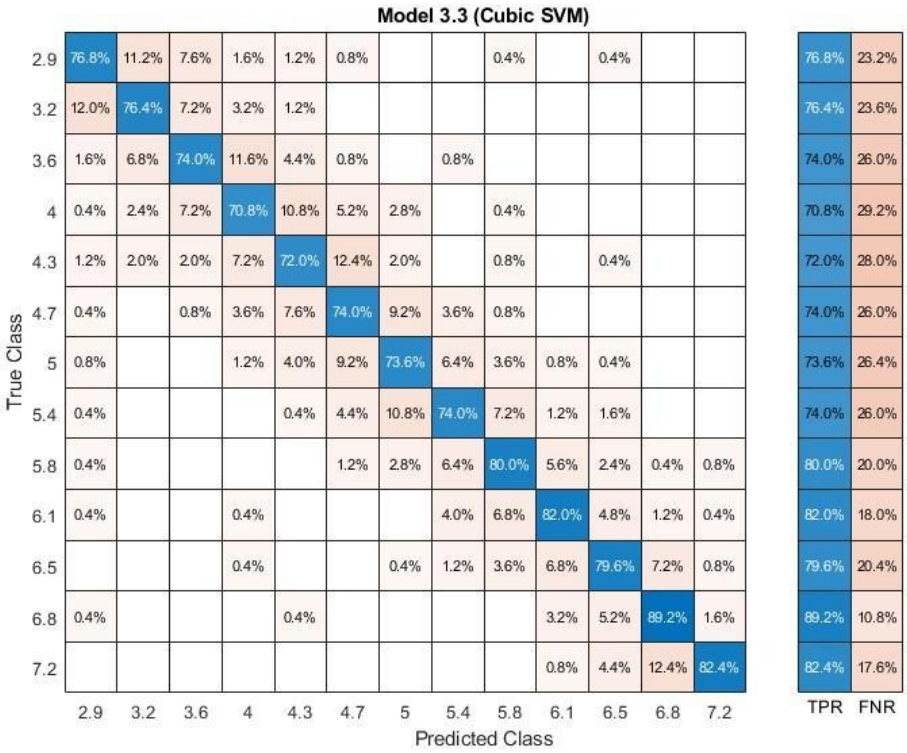

Figure 81. TPR-FNR test confusion matrix for left leg using complete time-series.

For left leg movement classification, the cubic SVM model achieved the highest accuracy, with performance notably increasing from training to the test set.

Table 12

Models' Accuracy For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	75.16923	77.29231
3.2	SVM	71.1	73.13846
4.6	KNN	68.9	72.12308
4.4	KNN	67.76923	70.58462
4.2	KNN	67.31538	70.21538
5.3	Neural Network	66.76923	68.95385
4.1	KNN	65.77692	68.18462
3.5	SVM	65.16923	67.63077
4.5	KNN	63.03846	65.53846
5.2	Neural Network	53.91538	56.33846
3.4	SVM	53.27692	54.8

Table 12 (cont.)

3.1	SVM	47.33077	47.10769
5.5	Neural Network	46.07692	46.98462
5.4	Neural Network	45.84615	48.4
4.3	KNN	45.66154	53.35385
5.1	Neural Network	43.47692	47.78462
3.6	SVM	36.39231	38.55385
2.1	Tree	29.15385	28.64615
2.2	Tree	22.49231	22.8
2.3	Tree	17.35385	17.2

4.2.1.2.2 Skipping speed odd pattern. After applying the skipped speeds preprocessing technique, these confusion matrices show the results of classifying the complete time series data in the left leg. Interestingly, the test confusion matrix exhibited a higher percentage of accurate classifications compared to the training confusion matrix.

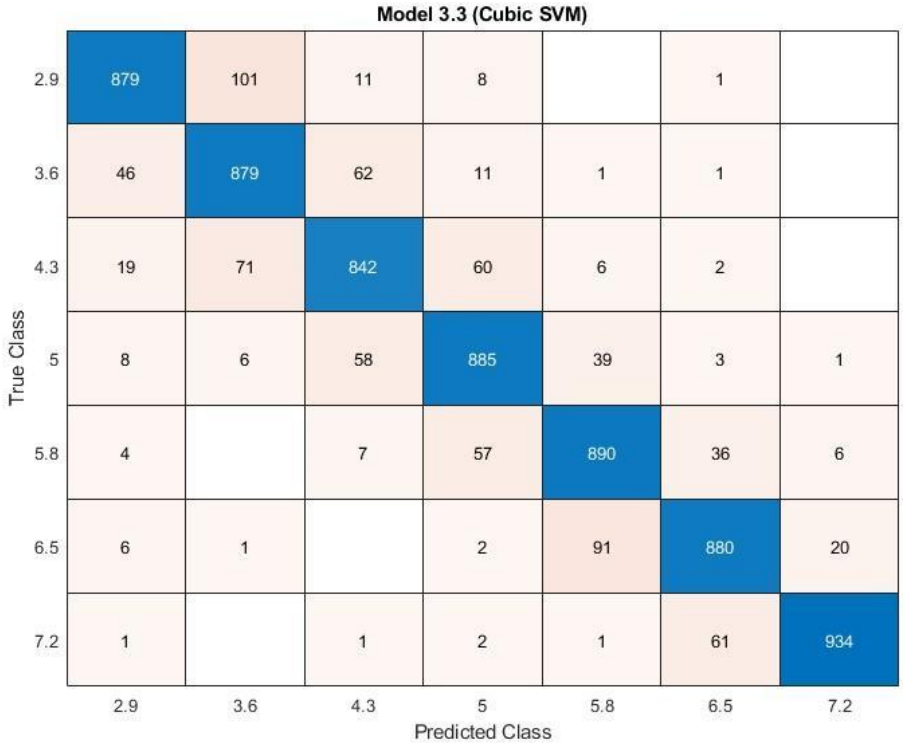

Figure 82. Training confusion matrix of subset 1 for left leg using complete time-series.

Figure 83. Test confusion matrix of subset 1 for left leg using complete time-series.

The confusion matrices illustrate the impressive performance of the classification method in accurately detecting speeds of the left leg, even with the skipped speeds preprocessing. High precision ensures minimal misclassification, while high sensitivity in the TPR plot confirms the reliable detection of left leg speed events.

Figure 84. PPV-FDR training confusion matrix of subset 1 for left leg using complete time-series.

Figure 85. PPV-FDR test confusion matrix of subset 1 for left leg using complete time-series.

Figure 86. TPR-FNR training confusion matrix of subset 1 for left leg using complete time-series.

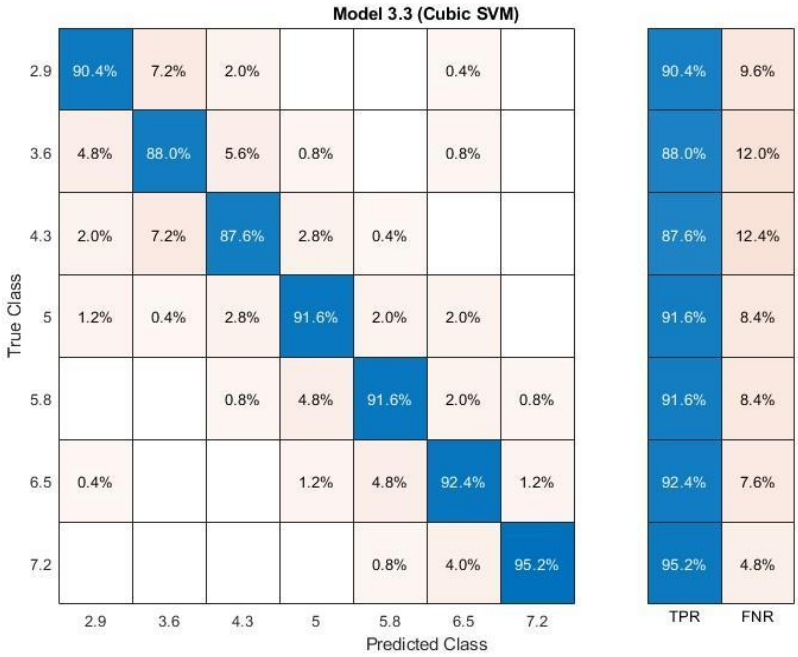

Figure 87. TPR-FNR test confusion matrix of subset 1 for left leg using complete time-series.

In analyzing left leg speed with skipped speeds, the cubic SVM model not only demonstrated high accuracy on the training data but also generalized exceptionally well to the test set, exhibiting even higher accuracy.

Table 13

Models' Accuracy Of Subset 1 For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	88.41429	90.97143
3.2	SVM	85.8	90.17143
5.3	Neural Network	85.1	88.17143
4.6	KNN	84.44286	86.51429
4.1	KNN	83.02857	84.62857
4.2	KNN	82.81429	85.6
4.4	KNN	82.2	86.17143
3.5	SVM	80.6	84.05714
5.2	Neural Network	79.42857	83.88571
4.5	KNN	78.48571	83.08571
5.5	Neural Network	69.2	71.37143
5.4	Neural Network	69.02857	70.97143
5.1	Neural Network	68.38571	68.22857
3.4	SVM	68.08571	70.34286
3.1	SVM	65.01429	66.17143

Table 13 (cont.)

3.6	SVM	54.52857	56.22857
4.3	KNN	53.91429	64.8
2.1	Tree	49.45714	53.37143
2.2	Tree	39.24286	41.02857
2.3	Tree	31.58571	32.68571

4.2.1.2.3 Skipped speeds by even pattern. The figures below demonstrate that, for the left leg, the skipped speeds method achieves a higher degree of speed-specific classification compared to using percentile data.

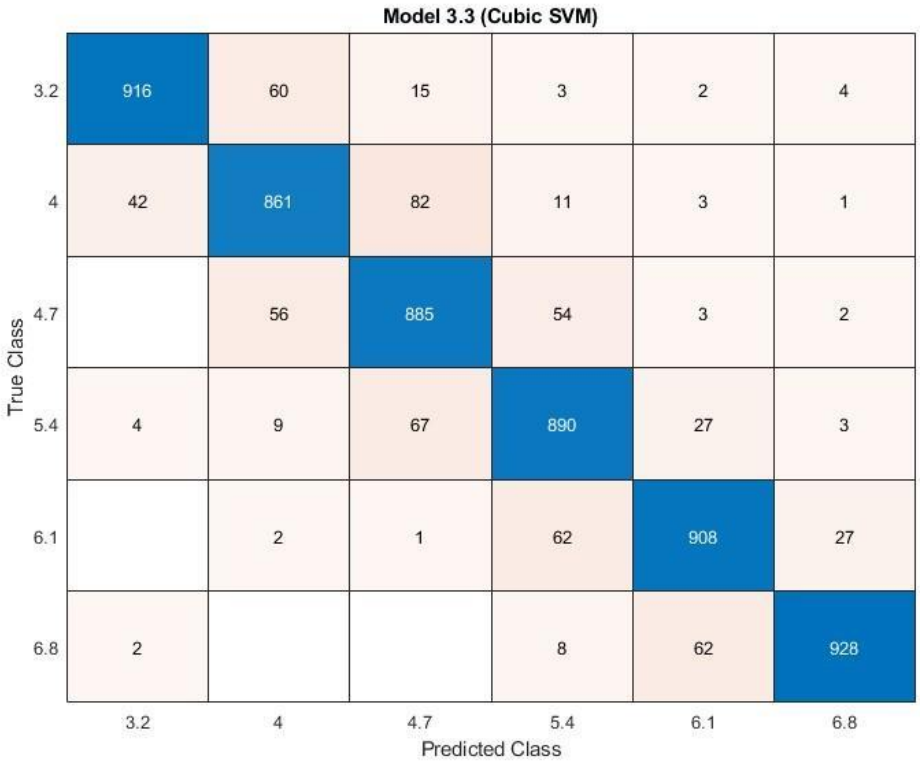

Figure 88. Training confusion matrix of subset 2 for left leg using complete time-series.

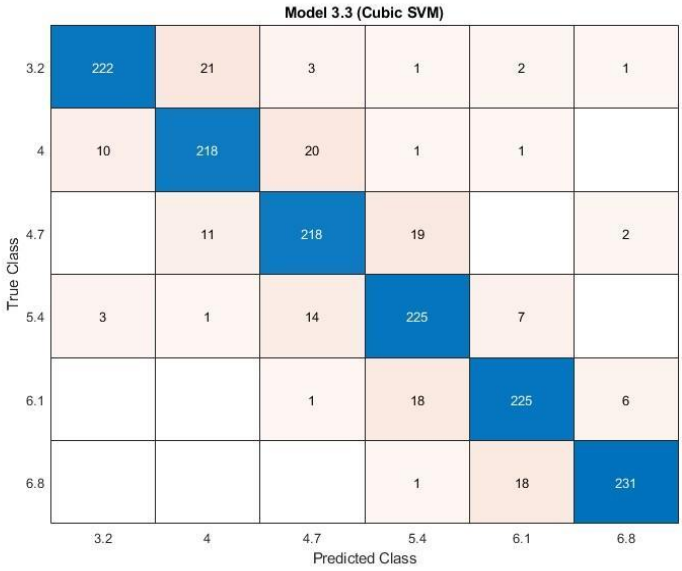

Figure 89. Test confusion matrix of subset 2 for left leg using complete time-series.

The confusion matrices illustrate the impressive performance of the classification method in accurately detecting speeds of the left leg, even with the skipped speeds preprocessing. High precision ensures minimal misclassification, while high sensitivity in the TPR plot confirms the reliable detection of left leg speed events.

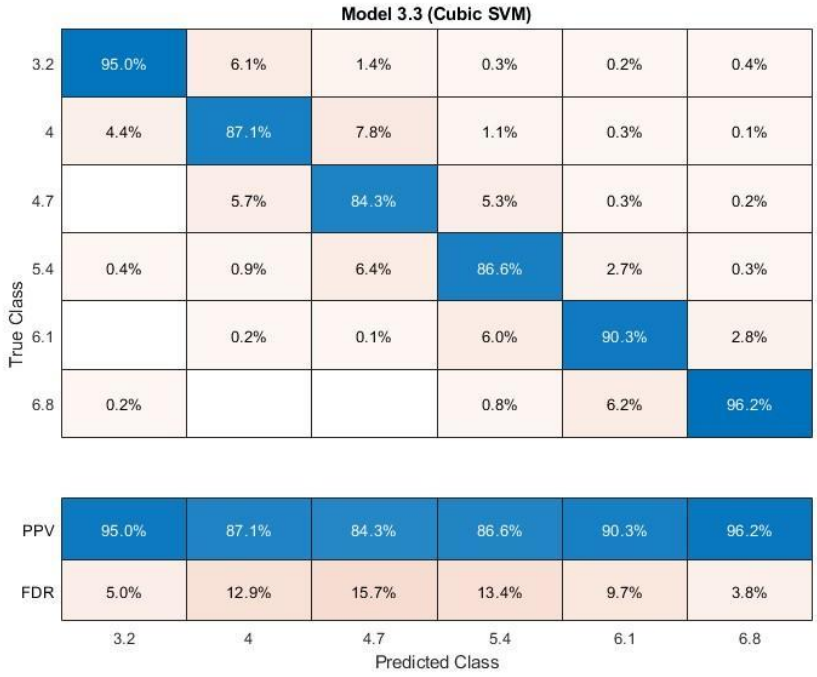

Figure 90. PPV-FDR training confusion matrix of subset 2 for left leg using complete time-series.

Figure 91. PPV-FDR test confusion matrix of subset 2 for left leg using complete time-series.

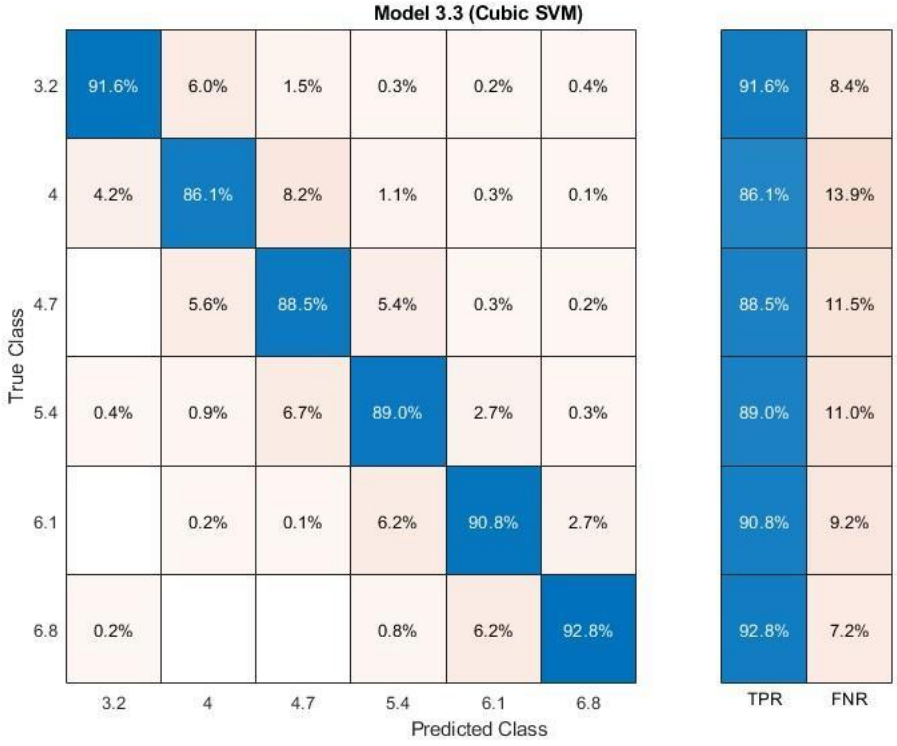

Figure 92. TPR-FNR training confusion matrix of subset 2 for left leg using complete time-series.

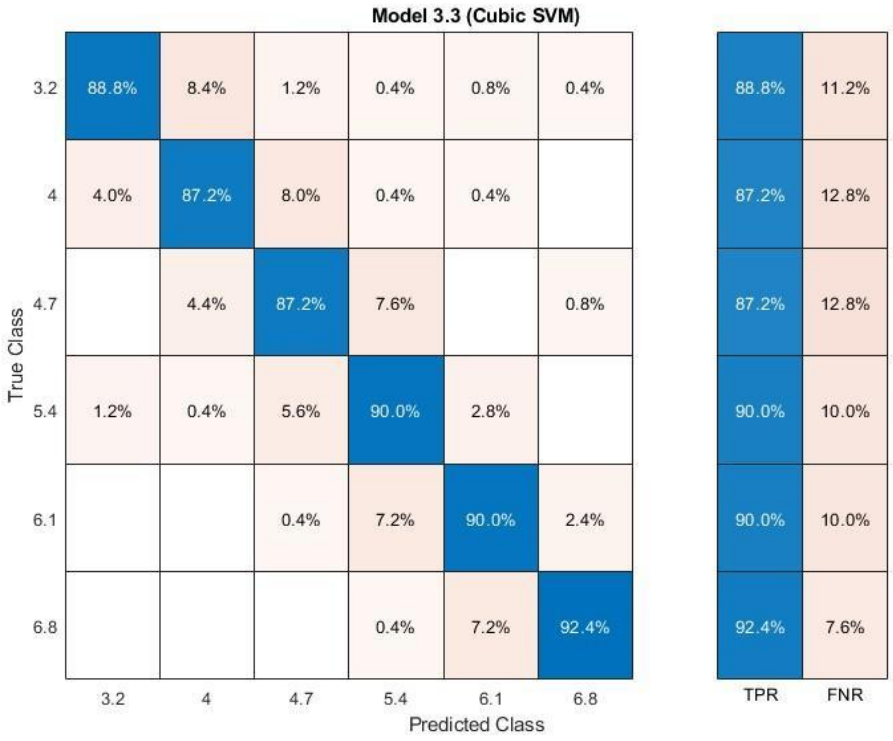

Figure 93. TPR-FNR test confusion matrix of subset 2 for left leg using complete time-series.

When classifying left leg speed using skipped speeds, the cubic SVM model not only demonstrated high accuracy on the training data but also generalized exceptionally well to the test set, exhibiting even higher accuracy.

Table 14

Models' Accuracy Of Subset 2 For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	89.8	89.26667
3.2	SVM	88.01667	88.26667
4.6	KNN	86.58333	86.4
5.3	Neural Network	86.58333	87.46667
4.4	KNN	86.26667	86.13333
4.1	KNN	85.45	86.4
4.2	KNN	85.45	85.73333
3.5	SVM	83.96667	85.2
4.5	KNN	82.1	81.33333
5.2	Neural Network	81.98333	81.46667
5.4	Neural Network	74.9	73.4

Table 14 (cont.)

5.5	Neural Network	73.81667	74.86667
5.1	Neural Network	70.98333	72.6
3.1	SVM	69.85	68.93333
3.4	SVM	68.81667	68.46667
3.6	SVM	58.28333	59.4
4.3	KNN	55.41667	63.46667
2.1	Tree	54.2	52.66667
2.2	Tree	43.88333	42.6
2.3	Tree	34.55	35.6

4.2.1.2.4 Remaining frames 50. Using only 50 remaining frames from the complete time series data for the left leg, these confusion matrices illustrate successful classification. Remarkably, the test confusion matrix showed a higher percentage of correct classifications than the training confusion matrix.

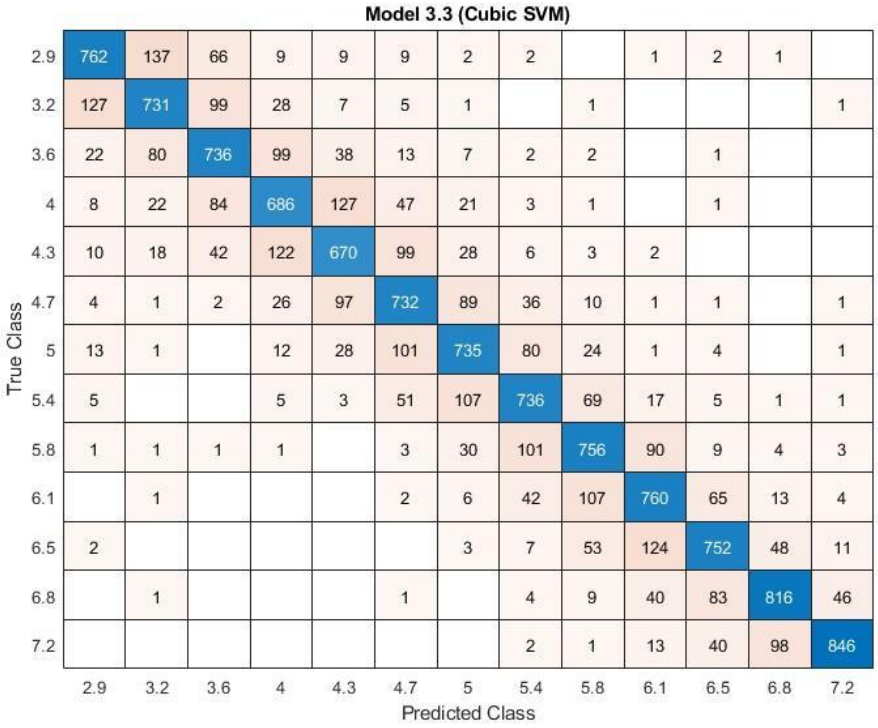

Figure 94. Training confusion matrix of 50 remaining frames for left leg using complete time-series.

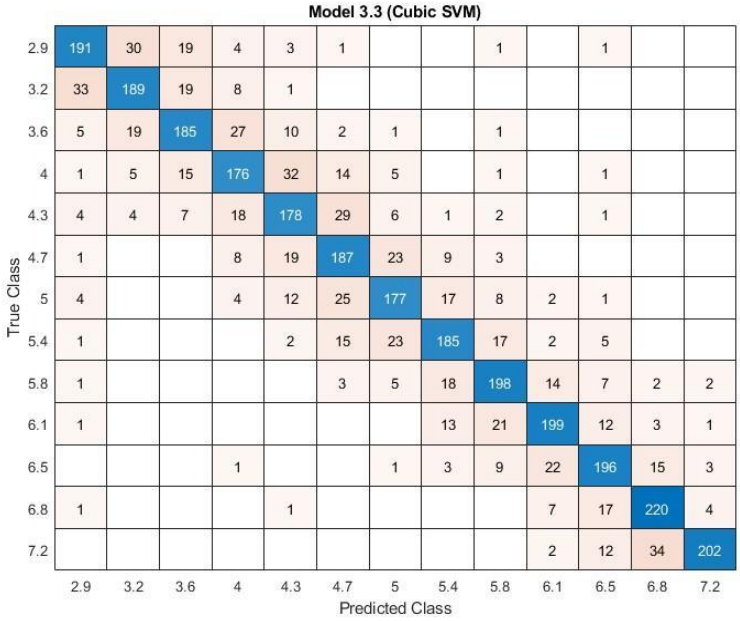

Figure 95. Test confusion matrix of 50 remaining frames for left leg using complete time-series.

Even with only 50 remaining frames, these confusion matrices demonstrate the high accuracy and precision of the classification method in detecting left leg speeds.

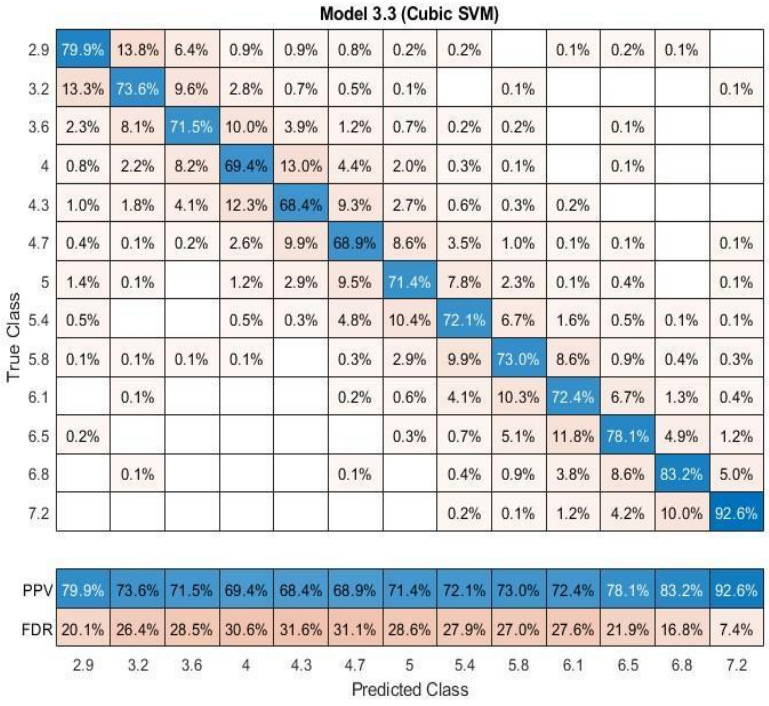

Figure 96. PPV-FDR training confusion matrix of 50 remaining frames for left leg using complete time-series.

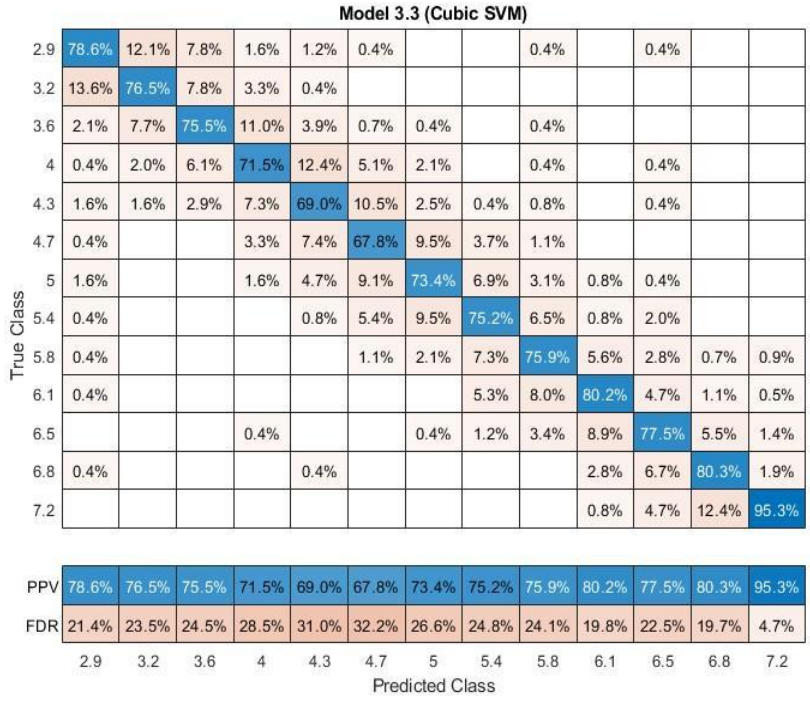

Figure 97. PPV-FDR test confusion matrix of 50 remaining frames for left leg using complete time-series.

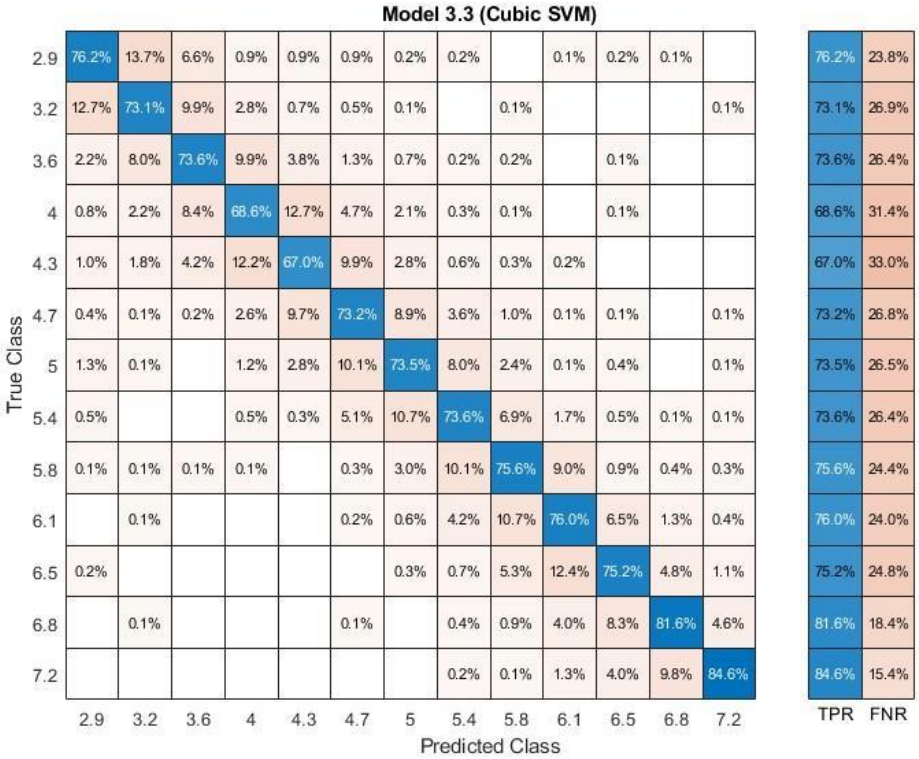

Figure 98. TPR-FNR training confusion matrix of 50 remaining frames for left leg using complete time-series.

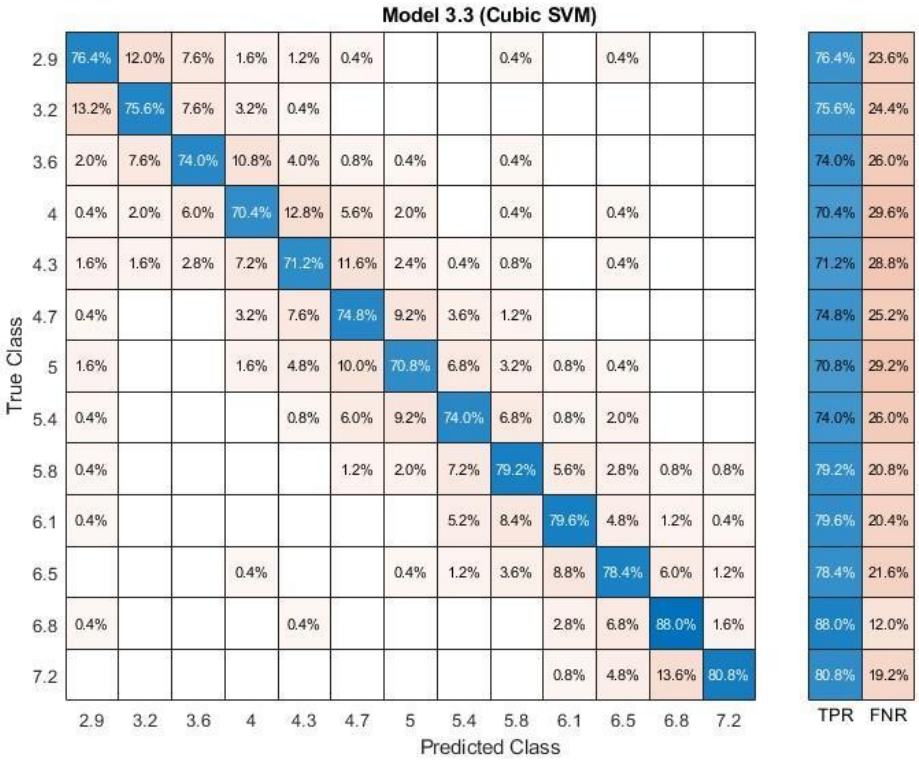

Figure 99. TPR-FNR test confusion matrix of 50 remaining frames for left leg using complete time-series.

Using just 50 remaining frames for left leg analysis, the cubic SVM model delivered the highest accuracy, demonstrating its efficiency and generalization ability.

Table 15

Models' Accuracy Of 50 Remaining Frames For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	74.75385	76.4
3.2	SVM	70.53077	72.64615
4.6	KNN	68.77692	71.38462
4.4	KNN	67.20769	70.4
5.3	Neural Network	67.07692	69.56923
4.2	KNN	66.92308	70.33846
4.1	KNN	65.52308	67.63077
3.5	SVM	64.11538	66.33846
4.5	KNN	62.55385	65.32308
5.2	Neural Network	55.10769	58.24615

Table 15 (cont.)

3.4	SVM	53.27692	55.29231
5.4	Neural Network	47.08462	50.4
5.5	Neural Network	46.9	49.44615
3.1	SVM	46.72308	47.26154
5.1	Neural Network	46.35385	47.04615
4.3	KNN	45.36923	52.8
3.6	SVM	36.3	38.33846
2.1	Tree	28.90769	30.73846
2.2	Tree	22.33077	22.86154
2.3	Tree	17.3	17.41538

4.2.1.2.5 Remaining frames 25. Despite utilizing only 25 remaining frames for left leg movement classification, the complete time series data was effectively classified, as shown in these confusion matrices. Unexpectedly, the test confusion matrix exhibited a higher percentage of accurate classifications compared to the training confusion matrix.

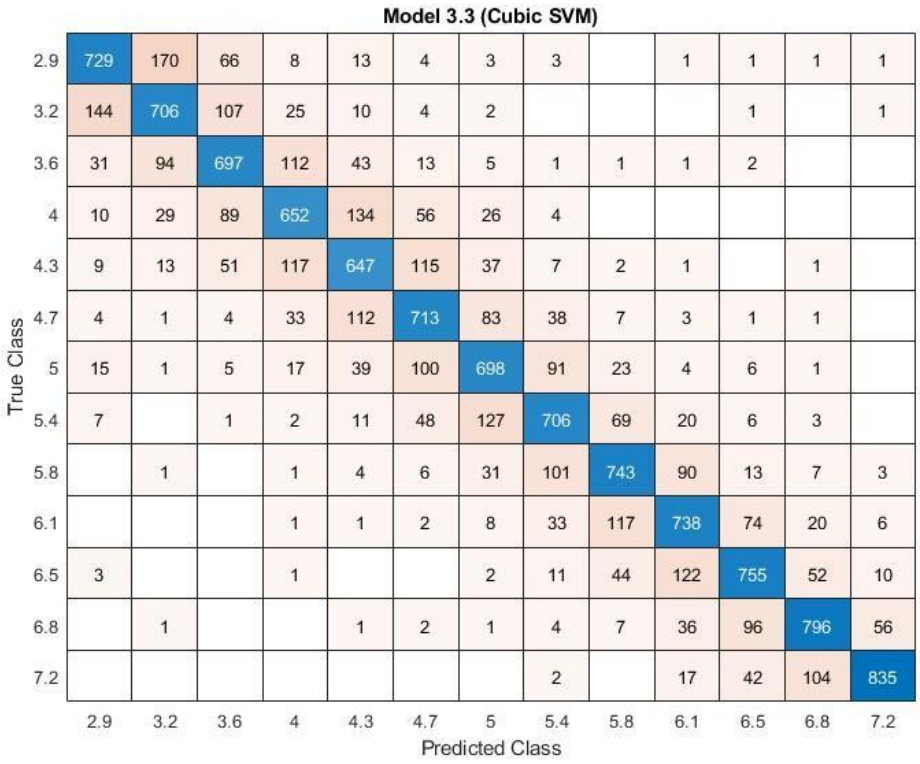

Figure 100. Training confusion matrix of 25 remaining frames for left leg using complete time-series.

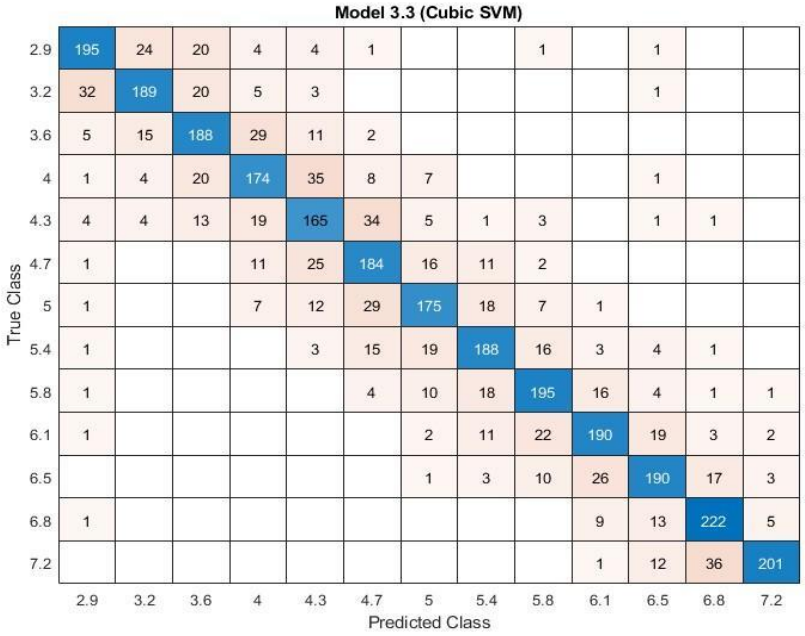

Figure 101. Test confusion matrix of 25 remaining frames for left leg using complete time-series.

The confusion matrices illustrate the impressive performance of the classification method in accurately detecting speeds of the left leg using only 25 remaining frames.

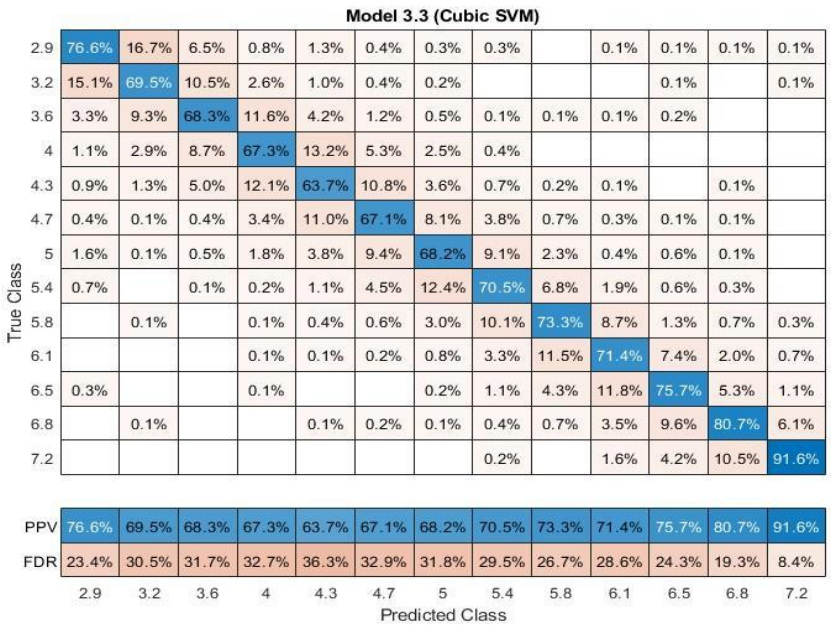

Figure 102. PPV-FDR training confusion matrix of 25 remaining frames for left leg using complete time-series.

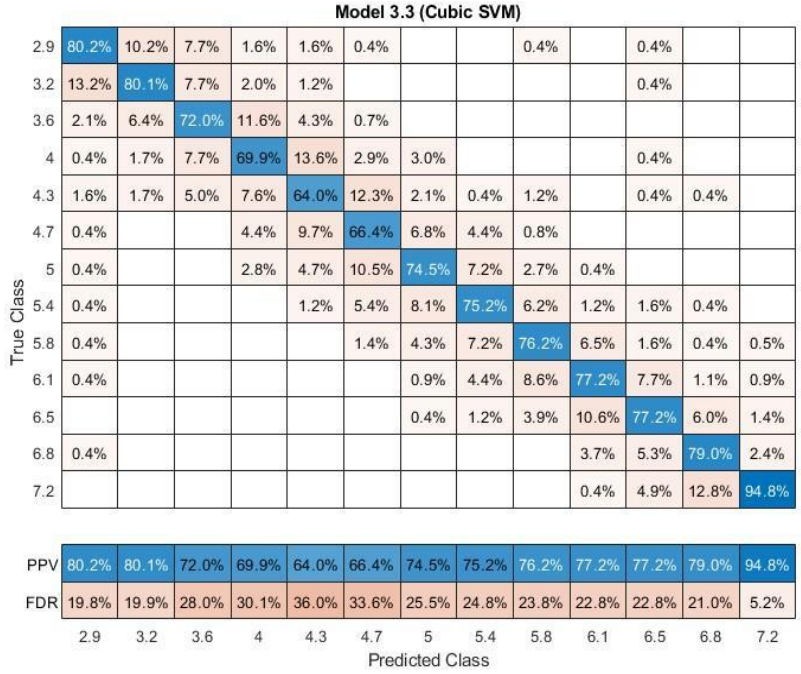

Figure 103. PPV-FDR test confusion matrix of 25 remaining frames for left leg using complete time-series.

Figure 104. TPR-FNR training confusion matrix of 25 remaining frames for left leg using complete time-series.

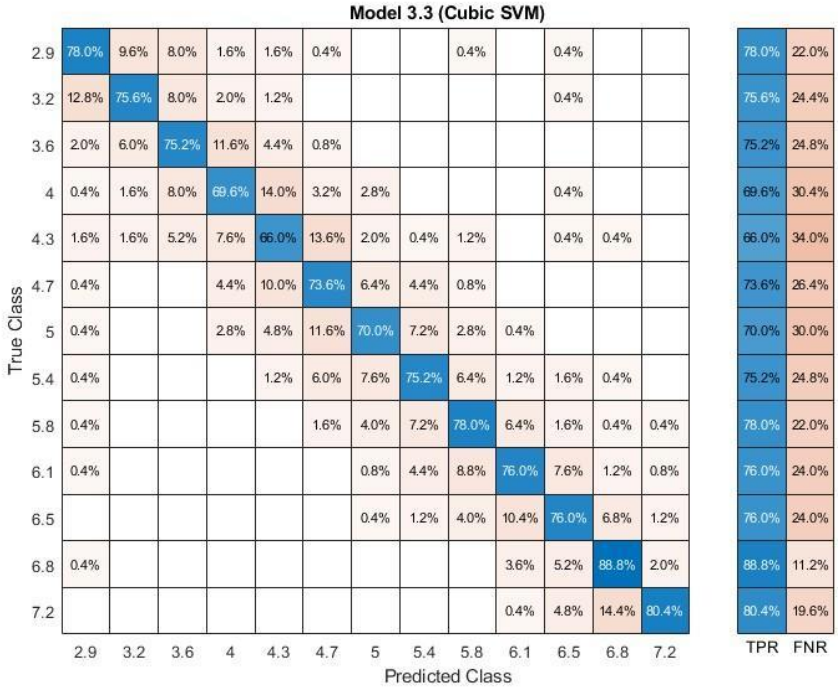

Figure 105. TPR-FNR test confusion matrix of 25 remaining frames for left leg using complete time-series.

In classifying left leg movement, the cubic SVM model not only demonstrated high accuracy on the training data but also generalized exceptionally well to the test set, exhibiting even higher accuracy.

Table 16

Models' Accuracy Of 25 Remaining Frames For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	72.42308	75.56923
3.2	SVM	68.66923	70.76923
4.6	KNN	67.76923	69.84615
4.2	KNN	66.18462	68.4
4.4	KNN	65.66154	68.61538
5.3	Neural Network	65.03077	68.46154
4.1	KNN	64.11538	64.89231
3.5	SVM	63.06154	66.36923
4.5	KNN	61.74615	63.66154

Table 16 (cont.)

5.2	Neural Network	56.20769	60.61538
3.4	SVM	52.73846	54.64615
5.4	Neural Network	47.66923	49.96923
5. 5	Neural Network	47.55385	50.55385
5.1	Neural Network	46.8	45.01538
4.3	KNN	44.71538	51.23077
3.1	SVM	44.27692	44.4
3.6	SVM	35.66154	37.78462
2.1	Tree	28.06923	28.61538
2.2	Tree	21.76923	22.15385
2.3	Tree	17.20769	16.33846

4.2.1.2.6 Remaining frames by 20. Despite utilizing only 20 remaining frames for left leg movement classification, the complete time series data was effectively classified, as shown in these confusion matrices. Unexpectedly, the test confusion matrix exhibited a higher percentage of accurate classifications compared to the training confusion matrix.

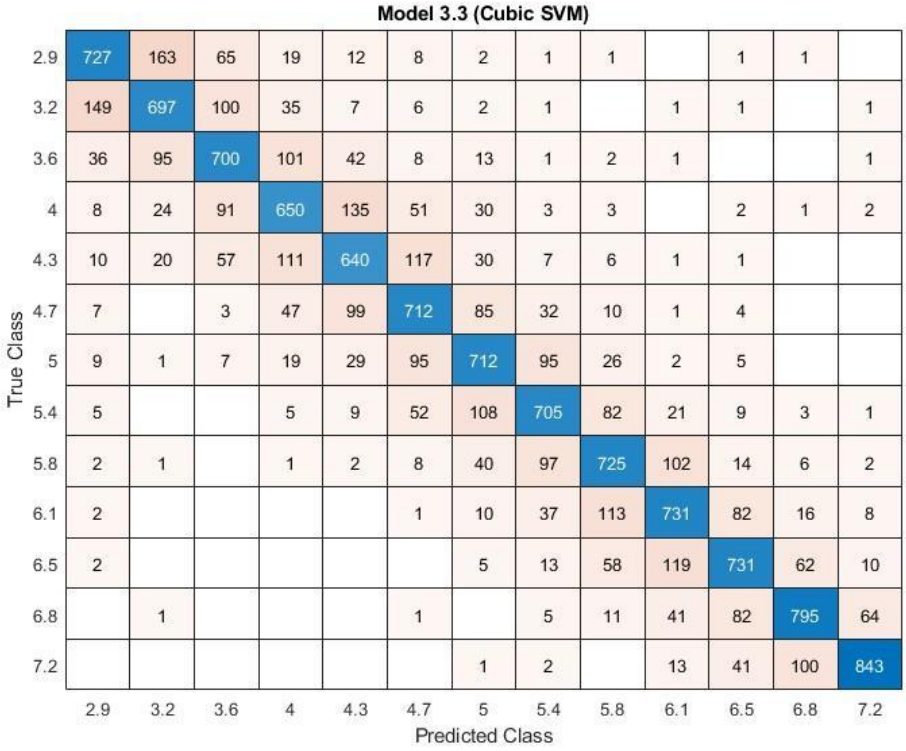

Figure 106. Training confusion matrix of 20 remaining frames for left leg using complete time-series.

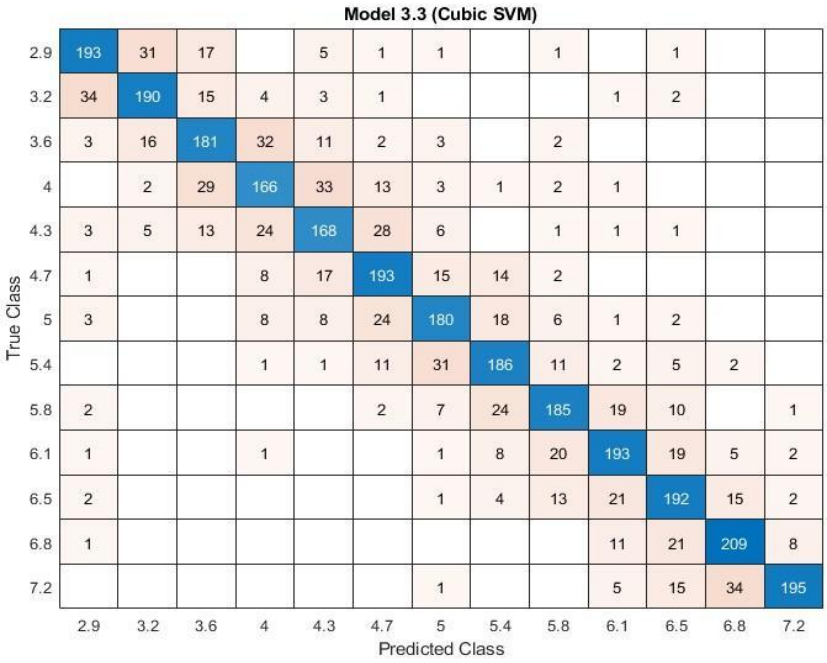

Figure 107. Test confusion matrix of 20 remaining frames for left leg using complete time-series.

The confusion matrices illustrate the impressive performance of the classification method in accurately detecting speeds of the left leg using only 20 remaining frames and showing parallel accuracy with the right leg.

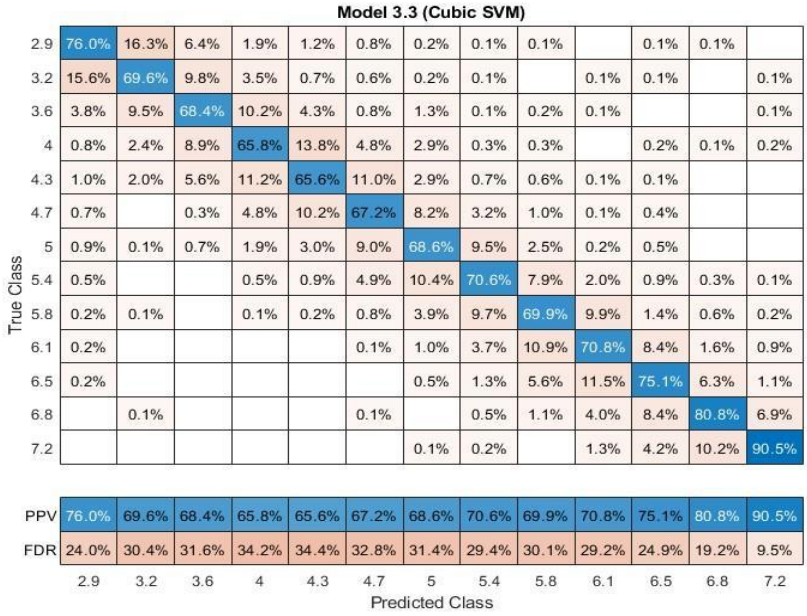

Figure 108. PPV-FDR training confusion matrix of 20 remaining frames for left leg using complete time-series.

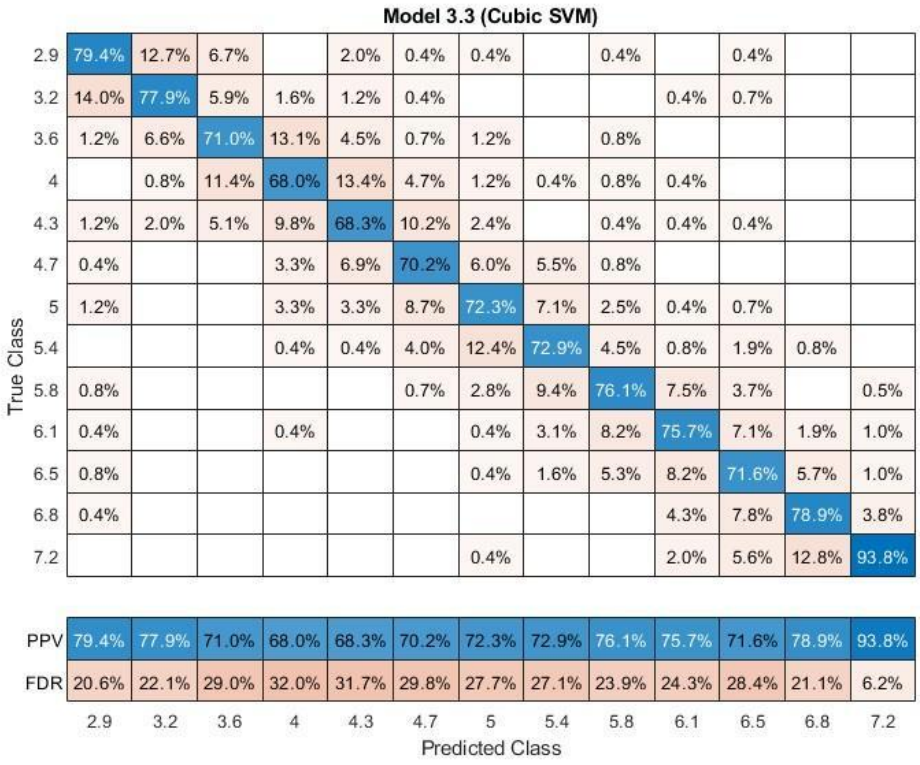

Figure 109. PPV-FDR test confusion matrix of 20 remaining frames for left leg using complete time-series.

Figure 110. TPR-FNR training confusion matrix of 20 remaining frames for left leg using complete time-series.

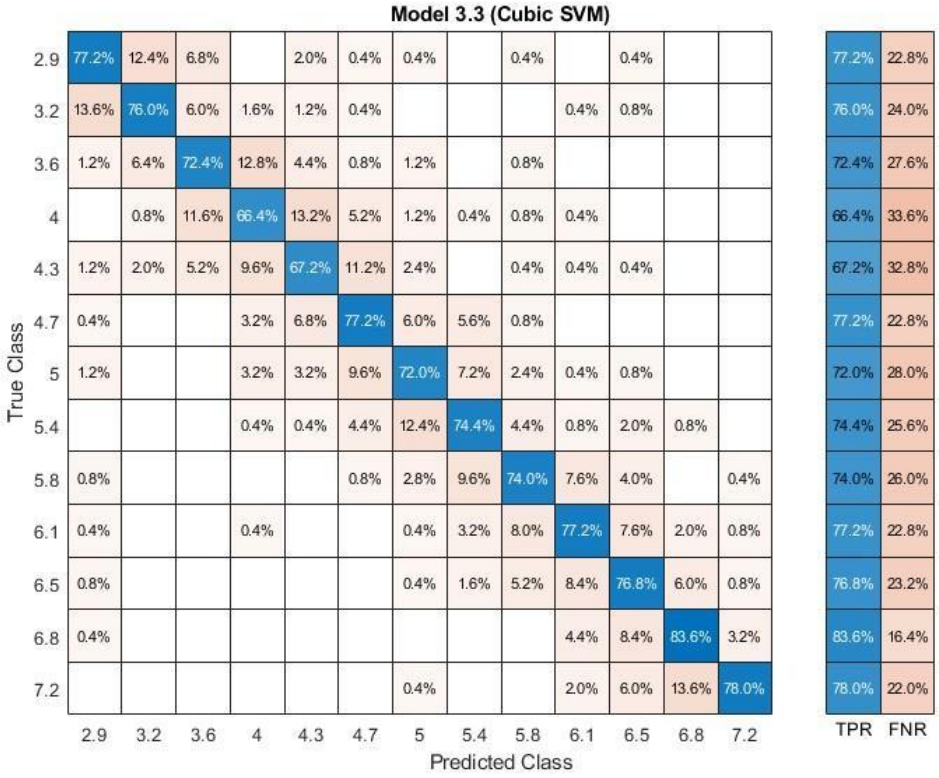

Figure 111. TPR-FNR test confusion matrix of 20 remaining frames for left leg using complete time-series.

The cubic SVM model's exceptional performance in classifying left leg movement with just 20 frames—showing increased accuracy on the test set—suggests its potential for accurate and efficient gait analysis in situations where data availability is limited.

Table 17

Models' Accuracy Of 20 Remaining Frames For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	72.06154	74.8
3.2	SVM	67.75385	71.41538
4.6	KNN	65.75385	68.4
4.2	KNN	64.42308	66.89231
5.3	Neural Network	63.7	65.32308
4.4	KNN	63.63846	66.09231
4.1	KNN	62.32308	64.09231
3.5	SVM	62.06923	64.30769

Table 17 (cont.)

4.5	KNN	60.22308	62.15385
5. 2	Neural Network	56.89231	59.07692
3.4	SVM	52.62308	53.44615
5.4	Neural Network	45.8	44.15385
5.5	Neural Network	45.7	46.43077
4.3	KNN	45.65385	52.03077
5.1	Neural Network	45.42308	45.72308
3.1	SVM	42.94615	43.23077
3.6	SVM	35.36923	36.64615
2.1	Tree	27.91538	29.6
2.2	Tree	21.79231	22.24615
2.3	Tree	17	17.56923

4.2.1.2.7 Remaining frames 10. Even with just 10 remaining frames from the complete time series data for the left leg, these confusion matrices illustrate successful classification. Remarkably, the test confusion matrix showed a higher percentage of correct classifications than the training confusion matrix.

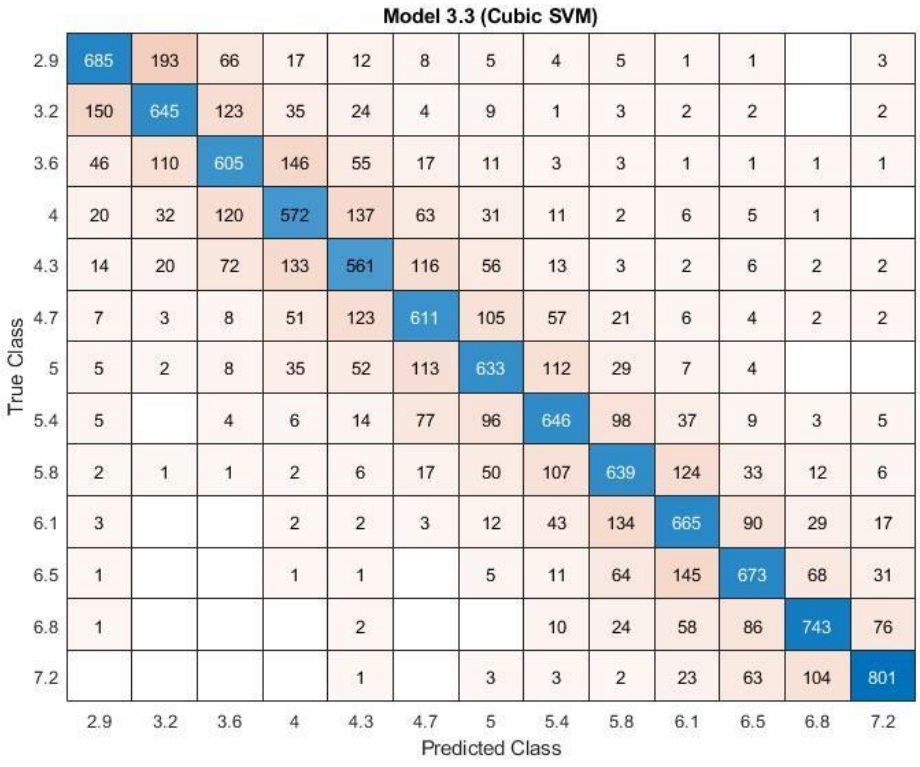

Figure 112. Training confusion matrix of 10 remaining frames for left leg using complete time-series.

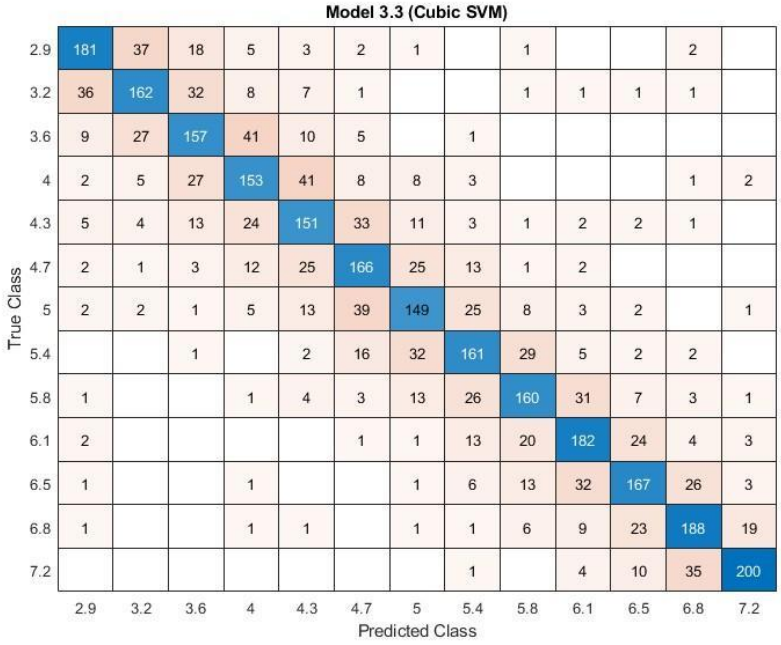

Figure 113. Test confusion matrix of 10 remaining frames for left leg using complete time-series.

Even with only 10 remaining frames, these confusion matrices demonstrate the high accuracy and precision of the classification method in detecting left leg speeds.

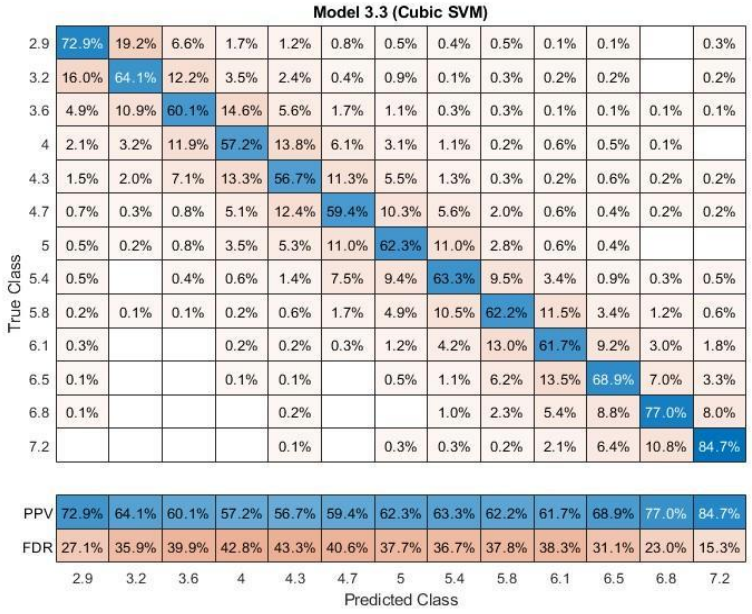

Figure 114. PPV-FDR training confusion matrix of 10 remaining frames for left leg using complete time-series.

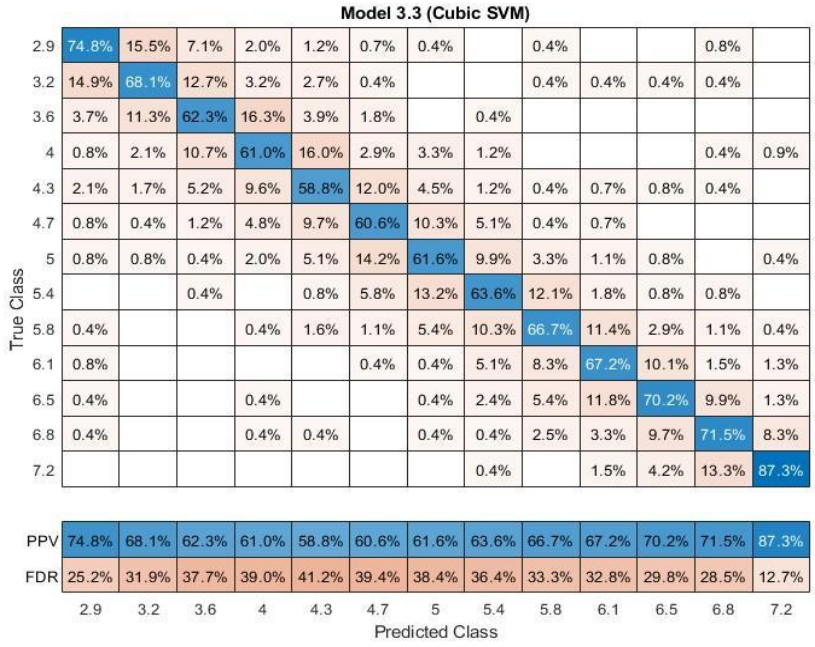

Figure 115. PPV-FDR test confusion matrix of 10 remaining frames for left leg using complete time-series.

Figure 116. TPR-FNR training confusion matrix of 10 remaining frames for left leg using complete time-series.

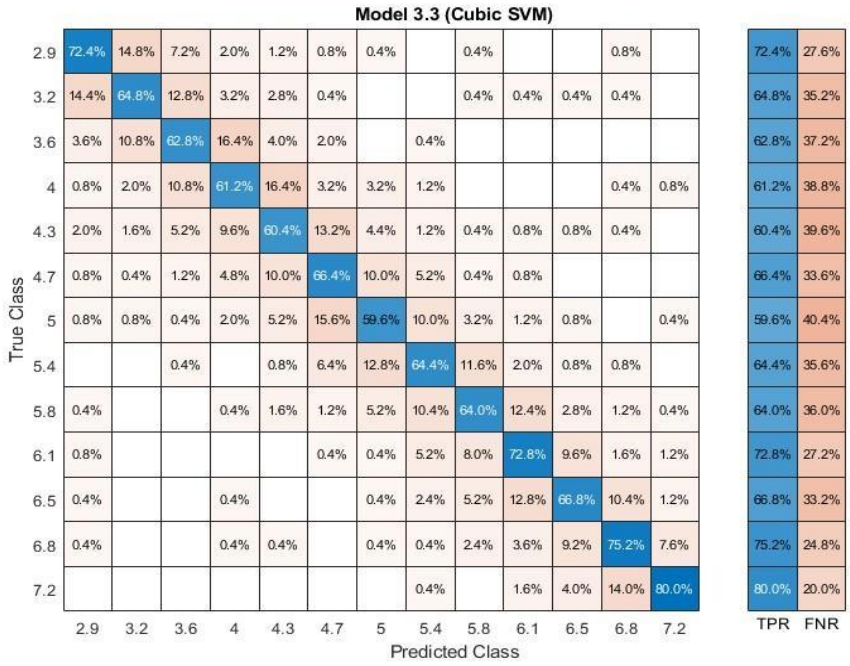

Figure 117. TPR-FNR test confusion matrix of 10 remaining frames for left leg using complete time-series.

Using just 10 remaining frames of left leg movement, the cubic SVM model delivered the highest accuracy, demonstrating its efficiency and generalization ability.

Table 18

Models' Accuracy Of 10 Remaining Frames For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	65.22308	66.98462
3.2	SVM	60.26923	64.70769
4.6	KNN	59.86923	60.95385
4.2	KNN	57.97692	58.76923
4.4	KNN	57.36154	58.95385
5.3	Neural Network	56.78462	60.67692
4.1	KNN	55.81538	57.2
4.5	KNN	54.28462	55.13846
3.5	SVM	54.2	58.70769
3.4	SVM	50.61538	52.86154
5.2	Neural Network	50.43077	50.46154
4.3	KNN	42.70769	47.75385
5.5	Neural Network	40.23077	39.04615

Table 18 (cont.)

5.4	Neural Network	39.17692	43.38462
5.1	Neural Network	38.16923	38.61538
3.1	SVM	34.52308	34.89231
3.6	SVM	31.38462	32.95385
2.1	Tree	26.06923	27.35385
2.2	Tree	21.30769	20.83077
2.3	Tree	16.50769	16.67692

4.2.1.2.8 Remaining frames 5. With just 5 frames of left leg data, the skipped frames method showed lower accuracy, differing significantly from the skipped speeds and full time series approaches.

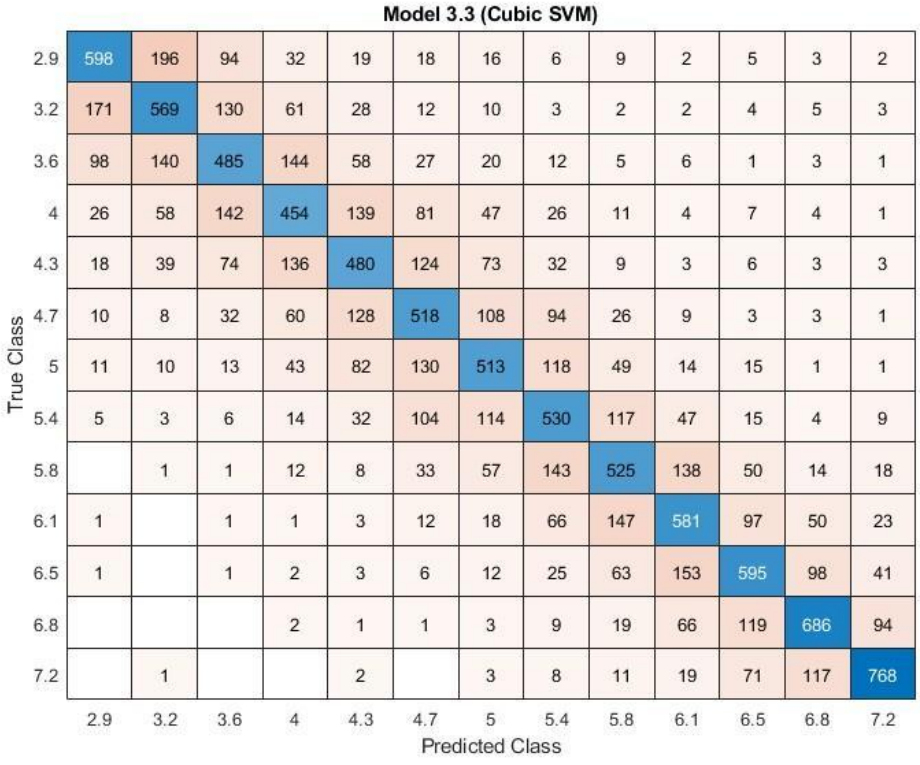

Figure 118. Training confusion matrix of 5 remaining frames for left leg using complete time-series.

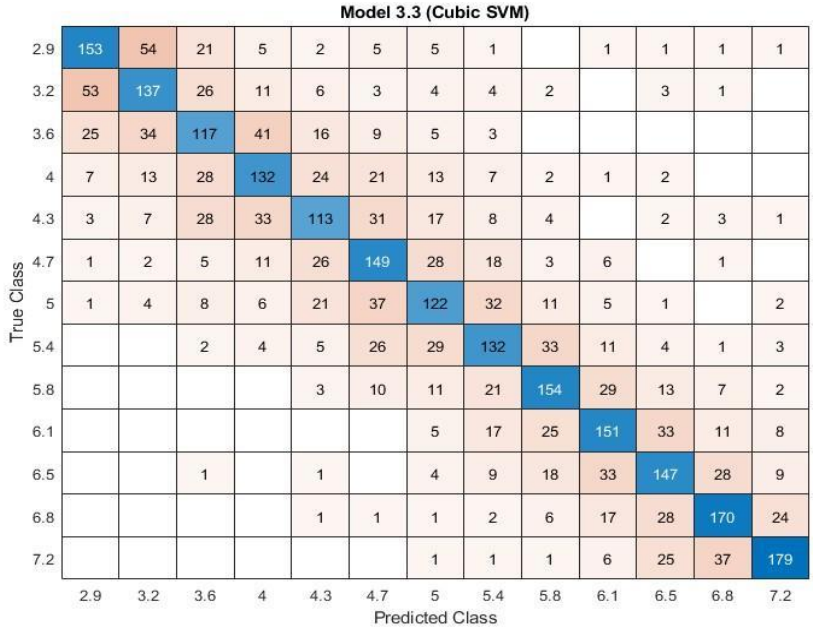

Figure 119. Test confusion matrix of 5 remaining frames for left leg using complete time-series.

Even with only 5 remaining frames, these confusion matrices demonstrate the high accuracy and precision of the classification method in detecting left leg speeds.

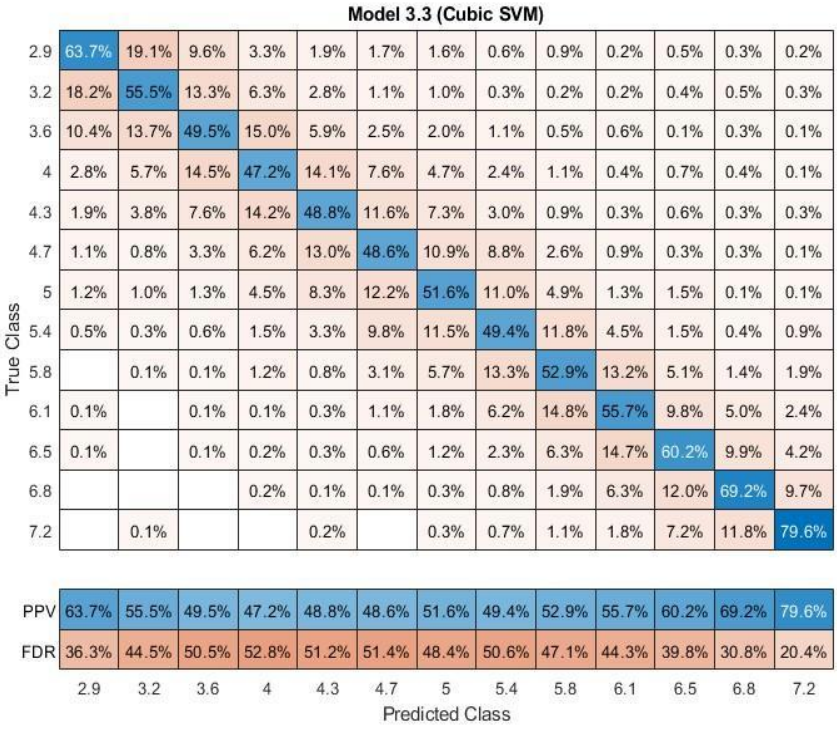

Figure 120. PPV-FDR training confusion matrix of 5 remaining frames for left leg using complete time-series.

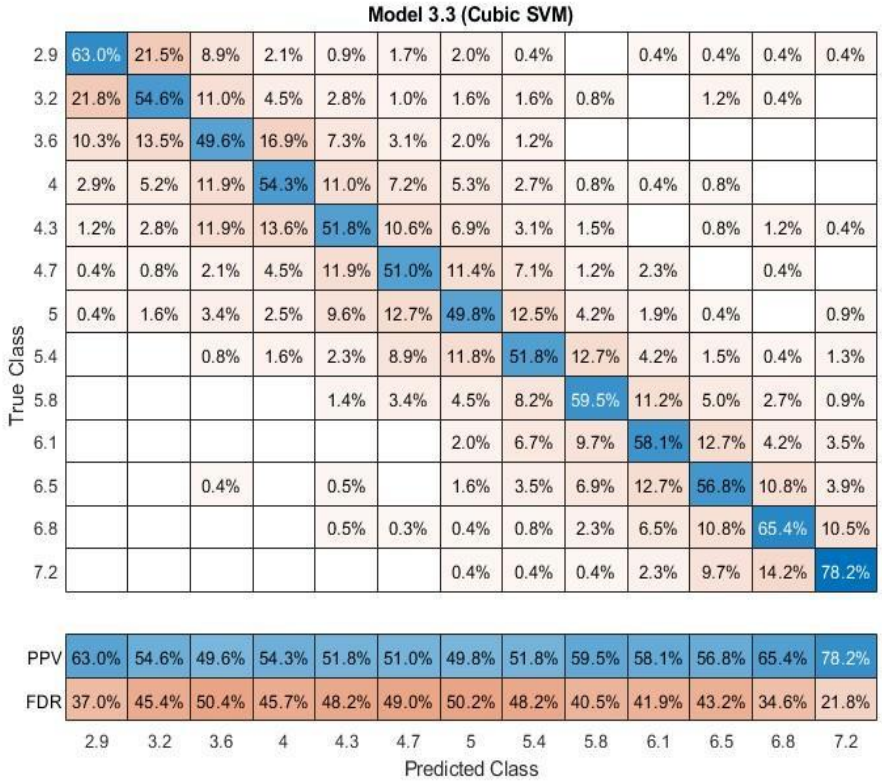

Figure 121. PPV-FDR test confusion matrix of 5 remaining frames for left leg using complete time-series.

Figure 122. TPR-FNR training confusion matrix of 5 remaining frames for left leg using complete time-series.

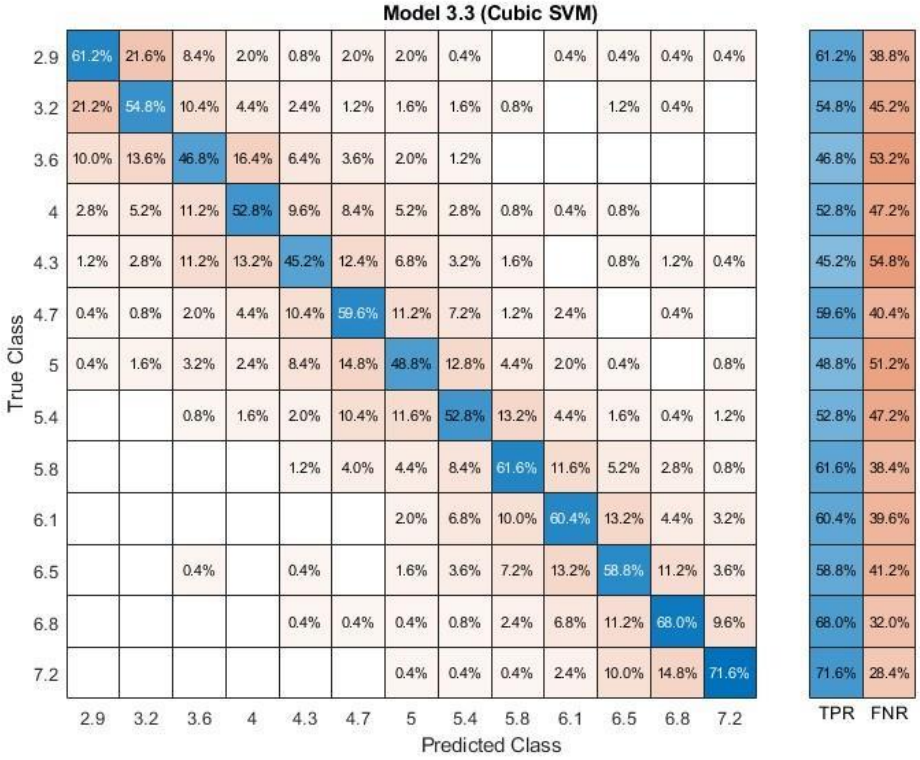

Figure 123. TPR-FNR test confusion matrix of 5 remaining frames for left leg using complete time-series.

Using just 5 remaining frames of left leg movement, the cubic SVM model delivered the highest accuracy, demonstrating its efficiency and generalization ability.

Table 19

Models' Accuracy Of 5 Remaining Frames For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	56.16923	57.10769
4.6	KNN	52.35385	55.53846
5.3	Neural Network	51.27692	52.95385
4.2	KNN	50.40769	53.29231
4.4	KNN	49.42308	50.70769
3.2	SVM	49.19231	51.2
4.1	KNN	48.60769	50.15385
4.5	KNN	47.80769	50.03077
3.4	SVM	47.51538	50.86154

Table 19 (cont.)

3.5	SVM	45.17692	48.67692
5.2	Neural Network	40.68462	42.86154
4.3	KNN	39.33846	43.81538
5.4	Neural Network	33.95385	34.49231
5.5	Neural Network	33.50769	32.92308
5.1	Neural Network	31.72308	31.84615
2.1	Tree	26.40769	26.27692
3.1	SVM	25.84615	26.15385
3.6	SVM	25.72308	27.10769
2.2	Tree	20.39231	19.93846
2.3	Tree	16.2	15.96923

4.2.1.2.9 Gait segmentation-first phase (stance phase). These confusion matrices illustrate the classification results obtained using the complete time series data. Notably, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

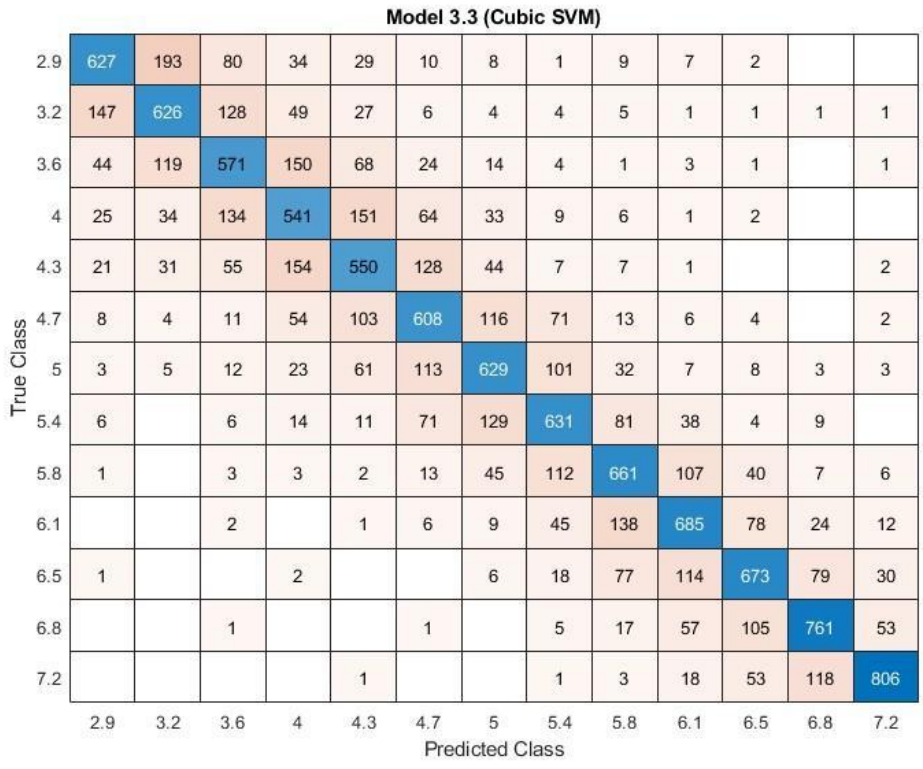

Figure 124. Training confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

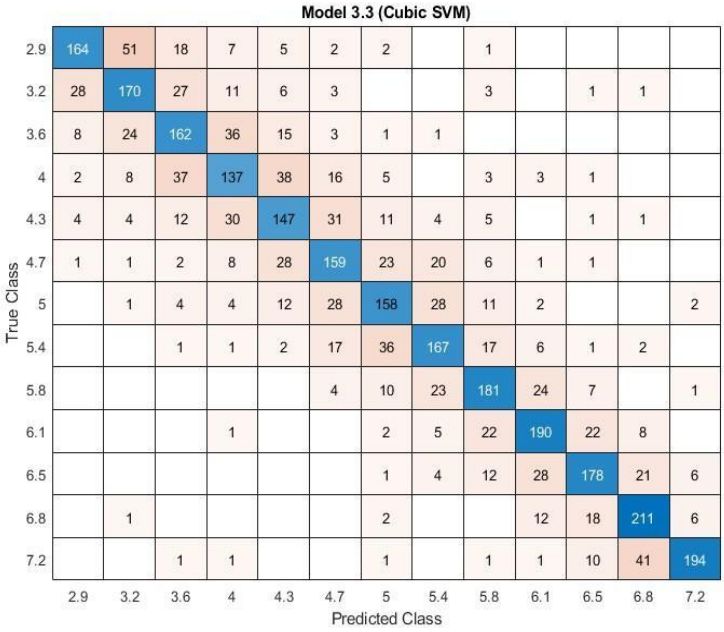

Figure 125. Test confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

These confusion matrices demonstrate the high accuracy and precision of the classification method in identifying the stance phase of gait.

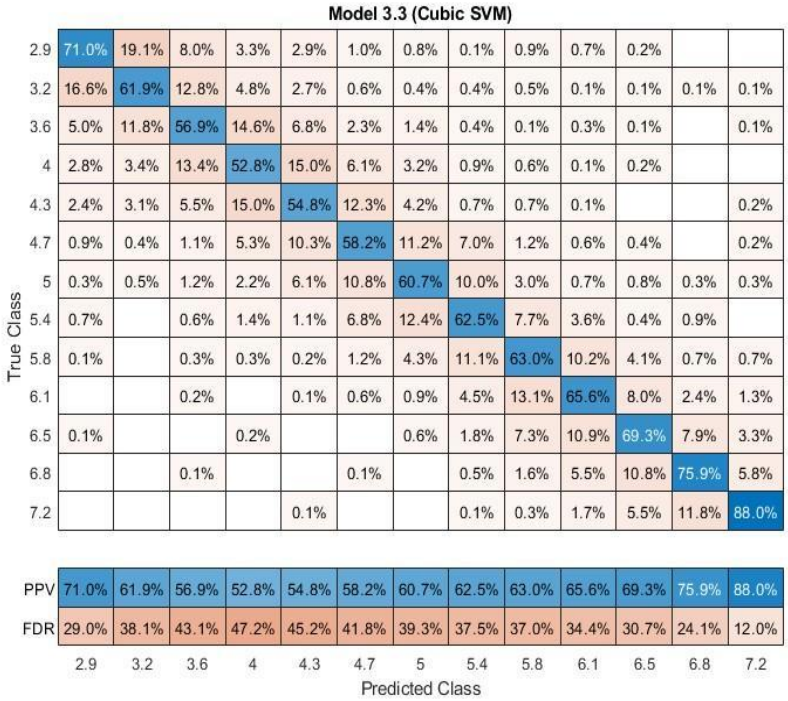

Figure 126. PPV-FDR training confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

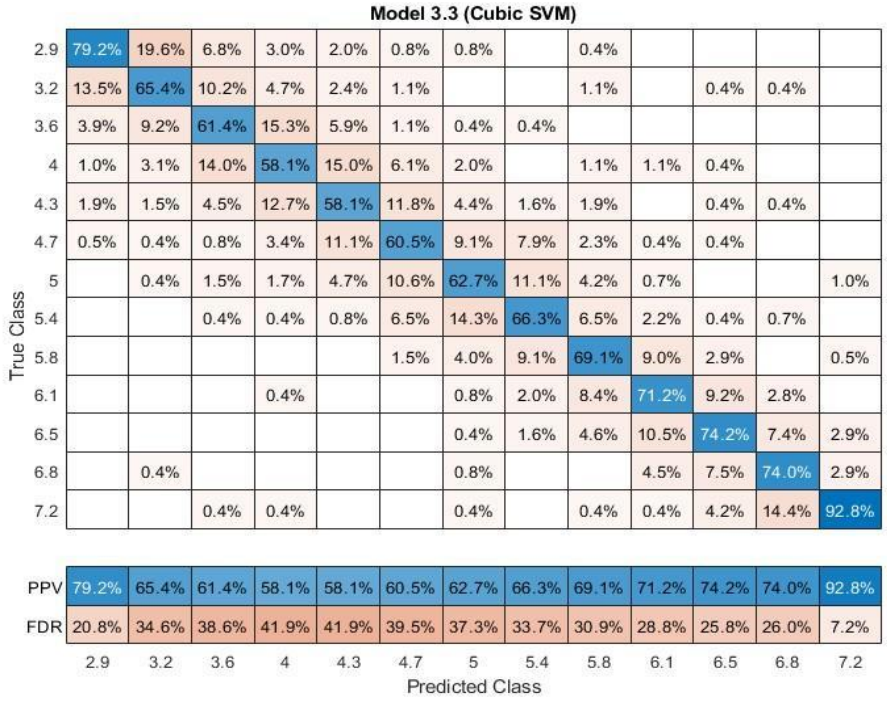

Figure 127. PPV-FDR test confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

Figure 128. TPR-FNR training confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

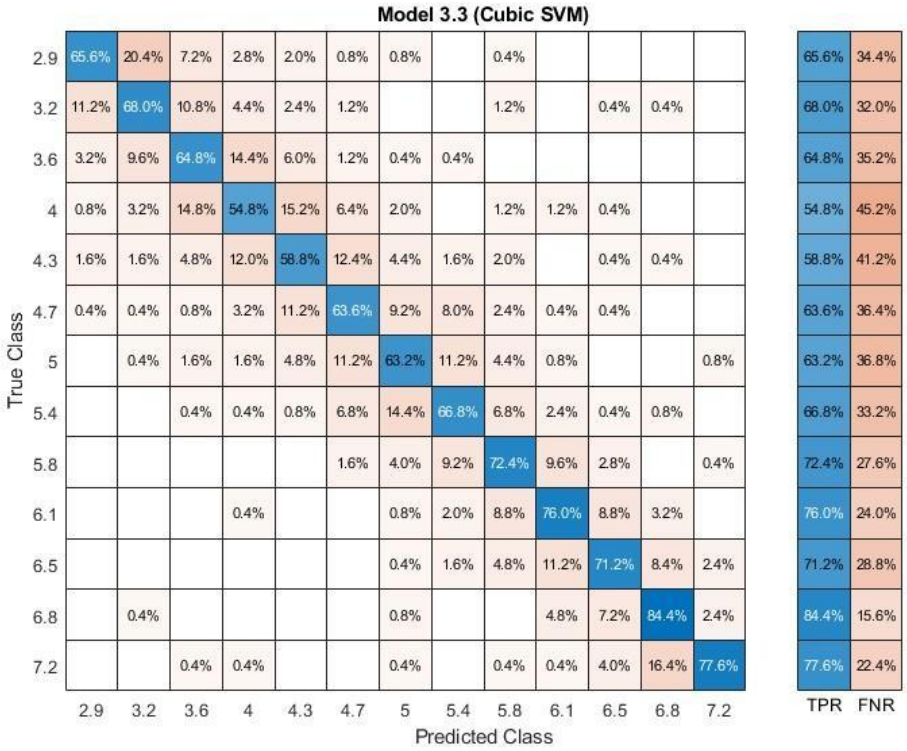

Figure 129. TPR-FNR test confusion matrix of gait segmentation (first phase) for left leg using complete time-series.

For classifying the stance phase of the left leg, the cubic SVM model achieved the highest accuracy, with performance notably increasing from the training to the test set.

Table 20

Models' Accuracy Of Gait Segmentation (First Phase) For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	64.37692	68.24615
4.6	KNN	62.43846	65.50769
4.2	KNN	60.96154	63.81538
4.4	KNN	60.24615	62.8
3.2	SVM	58.90769	61.38462
4.1	KNN	58.88462	60.73846
4.5	KNN	57.40769	58.67692
5.3	Neural Network	56.13077	57.29231
3.5	SVM	55.50769	58.73846
3.4	SVM	48.14615	50.21538

Table 20 (cont.)

5.2	Neural Network	44.35385	45.96923
4.3	KNN	41.9	47.84615
5.5	Neural Network	37.44615	38.52308
5.4	Neural Network	37.18462	38.55385
5.1	Neural Network	35.95385	35.26154
3.1	SVM	35.37692	36.30769
3.6	SVM	29.6	30.55385
2.1	Tree	27.13846	26.30769
2.2	Tree	19.83077	20.03077
2.3	Tree	16.10769	15.6

4.2.1.2.10 Gait segmentation second phase (swing phase). Mirroring the results of the stance phase analysis, the swing phase also exhibited the highest accuracy with the cubic SVM model, underscoring the model's consistent effectiveness across different gait phases.

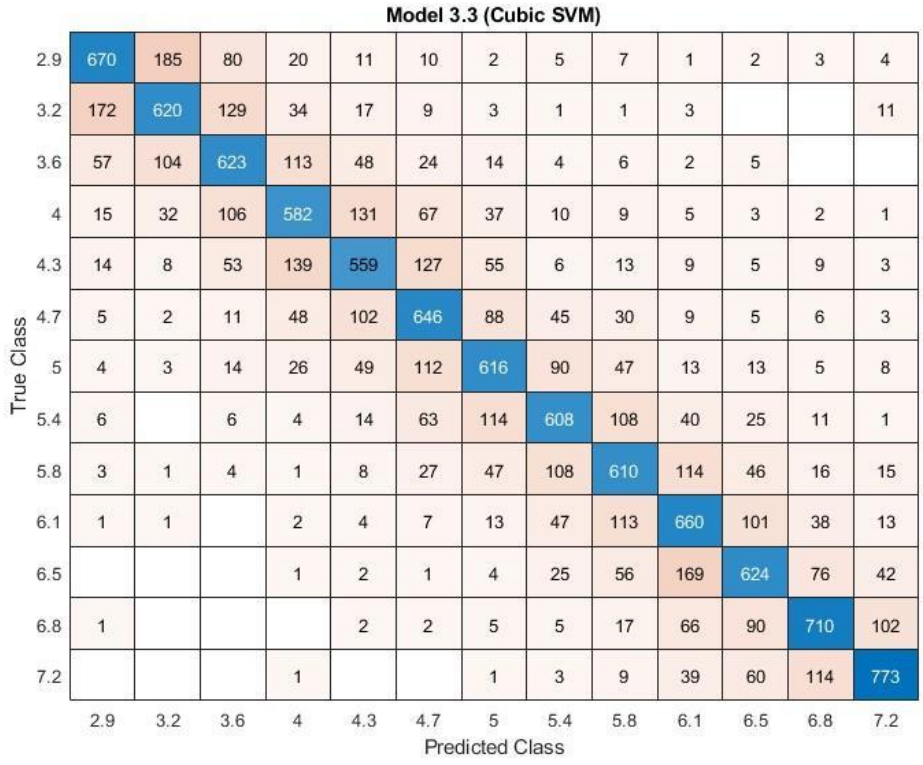

Figure 130. Training confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

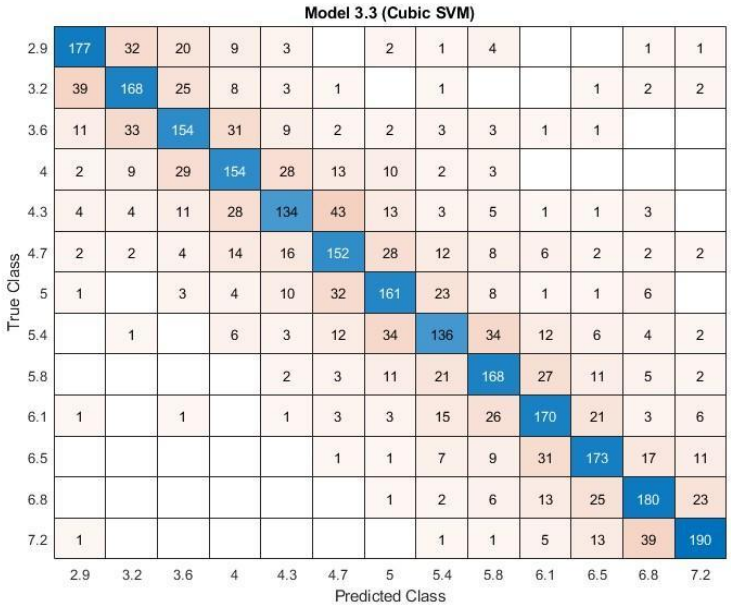

Figure 131. Test confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

Analysis of the confusion matrices reveals high precision in swing phase classification. This means the method accurately distinguishes the swing phase from other gait phases, minimizing errors that could affect gait analysis.

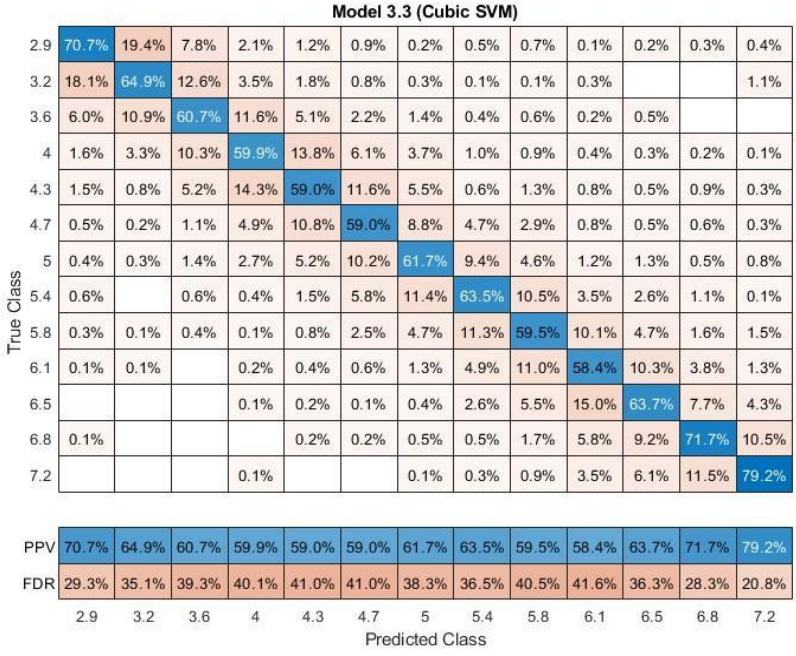

Figure 132. PPV-FDR training confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

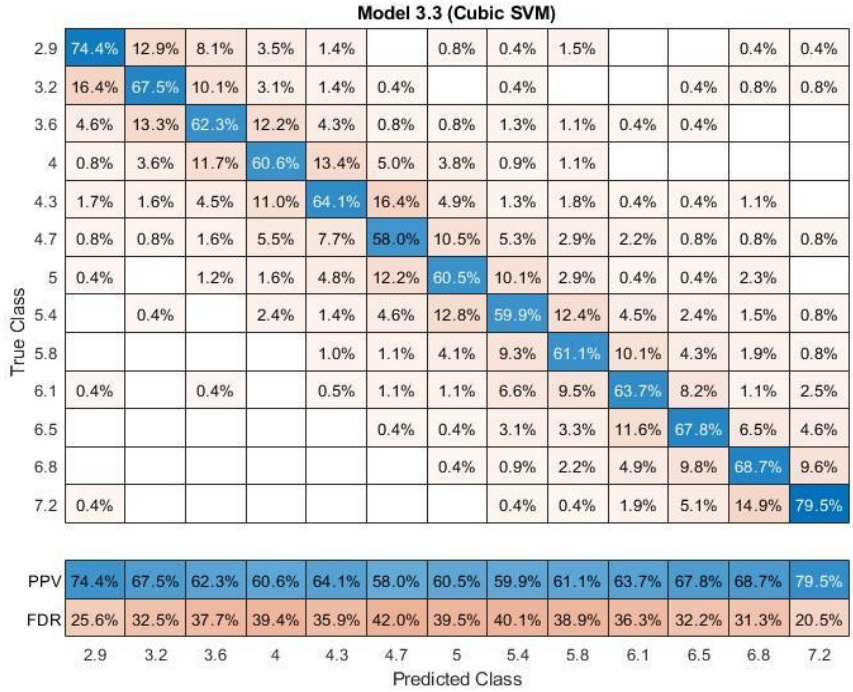

Figure 133. PPV-FDR test confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

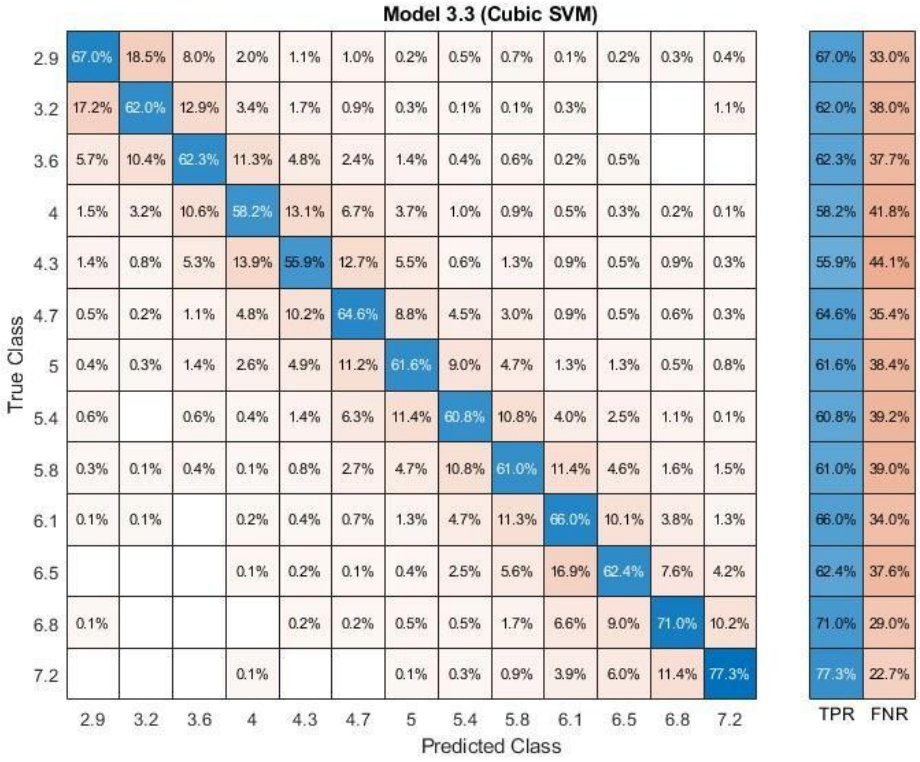

Figure 134. TPR-FNR training confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

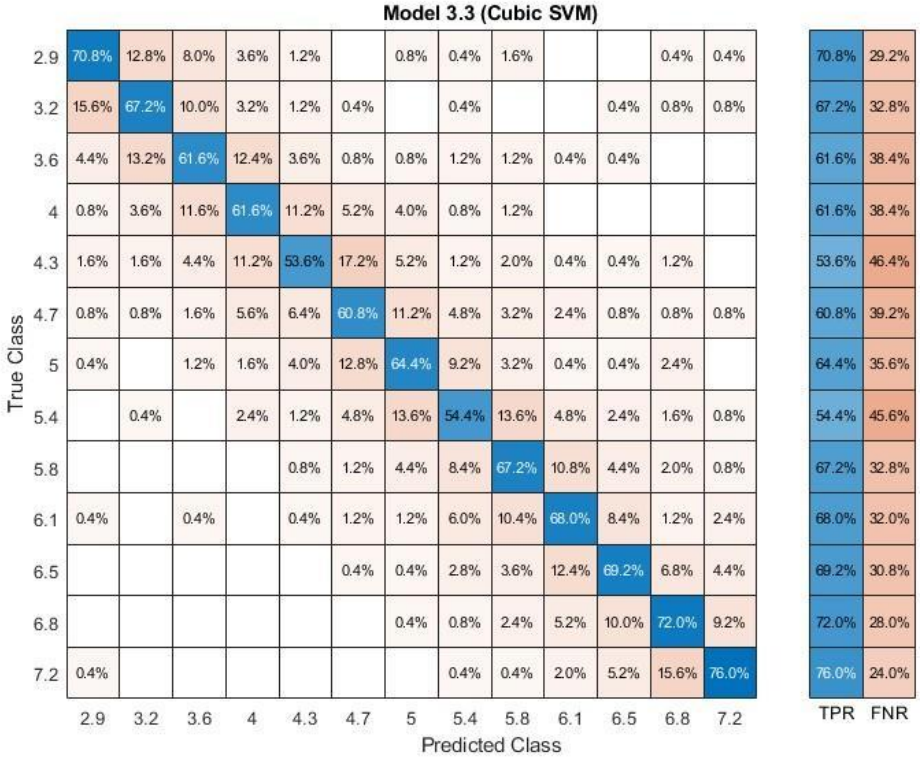

Figure 135. TPR-FNR test confusion matrix of gait segmentation (second phase) for left leg using complete time-series.

The cubic SVM model demonstrated exceptional performance in swing phase classification, not only achieving high accuracy on the training data but also generalizing well to the test set with even higher accuracy. although they have good accuracy, its accuracy doesn't not meet the unaltered time-series.

Table 21

Models' Accuracy Of Gait Segmentation (Second Phase) For Left Leg Using Complete Time-Series

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	63.85385	65.13846
4.6	KNN	63.69231	64.03077
4.2	KNN	61.66154	61.69231
4.1	KNN	60.52308	61.29231
4.4	KNN	60.1	61.10769
4.5	KNN	57.63846	57.56923
3.2	SVM	57.31538	58.33846
5.3	Neural Network	56.66154	57.87692
3.4	SVM	51.83846	53.53846

Table 21 (cont.)

3.5	SVM	51.73077	51.90769
5.2	Neural Network	46.71538	47.50769
4. 3	KNN	44.72308	49.63077
5.5	Neural Network	38.80769	39.26154
5.4	Neural Network	37.81538	37.90769
5.1	Neural Network	36.05385	37.01538
3.1	SVM	34.69231	35.72308
3.6	SVM	30.05385	30.43077
2.1	Tree	28.69231	28.4
2.2	Tree	22.10769	23.26154
2.3	Tree	17.02308	17.13846

4.3 Summary of Accuracy Variations with Incremental Frame Skipping

The plot provides a summary of how accuracy changes with incremental frame skipping in gait analysis, allowing for an assessment of the effects on both right and left leg classification performance.

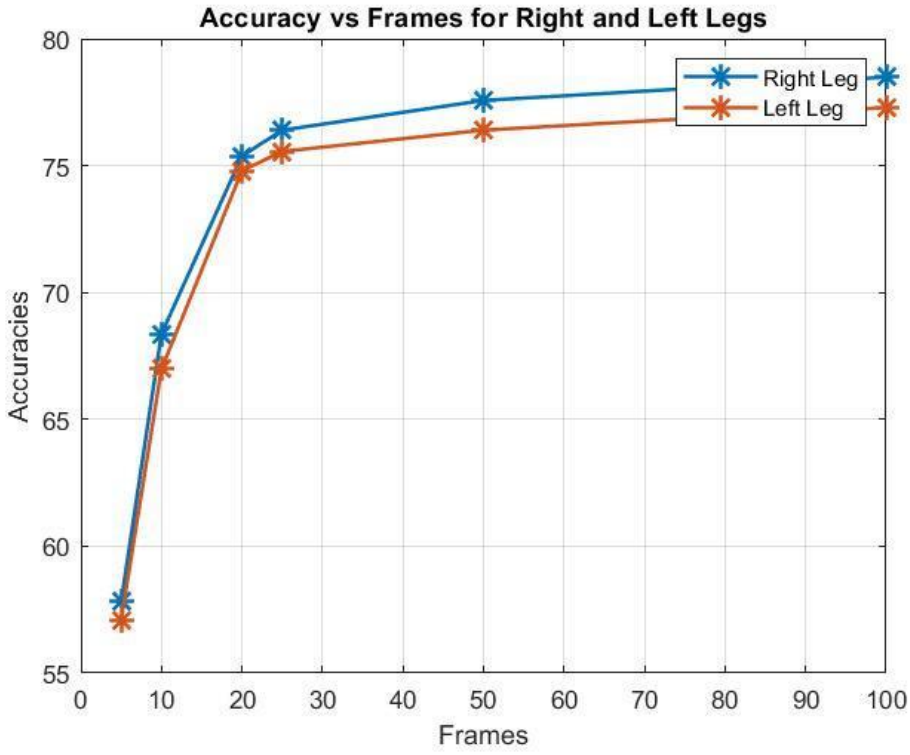

Figure 136. Accuracy vs frames skipping for right and left leg.

4.4 Feature Based Manipulations

4.4.1 Right leg analysis.

4.4.1.1 Complete features data.

These confusion matrices illustrate the classification results achieved using a feature-based approach applied to the complete time series data. Notably, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

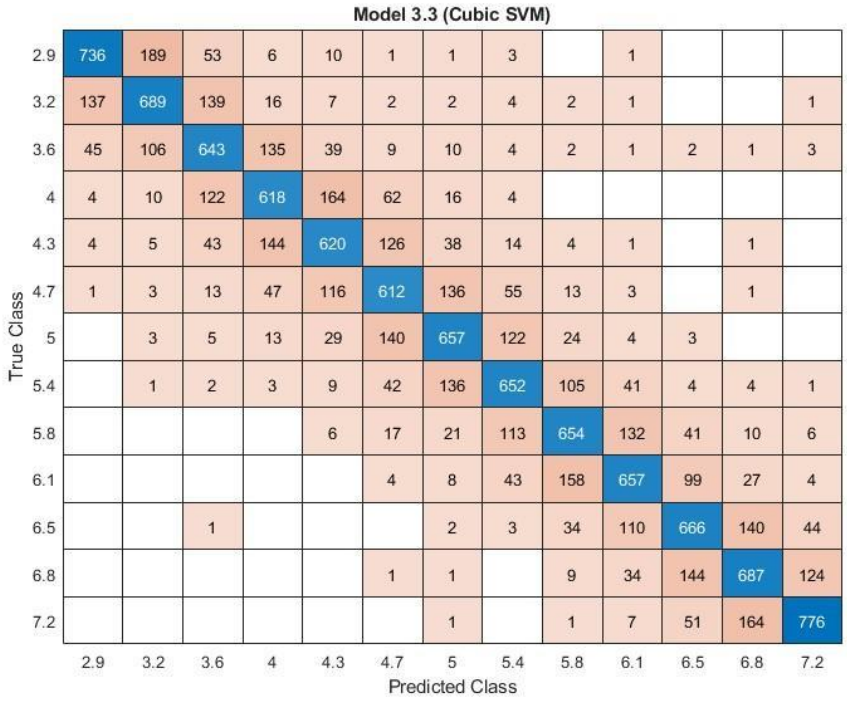

Figure 137. Training confusion matrix for right leg using complete features data.

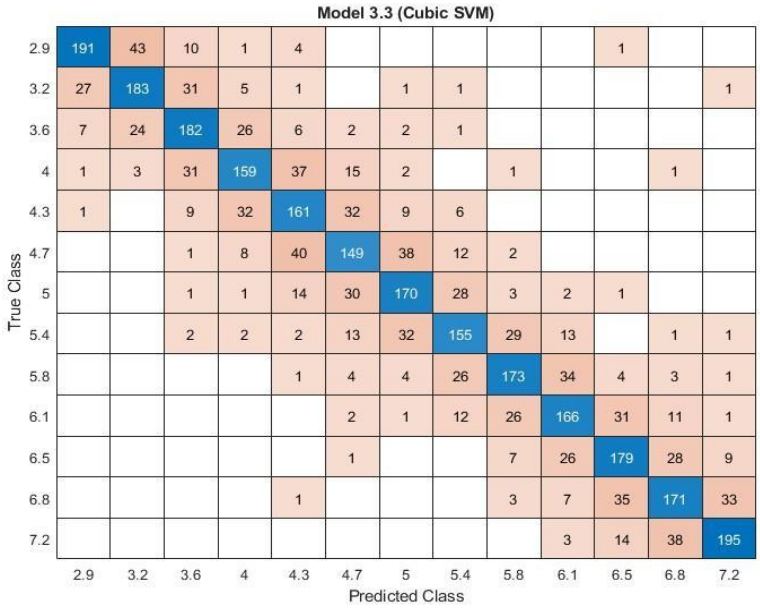

Figure 138. Test confusion matrix for right leg using complete features data.

These confusion matrices demonstrate the effectiveness of the selected features in achieving high accuracy and precision in the classification. The very high precision indicates that the features successfully capture the distinctive characteristics of different classes.

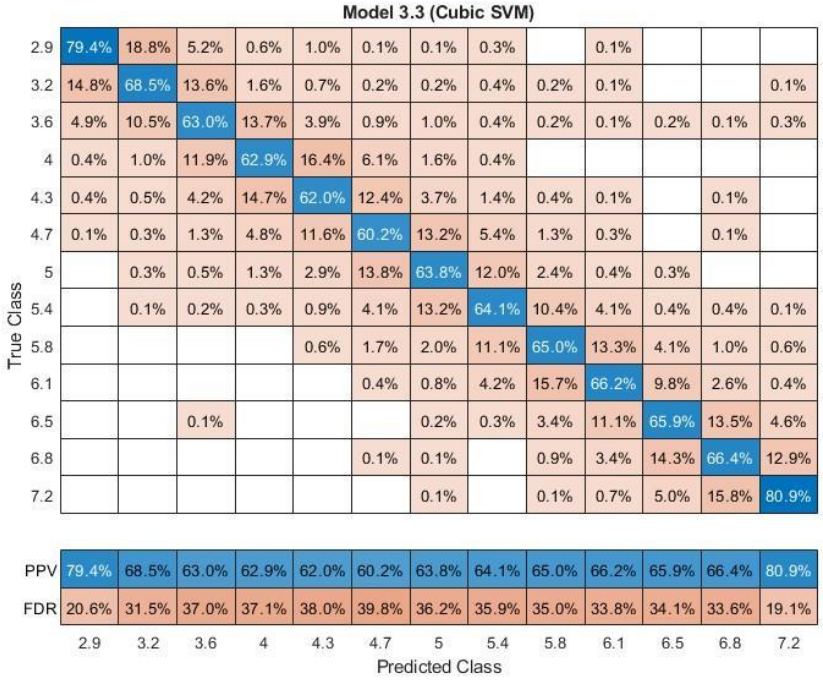

Figure 139. PPV-FDR training confusion matrix for right leg using complete features data.

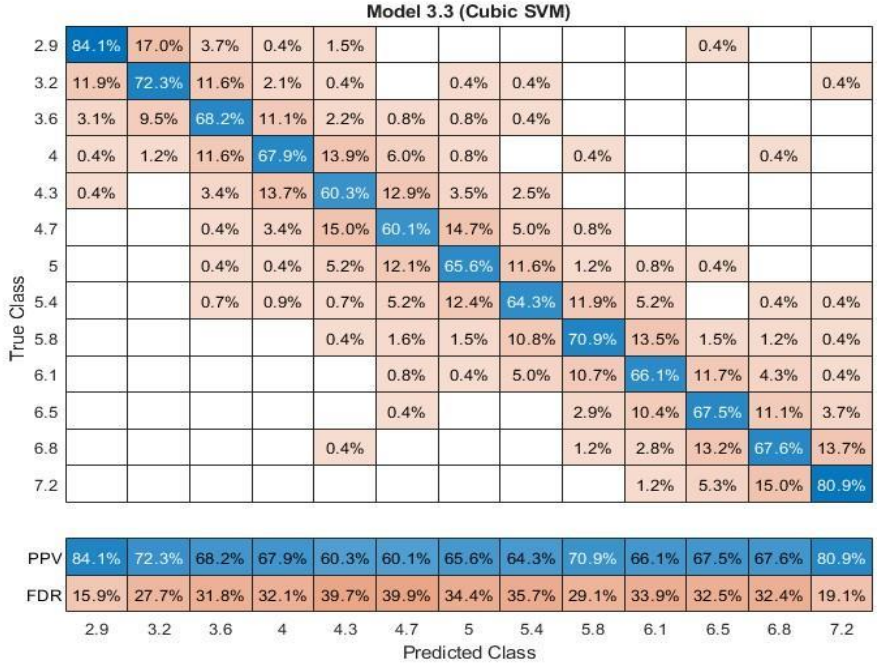

Figure 140. PPV-FDR test confusion matrix for right leg using complete features data.

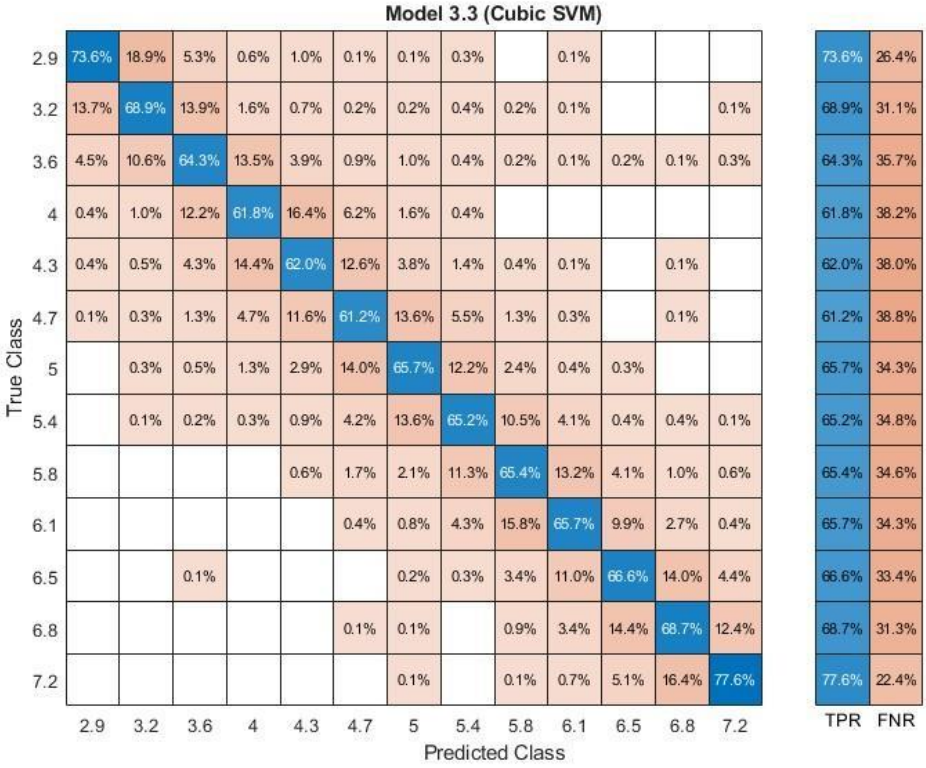

Figure 141. TPR-FNR training confusion matrix for right leg using complete features data.

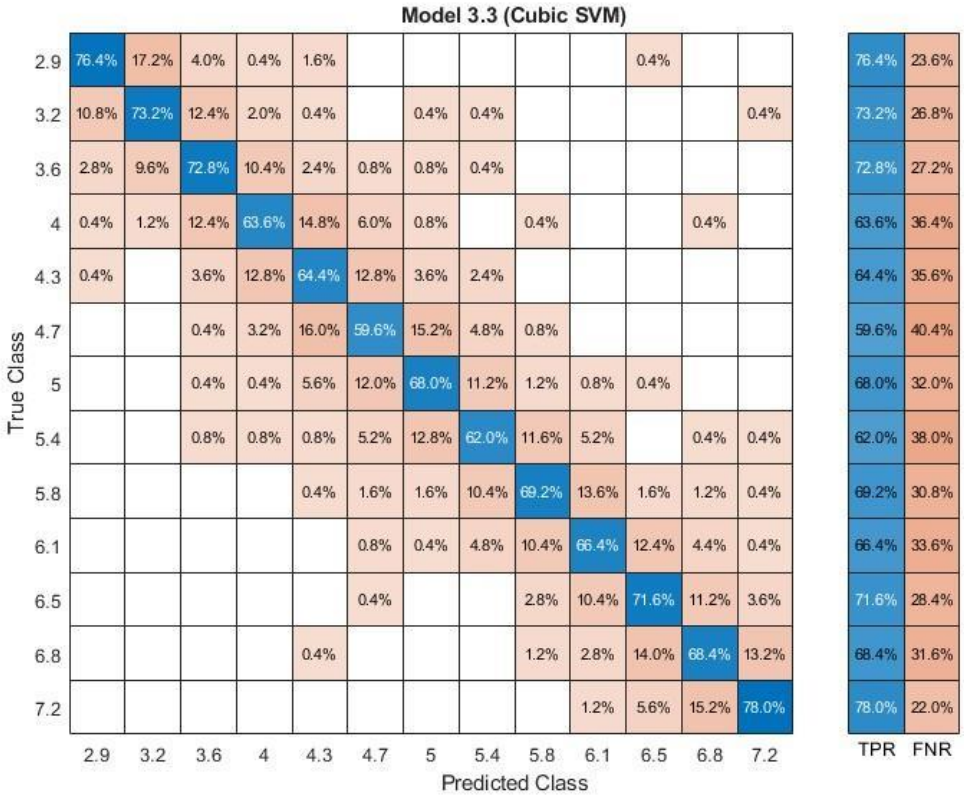

Figure 142. TPR-FNR test confusion matrix for right leg using complete features data.

Using carefully selected features, the cubic SVM model not only demonstrated high accuracy during training but also generalized exceptionally well to the test set, exhibiting even higher accuracy. This highlights the effectiveness of the chosen features in capturing the underlying patterns in the data.

Table 22

Models' Accuracy For Right Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	66.66923	68.73846
3.4	SVM	64.46154	66.64615
5.3	Neural Network	61.96923	64.92308
4.6	KNN	60.96154	61.78462
3.2	SVM	60.88462	61.84615
4.2	KNN	59.61538	61.6
4.4	KNN	58.38462	60.18462
4.5	KNN	56.76154	58.21538
4.1	KNN	56.11538	56.06154

Table 22 (cont.)

5.2	Neural Network	53.93846	54.30769
3.5	SVM	52.73846	54.76923
5.5	Neural Network	45.1	45.63077
5.4	Neural Network	44.57692	44.64615
4.3	KNN	43.98462	50.61538
5.1	Neural Network	42.84615	43.66154
3.1	SVM	31.46923	32.33846
2.1	Tree	28.70769	29.2
3.6	SVM	24.16154	23.90769
2.2	Tree	20.61538	21.69231
2.3	Tree	16.3	15.78462

4.4.1.2 Skipped speeds odd pattern (subset 1). The figures below demonstrate that the skipped speeds method, when applied to the extracted features, achieves a higher degree of speed-specific classification compared to using percentile data.

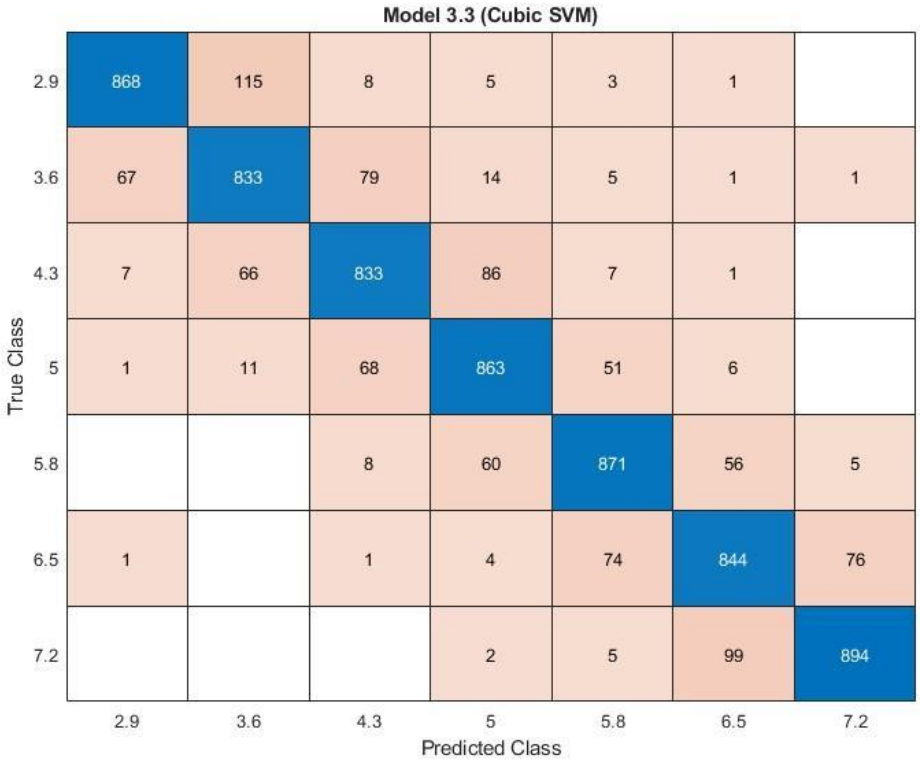

Figure 143. Training confusion matrix of subset 1 for right leg using complete features data.

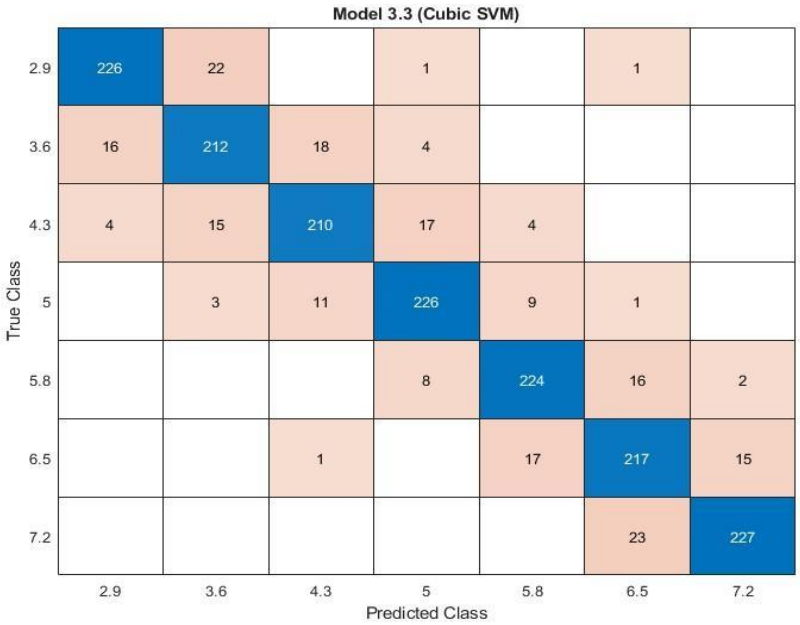

Figure 144. Test confusion matrix of subset 1 for right leg using complete features data.

The high accuracy and precision evident in these confusion matrices underscore the importance of careful feature selection for effective speed classification.

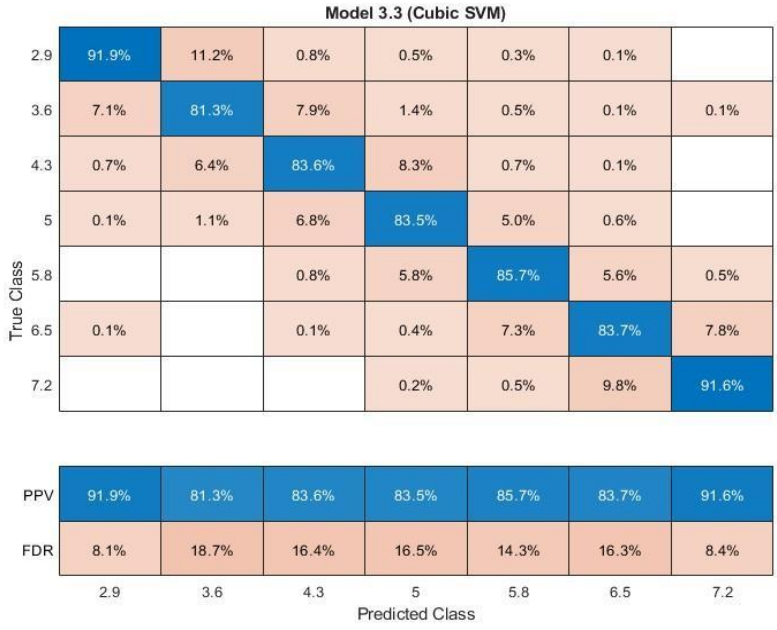

Figure 145. PPV-FDR training confusion matrix of subset 1 for right leg using complete features data.

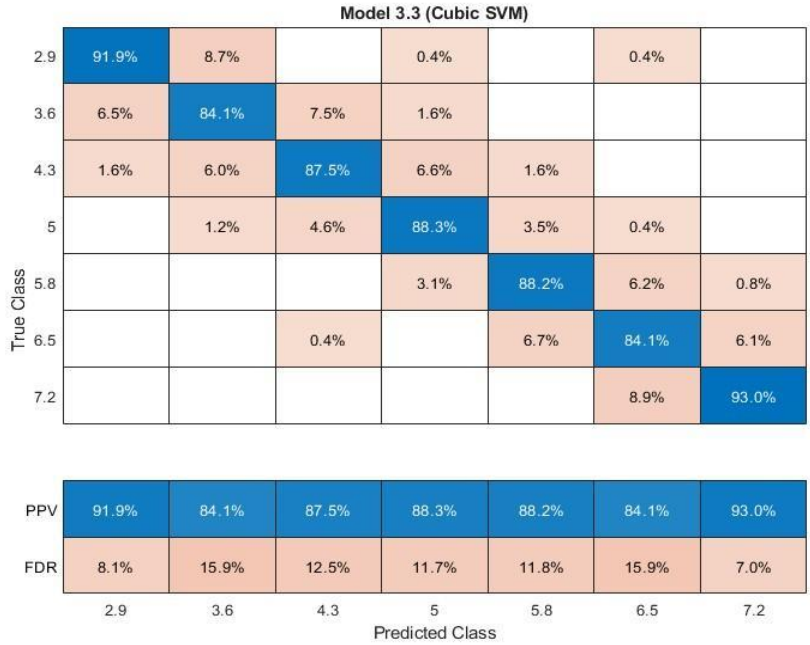

Figure 146. Test confusion matrix of subset 1 for right leg using complete features data.

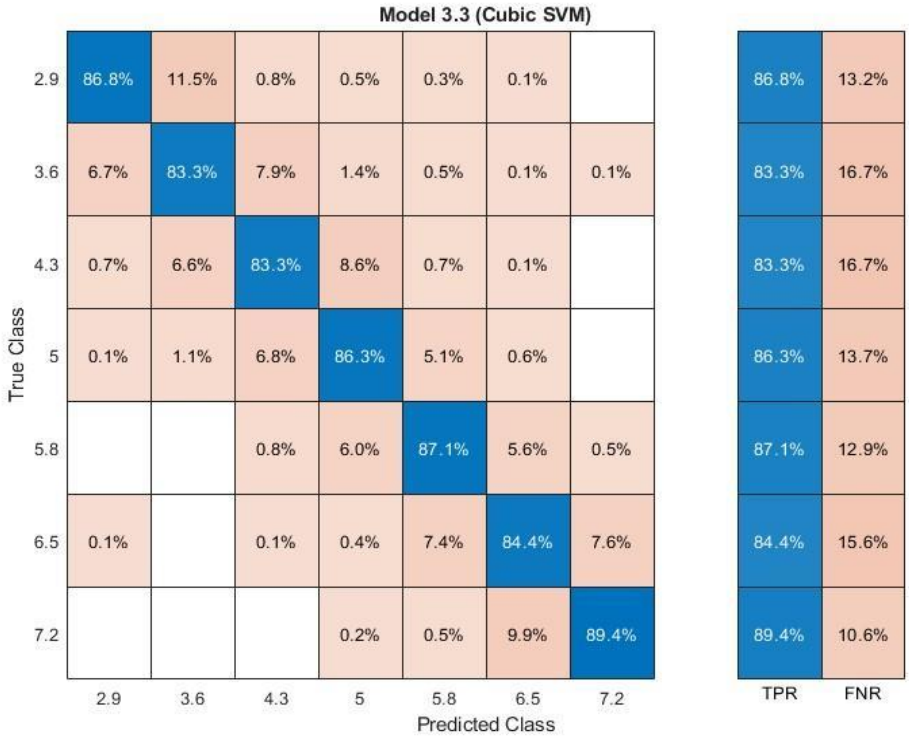

Figure 147. TPR-FNR training confusion matrix of subset 1 for right leg using complete features data.

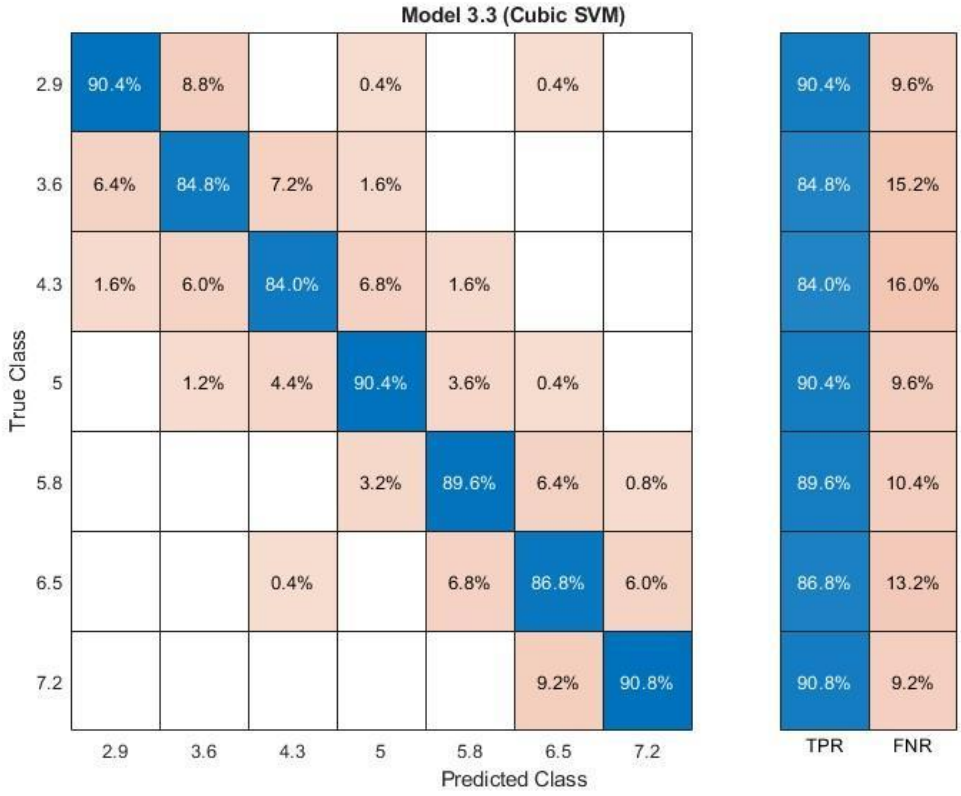

Figure 148. TPR-FNR test confusion matrix of subset 1 for right leg using complete features data.

Using carefully selected features, the cubic SVM model not only demonstrated high accuracy during training but also generalized exceptionally well to the test set, exhibiting even higher accuracy in speed classification. This highlights the effectiveness of the chosen features in capturing the underlying patterns in the data.

Table 23

Models' Accuracy Of Subset 1 For Right Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	85.8	88.11429
4.6	KNN	84.77143	86.28571
3.4	SVM	84.37143	88.28571
4.2	KNN	83.97143	85.54286
5.3	Neural Network	82.57143	85.08571
4.4	KNN	81.95714	83.82857
3.2	SVM	81.25714	82.17143
4.5	KNN	81.14286	82.74286

Table 23 (cont.)

4.1	KNN	80.15714	81.48571
5.2	Neural Network	78.9	79.71429
3.5	SVM	74.9	76.51429
5.4	Neural Network	72.68571	73.48571
5.5	Neural Network	72.12857	73.94286
5.1	Neural Network	71.38571	70.51429
3.1	SVM	54.51429	53.77143
4.3	KNN	54.34286	64.57143
2.1	Tree	52.27143	51.02857
3.6	SVM	44.51429	43.6
2.2	Tree	38.15714	36.85714
2.3	Tree	30.3	29.82857

4.4.1.3 Skipped speeds even pattern. By leveraging the informative features extracted from the data, the figures below show that the skipped speeds method provides more accurate speed-specific classification than using percentile data.

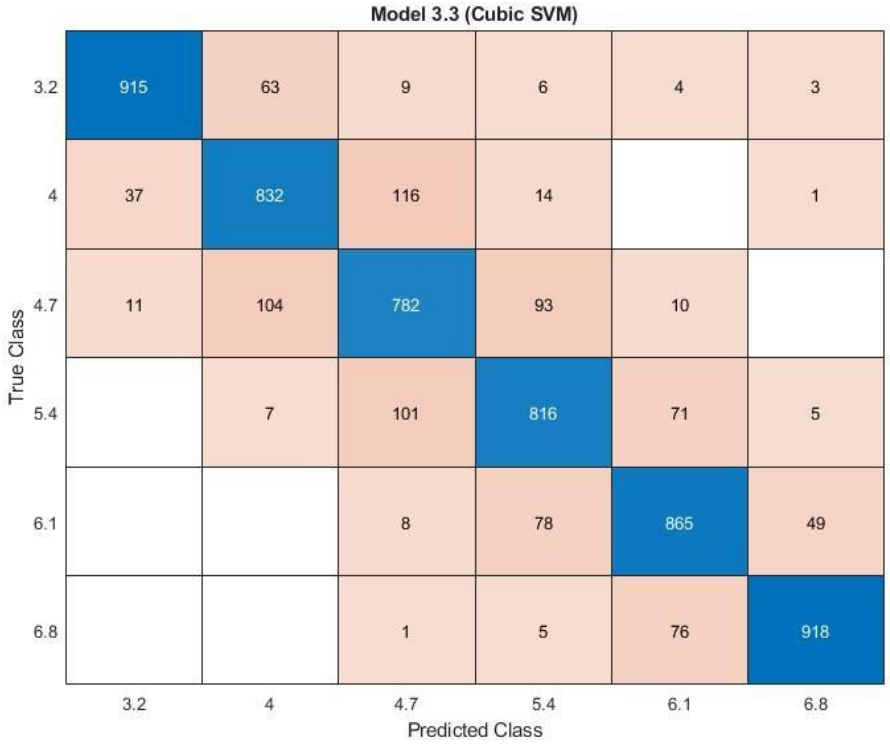

Figure 149. Training confusion matrix of subset 1 for right leg using complete features data.

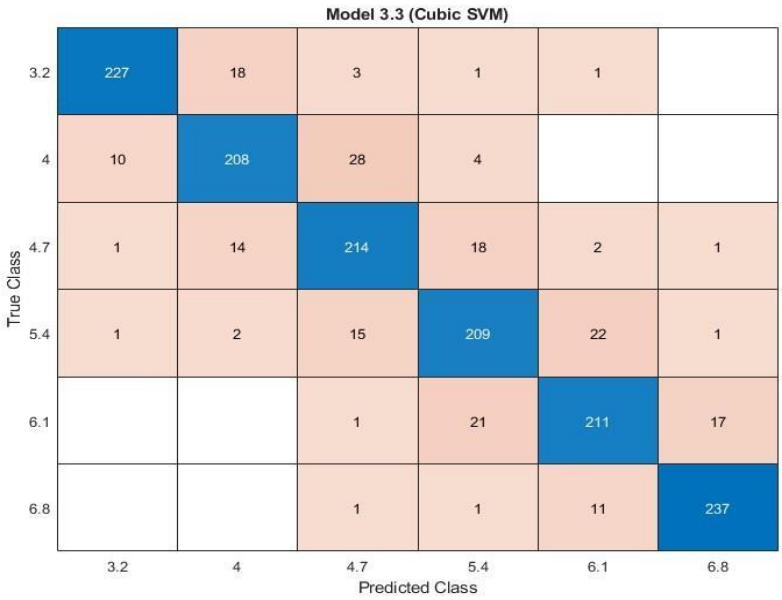

Figure 150. Test confusion matrix of subset 1 for right leg using complete features data.

The high accuracy and precision evident in these confusion matrices underscore the importance of careful feature selection for effective speed classification.

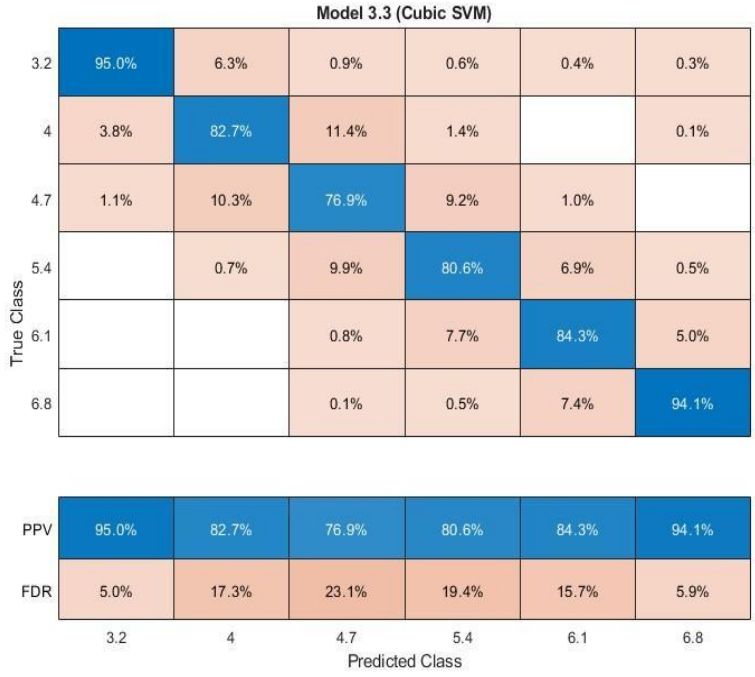

Figure 151. PPV-FDR training confusion matrix of subset 1 for right leg using complete features data.

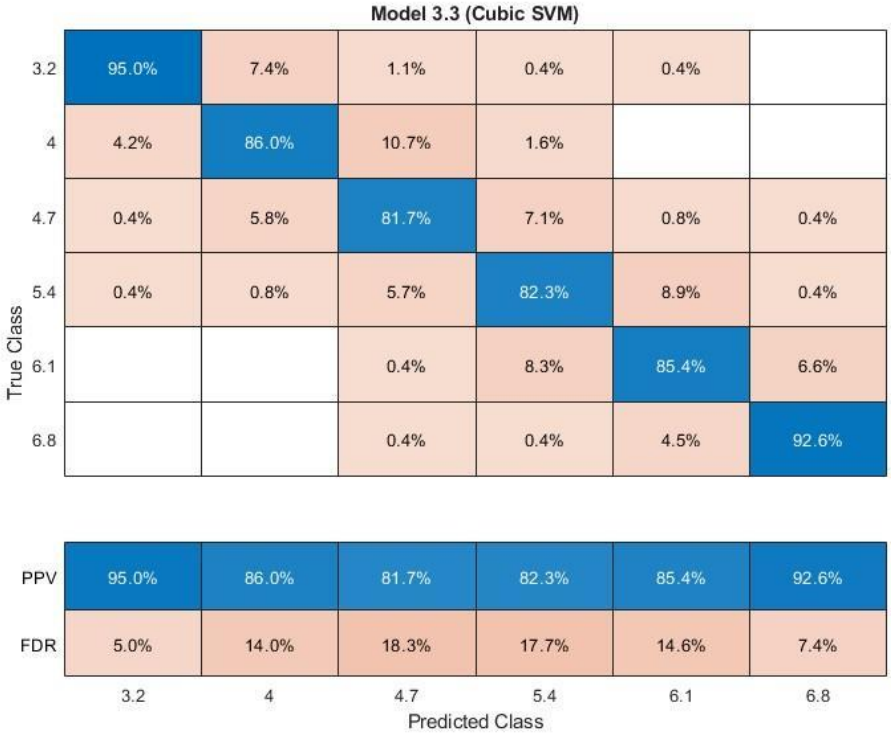

Figure 152. PPV-FDR test confusion matrix of subset 1 for right leg using complete features data.

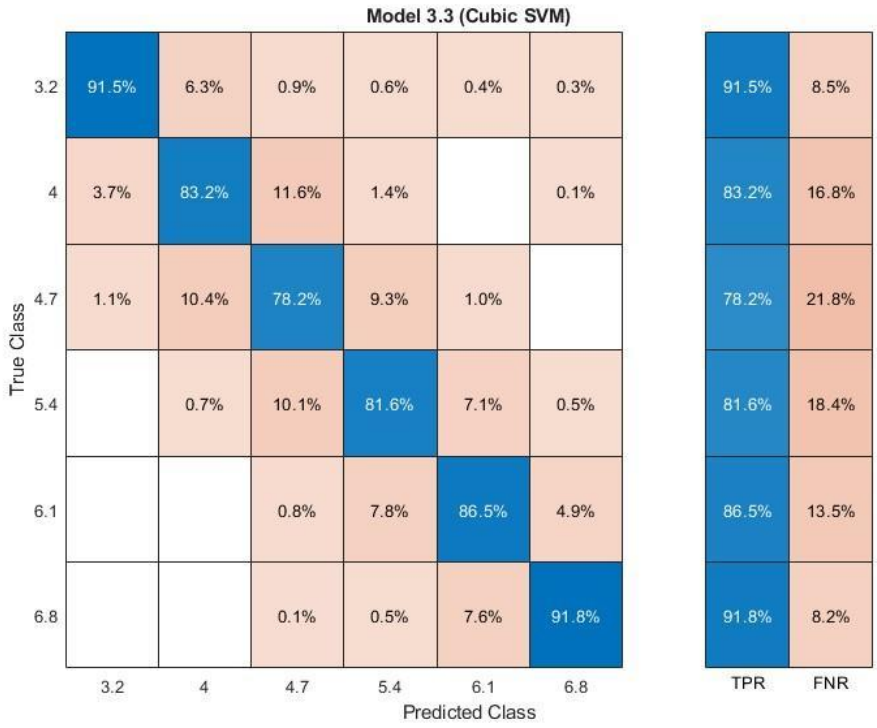

*Figure 153.*TPR-FNR training confusion matrix of subset 1 for right leg using complete features data.

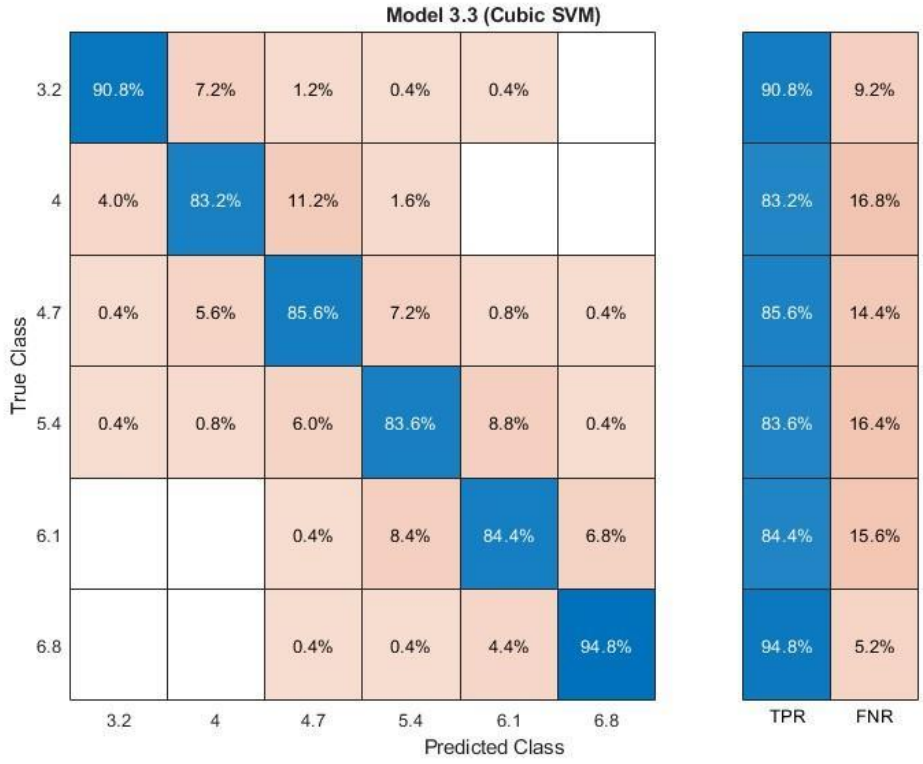

Figure 154. Test confusion matrix of subset 1 for right leg using complete features data.

Using carefully selected features, the cubic SVM model not only demonstrated high accuracy during training but also generalized exceptionally well to the test set, exhibiting even higher accuracy in speed classification. This highlights the effectiveness of the chosen features in capturing the underlying patterns in the data.

Table 24

Models' Accuracy Of Subset 1 For Right Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	85.46667	87.06667
3.4	SVM	84.96667	87.46667
4.6	KNN	83.93333	84.8
5.3	Neural Network	83.55	83.93333
4.2	KNN	82.93333	84.73333
3.2	SVM	82.6	82.33333
4.1	KNN	81.33333	80.46667
4.4	KNN	81.3	82.46667
4.5	KNN	79.55	81.2

Table 24 (cont.)

5.2	Neural Network	79.53333	81.86667
3.5	SVM	76.65	79.93333
5.5	Neural Network	74.96667	76.13333
5.4	Neural Network	74.8	76.73333
5.1	Neural Network	73.46667	75.06667
2.1	Tree	56.4	57.33333
4.3	KNN	56.23333	64.46667
3.1	SVM	52.73333	56.06667
3.6	SVM	44.26667	46.4
2.2	Tree	42.96667	40.86667
2.3	Tree	32.73333	30.73333

4.4.2 Left leg analysis

4.4.2.1 Complete features data. These confusion matrices illustrate the classification results achieved using the complete time series data of the left leg and the extracted features. Notably, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

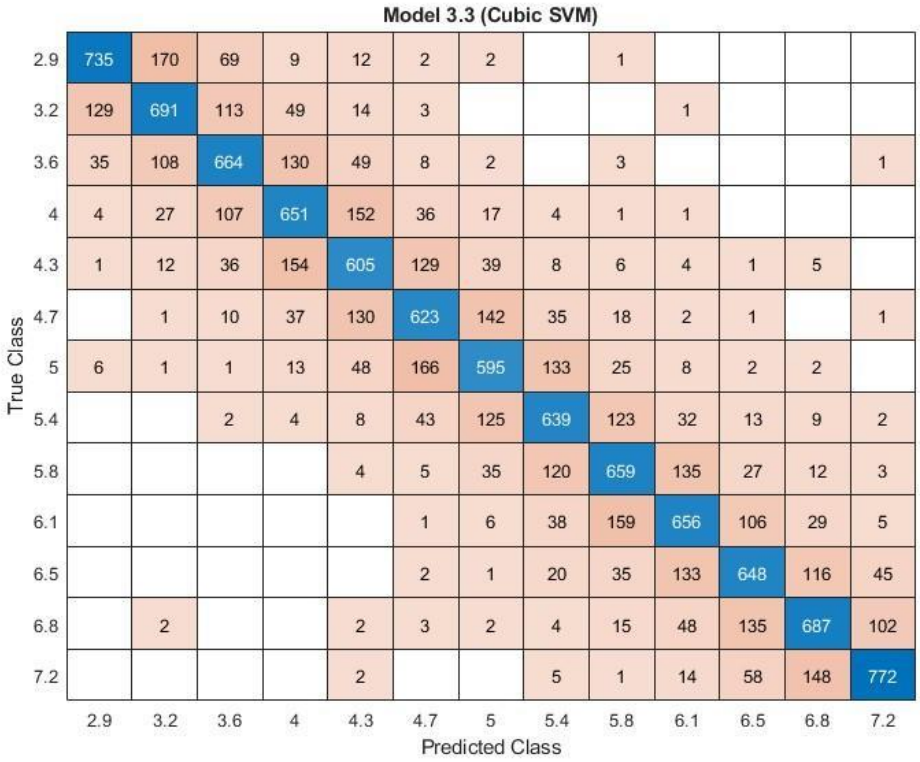

Figure 155. Training confusion matrix for left leg using complete features data.

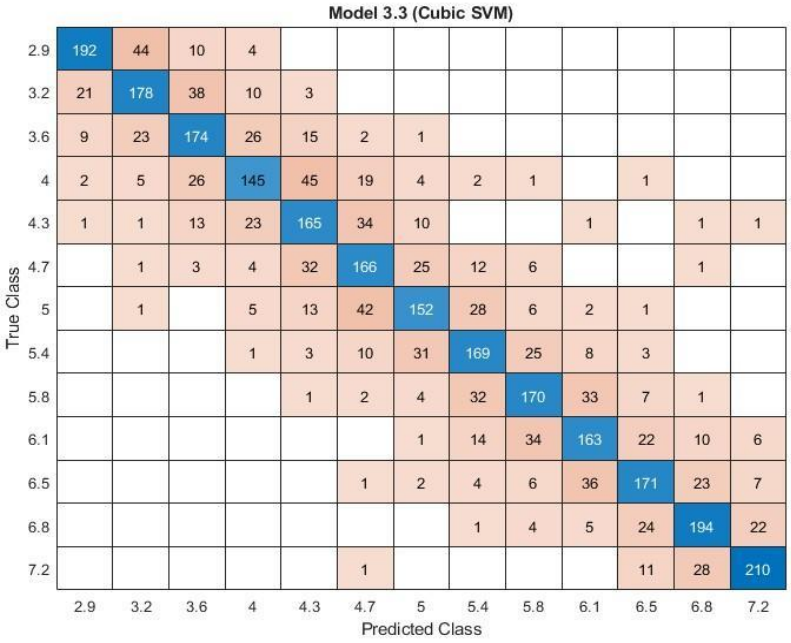

Figure 156. Test confusion matrix for left leg using complete features data.

These confusion matrices demonstrate the effectiveness of the selected features in achieving high accuracy and precision in speed classification. A high precision indicates successful feature capturing of distinctive characteristics.

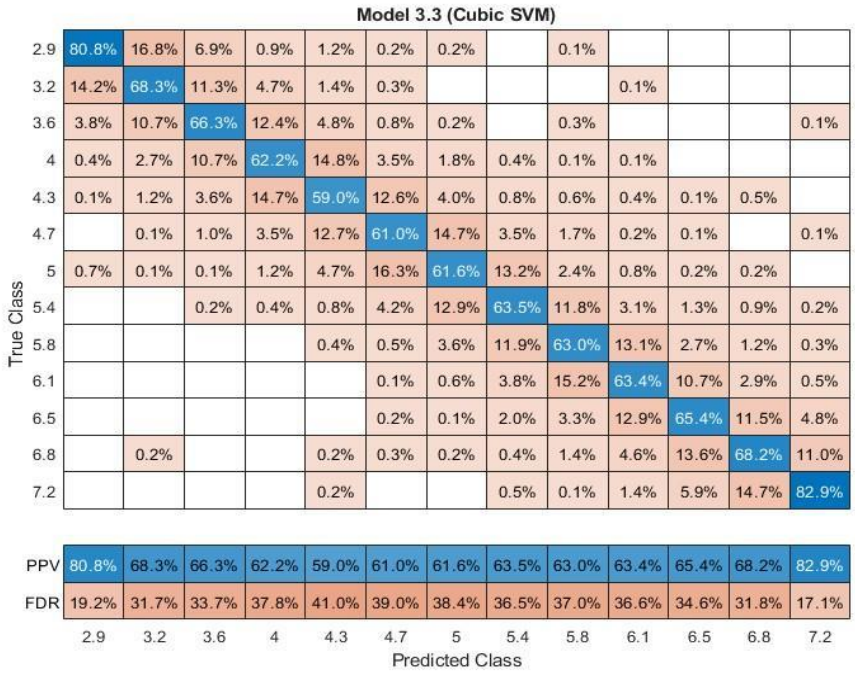

Figure 157. PPV-FDR training confusion matrix for left leg using complete features data.

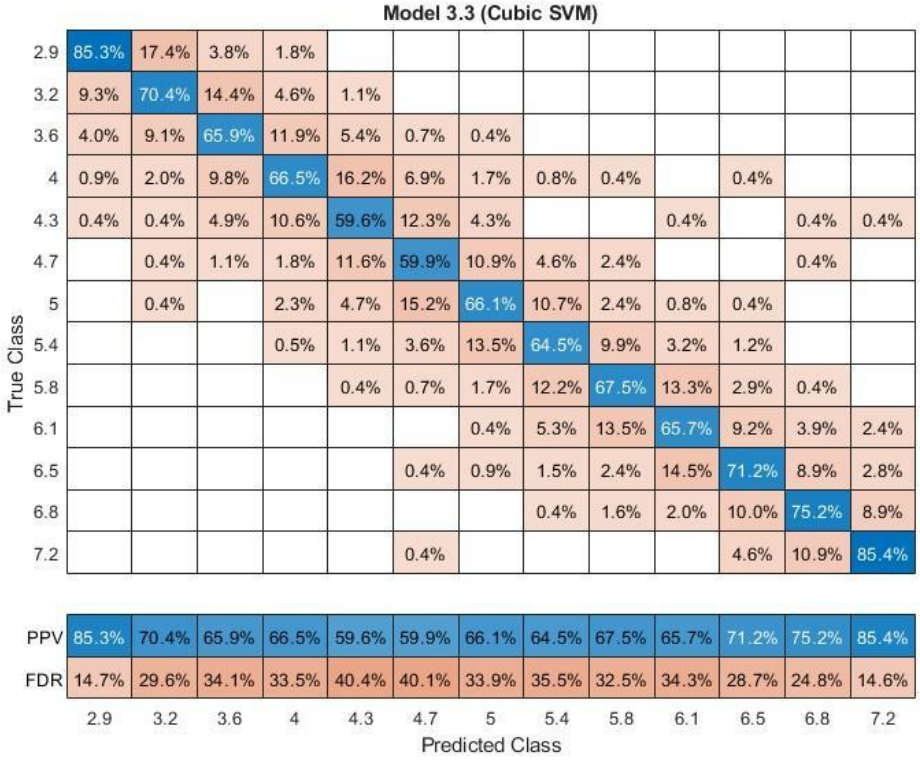

Figure 158. PPV-FDR test confusion matrix for left leg using complete features data.

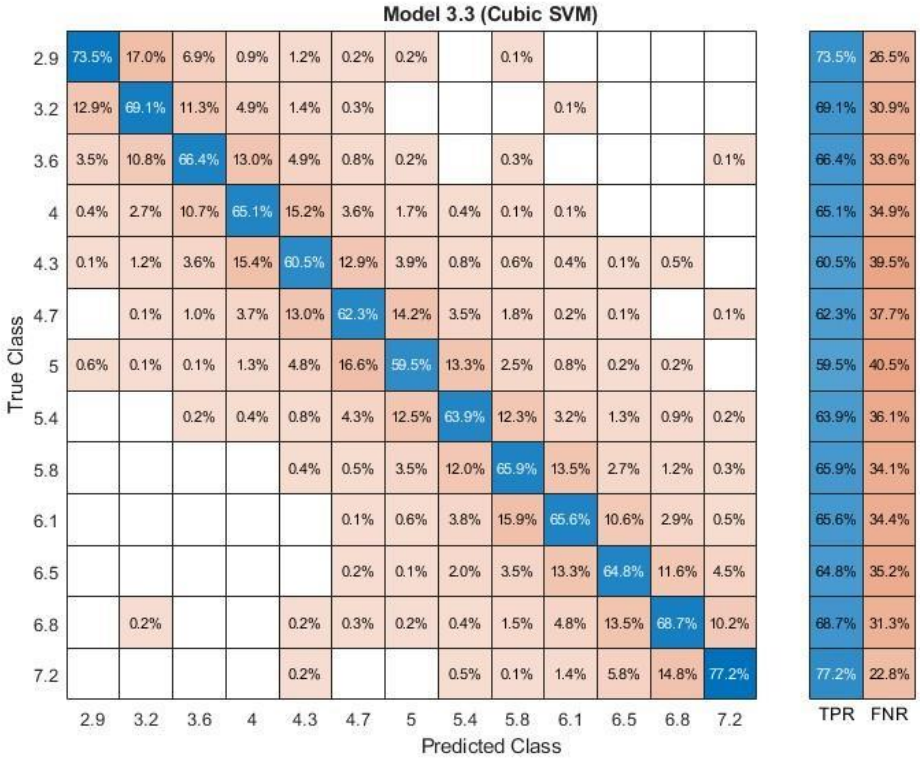

Figure 159. TPR-FNR training confusion matrix for left leg using complete features data.

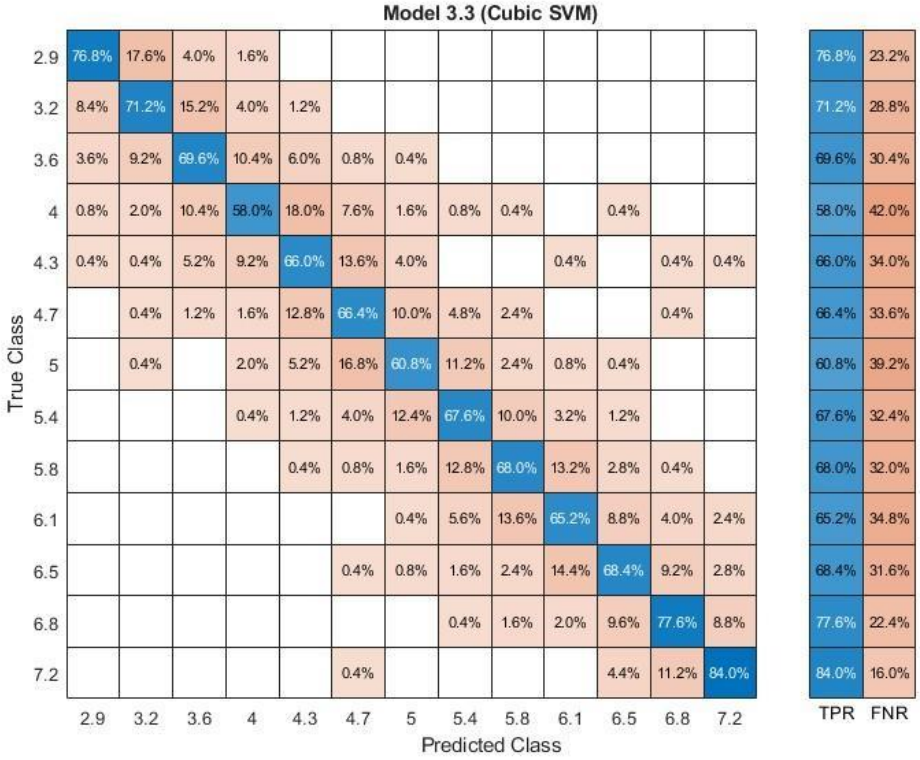

Figure 160. TPR-FNR test confusion matrix for left leg using complete features data.

Using carefully selected features, the cubic SVM model not only demonstrated high accuracy during training but also generalized exceptionally well to the test set, exhibiting even higher accuracy in speed classification. This highlights the effectiveness of the chosen features in capturing the underlying patterns in the data.

Table 25

Models' Accuracy For Left Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	66.34615	69.2
5.3	Neural Network	63.32308	66.15385
3.4	SVM	63.22308	66.12308
3.2	SVM	60.63077	62.70769
4.6	KNN	60.60769	63.10769
4.2	KNN	58.7	61.53846
4.4	KNN	57.94615	61.44615
5.2	Neural Network	56.38462	58.46154
4.1	KNN	56.11538	56.67692

Table 25 (cont.)

4.5	KNN	55.65385	58.24615
3.5	SVM	50.26923	53.04615
5.5	Neural Network	46.26923	44.76923
5.4	Neural Network	45.56923	48.06154
5.1	Neural Network	45.3	45.38462
4.3	KNN	43.72308	50.55385
3.1	SVM	36.59231	36.46154
3.6	SVM	29.06923	30.55385
2.1	Tree	28.12308	29.29231
2.2	Tree	20.77692	22.12308
2.3	Tree	16.30769	16.89231

4.4.2.2 Skipped speeds odd pattern (subset 1). These confusion matrices illustrate the classification results achieved using a feature-based approach applied to the skipped speeds in odd pattern. Notably, the test confusion matrix demonstrated a higher percentage of correct classifications than the training confusion matrix.

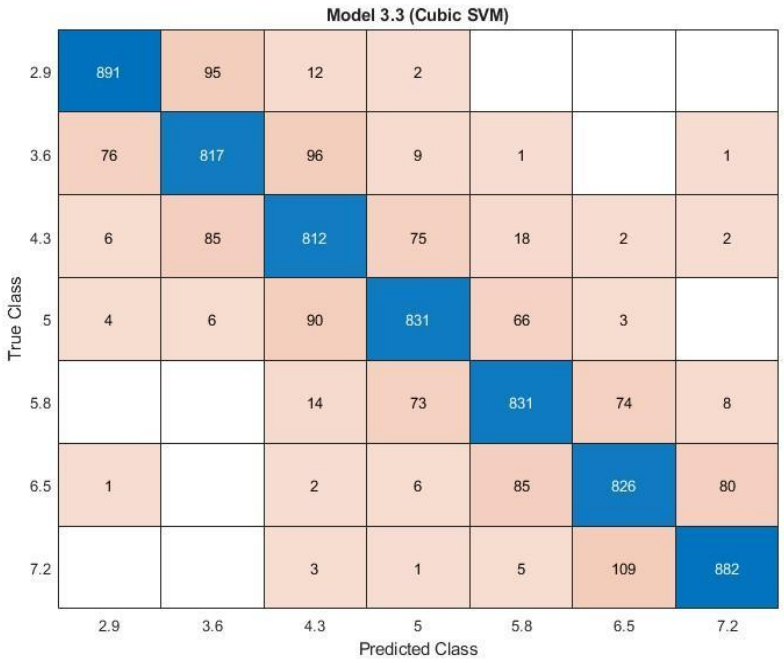

Figure 161. Training confusion matrix of subset 1 for left leg using complete features data.

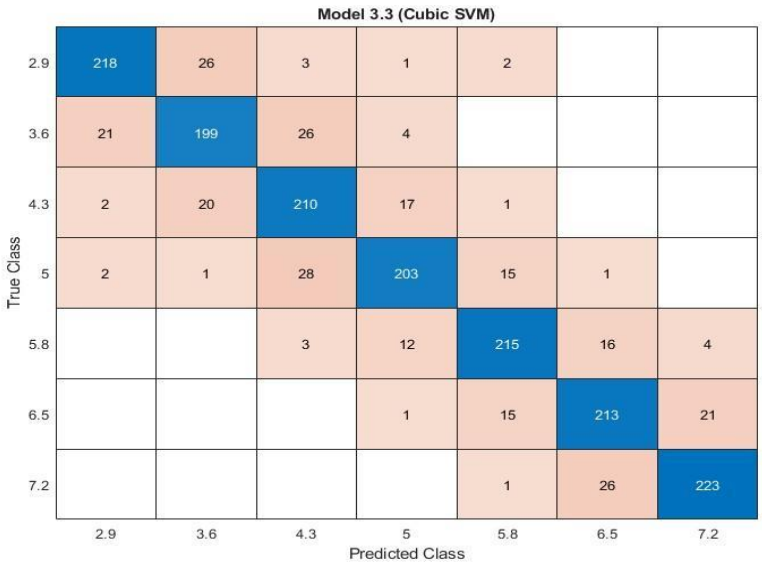

Figure 162. Test confusion matrix of subset 1 for left leg using complete features data.

These results highlight the power of well-chosen features in achieving accurate and precise speed classification.

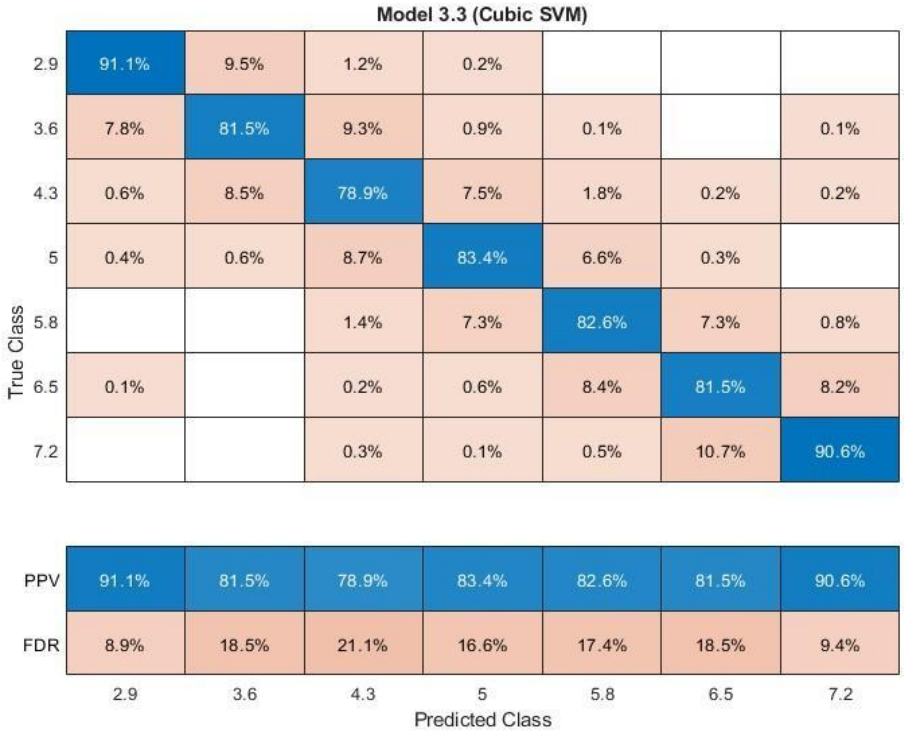

Figure 163. PPV-FDR training confusion matrix of subset 1 for left leg using complete features data.

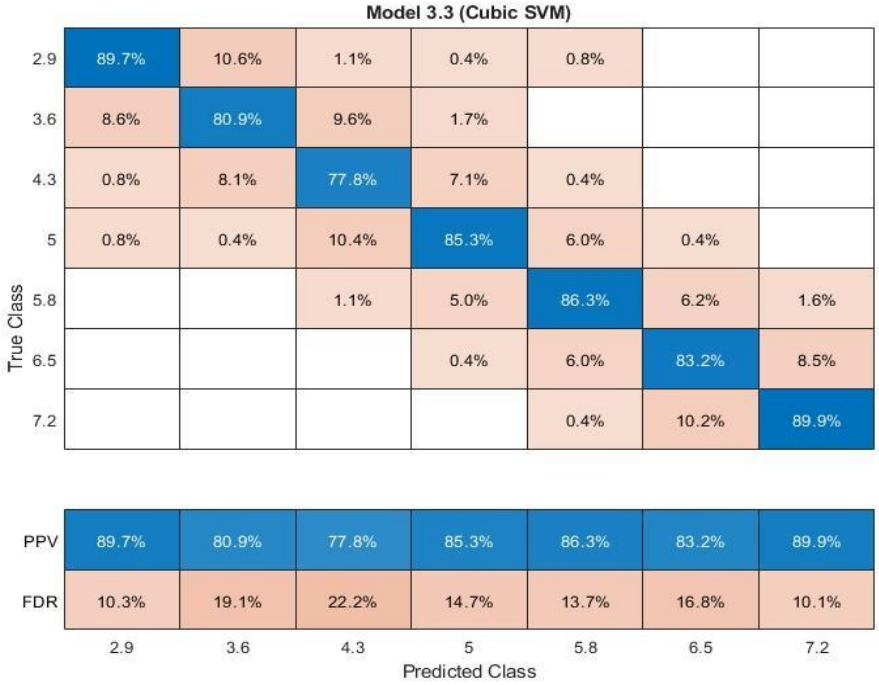

Figure 164. PPV-FDR test confusion matrix of subset 1 for left leg using complete features data.

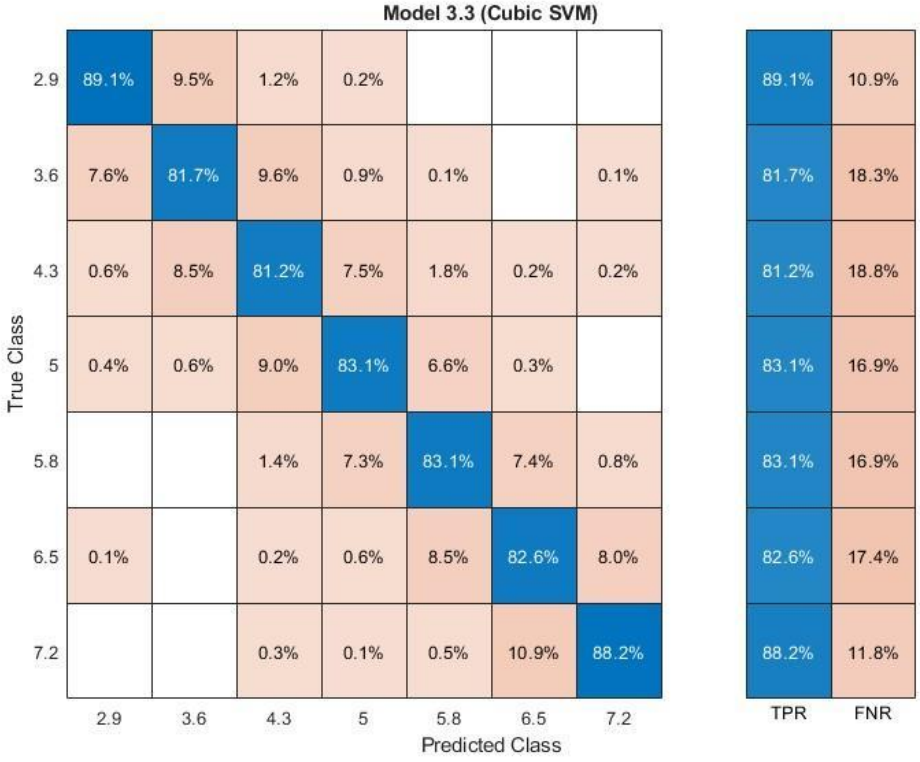

Figure 165. TPR-FNR training confusion matrix of subset 1 for left leg using complete features data.

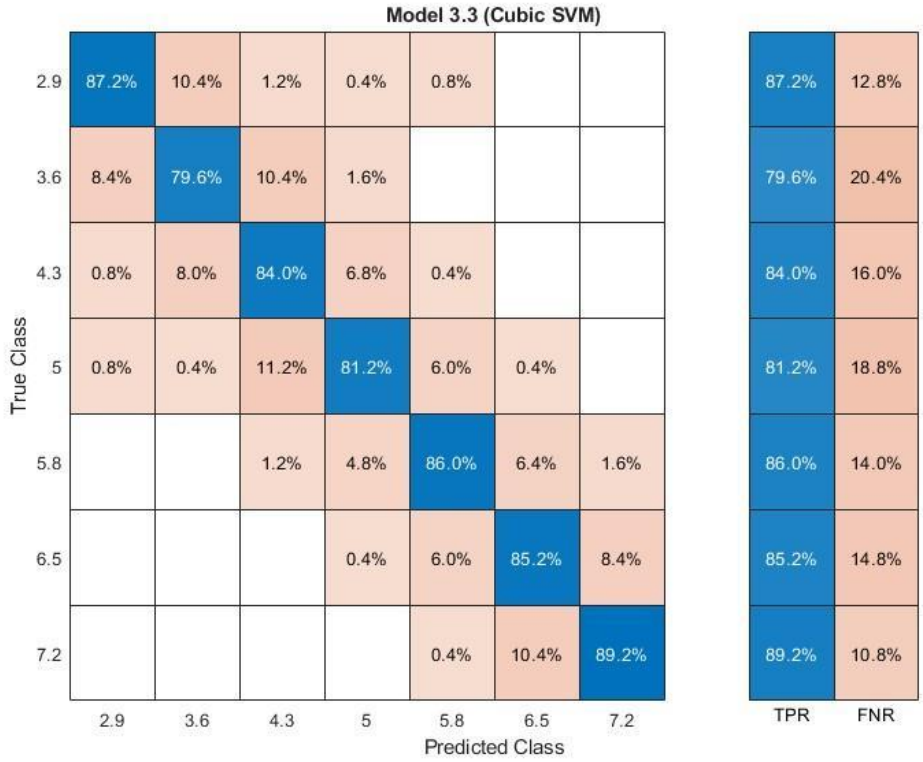

Figure 166. TPR-FNR test confusion matrix of subset 1 for left leg using complete features data.

The cubic SVM model, trained on the extracted features, achieved the highest accuracy in speed classification while skipping speeds. Its performance was remarkable, with very high accuracy on the training set and further improvement on the test set, demonstrating the effectiveness of the chosen features.

Table 26

Models' Accuracy Of Subset 1 For Left Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	84.14286	84.62857
3.4	SVM	83.85714	84.62857
5.3	Neural Network	82.81429	84.05714
4.6	KNN	81.95714	83.65714
3.2	SVM	80.7	82.57143
4.2	KNN	80.67143	82.91429
4.4	KNN	79.4	81.25714
5.2	Neural Network	79.37143	81.48571
4.1	KNN	78.87143	79.6

Table 26 (cont.)

4.5	KNN	77.24286	78.91429
5.4	Neural Network	74.31429	75.08571
5.5	Neural Network	73.95714	76.51429
5.1	Neural Network	72.87143	74.45714
3.5	SVM	72.02857	73.48571
3.1	SVM	58.72857	57.65714
4.3	KNN	57.38571	63.02857
2.1	Tree	50.48571	50.57143
3.6	SVM	49.88571	49.31429
2.2	Tree	39.55714	38.34286
2.3	Tree	31.12857	30.74286

4.4.2.3 Skipped even pattern (subset 2). For speed-specific classification based on the extracted features, the figures below demonstrate the superior performance of the skipped speeds method compared to using percentile data.

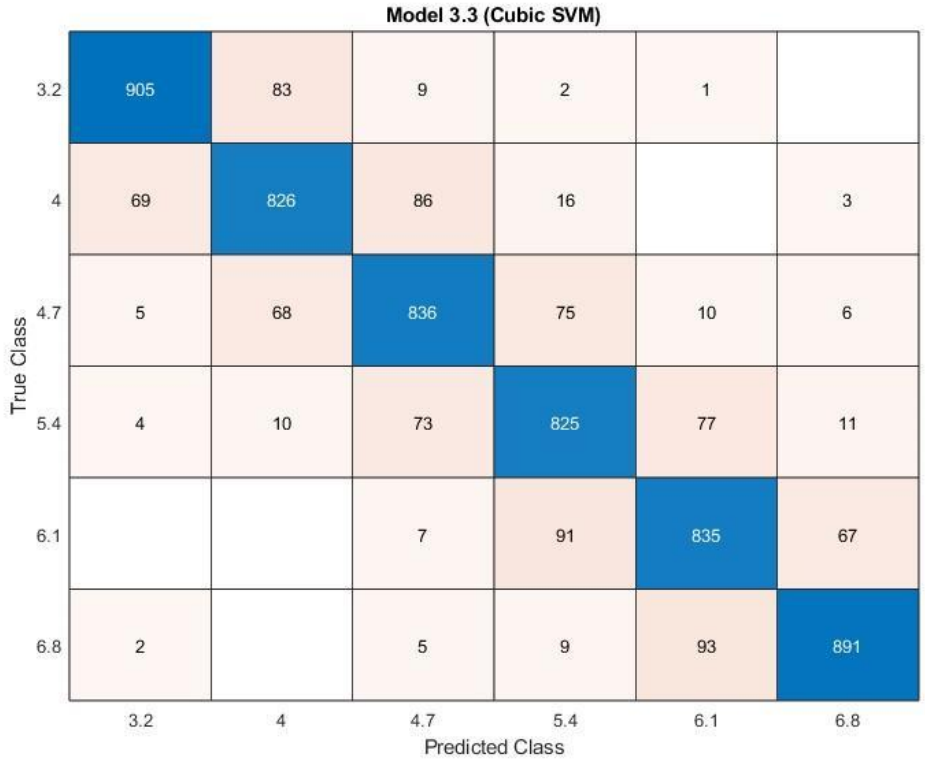

Figure 167. Training confusion matrix of subset 2 for left leg using complete features data.

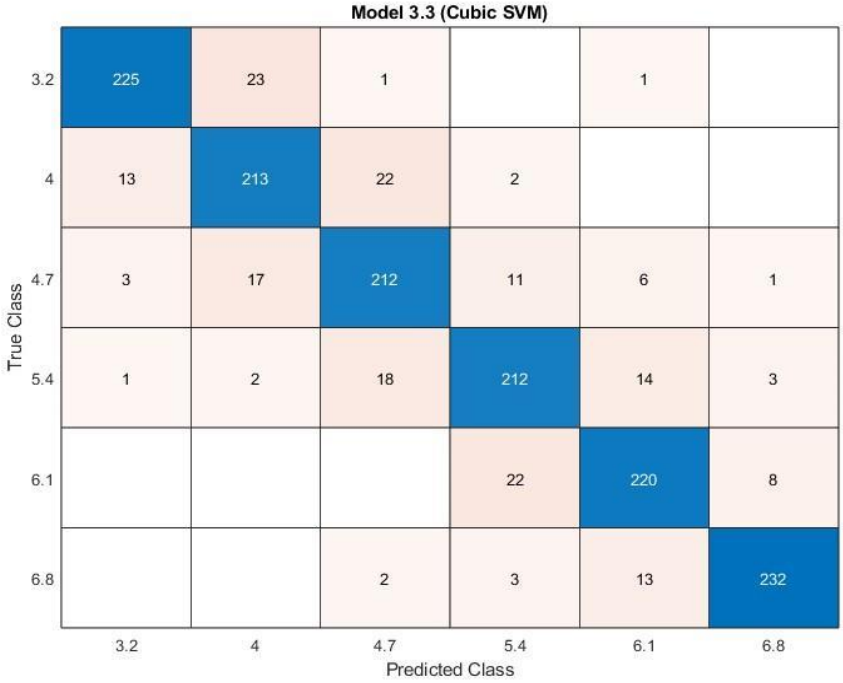

Figure 168. Test confusion matrix of subset 2 for left leg using complete features data.

These results highlight the power of well-chosen features in achieving accurate and precise speed classification.

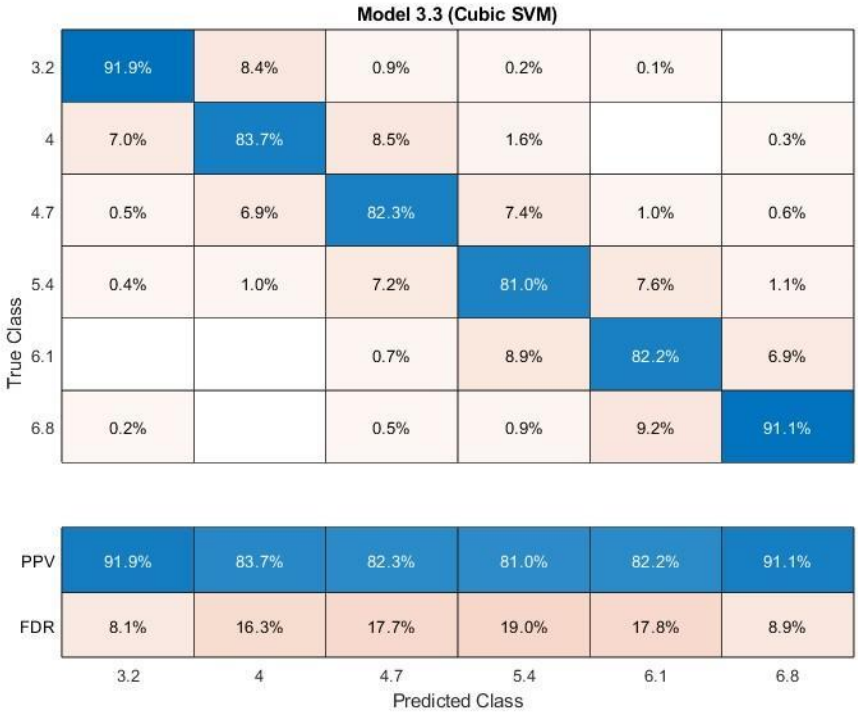

Figure 169. PPV-FDR training confusion matrix of subset 2 for left leg using complete features data.

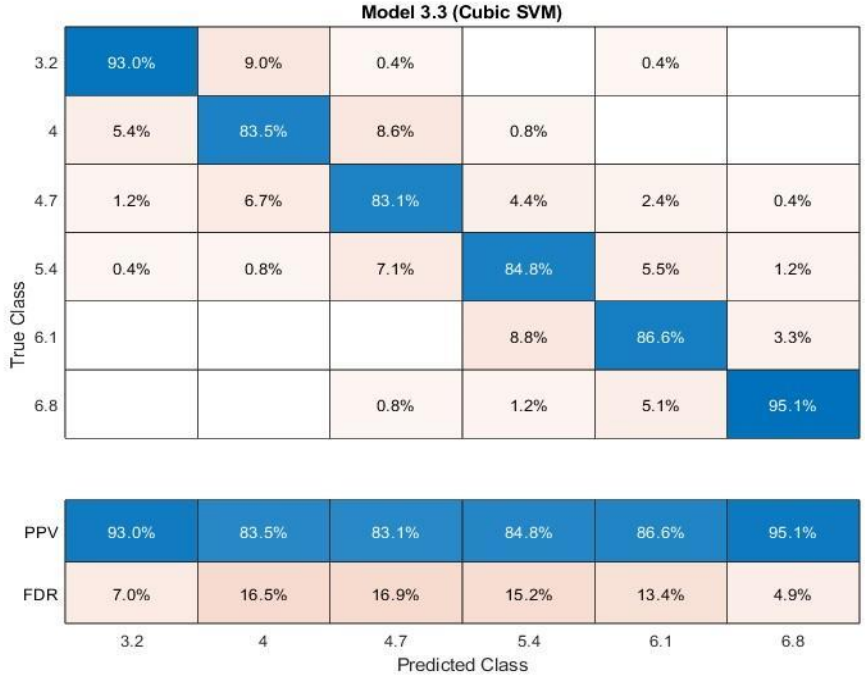

Figure 170. PPV-FDR test confusion matrix of subset 2 for left leg using complete features data.

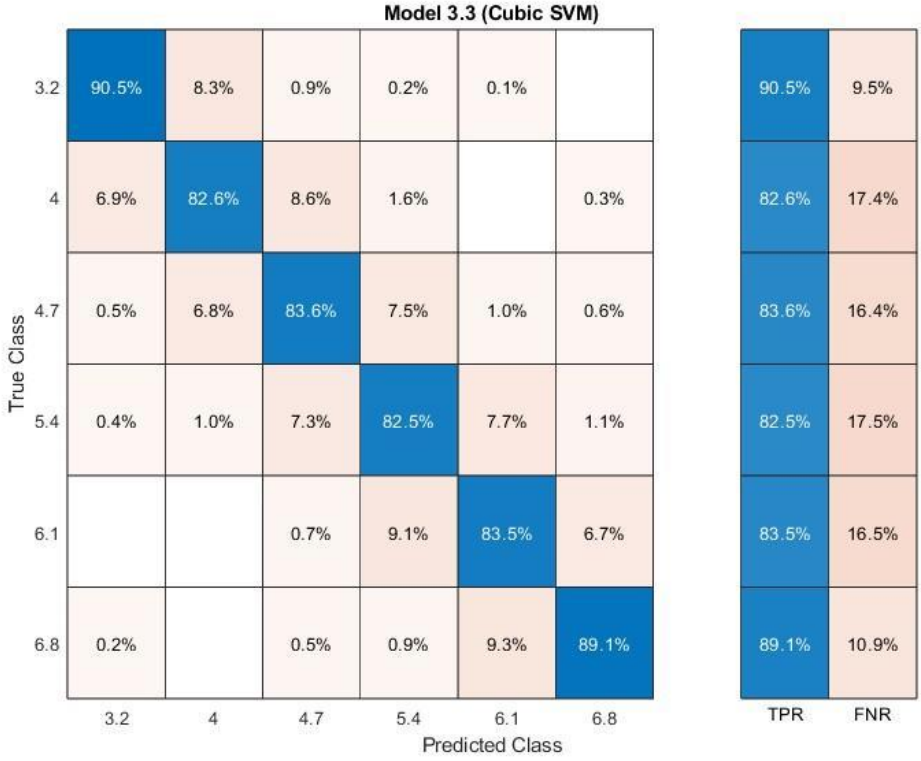

Figure 171. TPR-FNR training confusion matrix of subset 2 for left leg using complete features data.

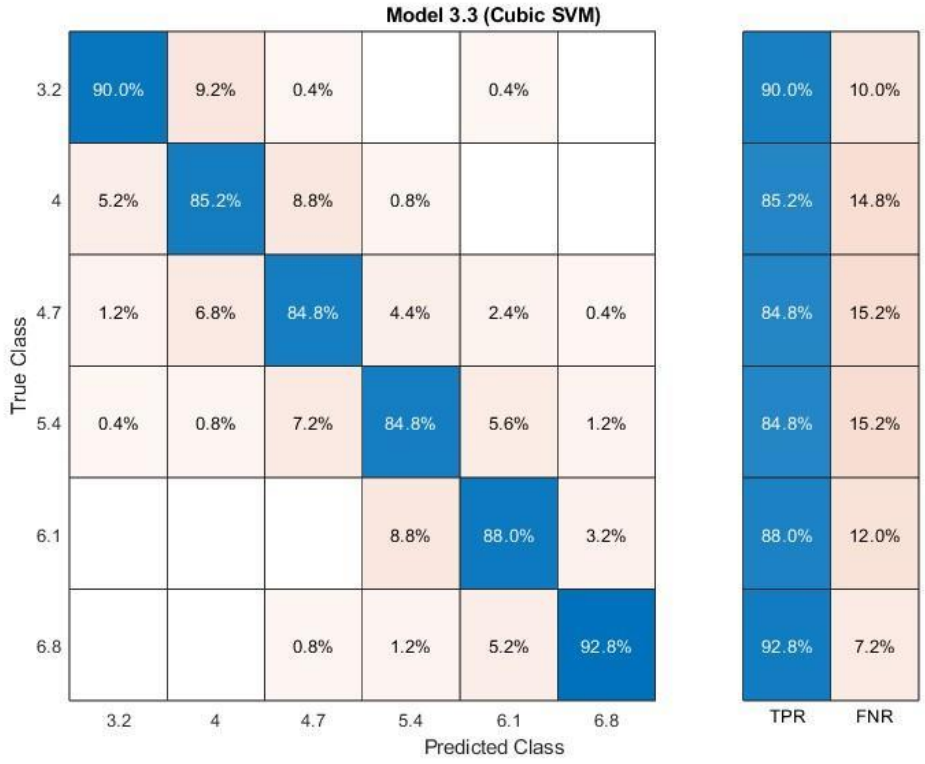

Figure 172. TPR-FNR test confusion matrix of subset 2 for left leg using complete features data.

The cubic SVM model, leveraging the discriminative power of features extracted after applying the skipped speeds method, yielded the highest accuracy in speed classification, with its performance exceeding expectations on the test set.

Table 27

Models' Accuracy Of Subset 2 For Left Leg Using Complete Features Data

Model Number	Model Type	Accuracy % (Validation)	Accuracy % (Test)
3.3	SVM	85.3	87.6
3.4	SVM	85.03333	86.46667
4.6	KNN	84.13333	84.66667
5.3	Neural Network	83.63333	85.06667
4.2	KNN	82.93333	83.26667
4.4	KNN	82.25	83.46667
3.2	SVM	82.11667	84.73333
5.2	Neural Network	80.41667	82.46667
4.1	KNN	80	82.33333
4.5	KNN	79.38333	81

Table 27 (cont.)

5.4	Neural Network	76.4	78.06667
5.1	Neural Network	75.53333	78
5.5	Neural Network	74.53333	78.73333
3.5	SVM	73.03333	75.93333
3.1	SVM	58.33333	60.4
2.1	Tree	56.81667	56.46667
4.3	KNN	56.28333	63.46667
3.6	SVM	49.43333	50.33333
2.2	Tree	42.86667	44.53333
2.3	Tree	34.25	35

Chapter 5

Discussion

This study delved into the intricate relationship between lower body kinematics and EMG activity during treadmill walking, aiming to classify gait patterns at distinct speeds. As detailed in the abstract, data were collected from 25 healthy subjects walking on a treadmill at speeds ranging from 0.8 m/s to 2 m/s. Surface EMG signals were recorded from 10 key muscles in the lower extremities, complemented by kinematic data captured using Xsens sensors.

A crucial step in this analysis involved identifying the most salient features for accurately classifying gait patterns. Initially, a comprehensive set of 27 features was derived from the EMG and kinematic data, encompassing muscle activation amplitudes, joint angles, and their respective derivatives. However, recognizing that high dimensionality can introduce noise and redundancy, a feature selection process was employed. Spearman's rank-order correlation was utilized to assess the relationships between these features, ultimately resulting in a reduced feature set of 14 variables. This refined set retained the most informative and discriminative characteristics of lower limb muscle activity and kinematics. These features represent key aspects of lower limb neuromuscular control and joint coordination during walking, expected to vary across different speeds.

Table 28

Summary Of Different Alterations And Their Accuracies

Manipulation	R_Accuracy (%)	L_Accuracy (%)
Full Data	78.49	77.29
Odd Speeds	92.23	90.97
Even Speeds	92.00	89.27
Skip 1 Frames (50 pts)	77.57	76.40
Skip 4 Frames (25 pts)	76.40	75.57

Table 28 (cont.)

Skip 5 Frames (20 pts)	75.38	74.80
Skip 10 Frames (10 pts)	68.37	66.98
Skip 20 Frames (5 pts)	57.82	57.10
Stance Phase (60 pts)	67.78	68.25
Swing Phase (40 pts)	67.26	65.14

This study investigated the performance of a cubic Support Vector Machine (SVM) model in classifying walking speeds based on surface electromyography (sEMG) data collected from five lower limb muscles. The analysis focused on evaluating the model's ability to fit the training data under various data manipulations, including alterations in the number of speeds considered, the temporal resolution of the sEMG signals, and the selection of specific gait phases.

5.1 Baseline Performance with Full Data

While using the complete dataset (all 13 speeds and 100 data points per muscle), the cubic SVM achieved accuracy of 78.49% for the right leg and 77.29% for the left leg. These values represent the baseline performance when the model is trained and tested on the full complement of available data. Although it is reasonably high, these accuracies suggest that the inherent complexity of the relationship between sEMG signals and walking speed presents a challenge for the model, even with comprehensive data. Potential sources of this complexity include inter-subject variability in muscle activation patterns, subtle variations in gait biomechanics at different speeds, and the inherent non-linearity of the sEMG-speed relationship.

5.2 Impact of Speed Subset Selection (Odd vs. Even Speeds)

Interestingly, when the model was trained and tested using only odd-numbered or even-numbered speeds, the accuracies were notably higher (92.23% and 90.97% for the right leg in odd and even, respectively; 89.27% and 92.00% for the left leg in odd and even, respectively) compared to the full data scenario. This improved performance suggests that when the model is presented with more widely spaced speeds, the differences in sEMG patterns between these speeds become more distinct,

allowing the model to fit the data more effectively. This observation highlights the potential for non-monotonic relationships between muscle activation and walking speed, where certain speeds might exhibit more distinguishable sEMG signatures than others.

5.3 Effects of Reduced Temporal Resolution

Progressively reducing the temporal resolution of the sEMG signals by skipping frames (from 50 data points down to 5 data points per muscle) resulted in a gradual decline in classification accuracy. This trend was consistent for both legs. While skipping 2, 4, and 5 frames only marginally affected the accuracy for the right and left legs, more substantial reductions (skipping 10 and 20 frames) led to more obvious drops in performance. These findings demonstrate the importance of temporal information within the sEMG signal for accurate speed classification. The higher-frequency components of the signal, which are lost when frames are skipped, appear to contain information that is valuable for the model to discern between different walking speeds. This reduction in data points might be eliminating information related to rapid changes in muscle activity that are crucial for differentiating speeds.

5.4 Stance vs. Swing Phase Contributions

The analysis of the stance and swing phases separately revealed that both phases yielded similar accuracies, with the stance phase performing slightly better for the left leg (68.25% vs. 65.14%) and the right leg performing slightly better on the swing phase (67.78% vs 67.26%). This suggests that both phases contain information relevant to speed classification, although the relative contributions of each phase might be subtle and potentially subject-dependent. These results indicate that the model is able to fit the data from both phases relatively well but also suggest that neither phase alone provides a complete picture of the sEMG-speed relationship.

5.5 Findings summary

This study investigated the use of a cubic Support Vector Machine (SVM) for classifying walking speeds based on surface electromyography (sEMG) data, exploring the model's ability to fit training data under various conditions. The baseline performance, achieved using all 13 speeds and the full temporal resolution of the

sEMG signals, reached 78.49% accuracy for the right leg and 77.29% for the left leg. These accuracies, while reasonable, suggest inherent complexities in the relationship between sEMG patterns and walking speed, possibly stemming from inter-subject variability and non-linear muscle activation patterns.

A key finding was the significantly improved accuracy when the model was trained and tested on either odd-numbered or even-numbered speeds. This result strongly indicates a non-linear relationship between walking speed and sEMG patterns, where certain speeds are characterized by more distinguishable muscle activation signatures than others. The cubic SVM, capable of modeling complex decision boundaries, appears particularly adept at capturing these non-linearities when provided with more widely spaced speed intervals. Furthermore, the study demonstrated the importance of temporal information within the sEMG signal. Progressively reducing the temporal resolution led to a decline in accuracy, emphasizing that higher-frequency components, reflecting rapid changes and finer details of muscle activity, are crucial for accurate speed classification.

Finally, the analysis of stance and swing phases individually revealed that both contribute valuable information for speed classification, yielding comparable accuracies. This supports the biomechanical understanding that both phases play distinct yet interconnected roles in the gait cycle. The slightly higher accuracies seen in the stance phase for the left leg and the swing phase for the right leg potentially indicate minor differences in how each leg's muscle patterns map to speed information. Overall, these results underscore the complex interplay between neuromuscular control, biomechanics, and walking speed. Future work should focus on evaluating the model's ability to generalize new data and exploring feature engineering to further enhance its accuracy and clinical applicability in gait analysis.

To evaluate the effectiveness of this feature selection process and explore the impact of speed variations on classification accuracy, a series of feature-based classification analyses were conducted. First, a baseline analysis was performed using the reduced feature set, achieving a classification accuracy of approximately 66% for both left and right leg datasets. Interestingly, similar accuracy levels were observed when utilizing the complete feature set, indicating that the reduced set effectively captured the essential information for gait classification.

However, a significant improvement in classification accuracy was achieved by introducing a novel approach: strategically skipping speeds during training. This strategy involved creating two distinct training datasets:

- **Odd Speed Set:** Included data from speeds 0.8 m/s, 1.0 m/s, 1.2 m/s, 1.4 m/s, and so on, effectively skipping every other speed.
- **Even Speed Set:** Included data from speeds 0.9 m/s, 1.1 m/s, 1.3 m/s, 1.5 m/s, and so on, similarly skipping every other speed.

This deliberate selection of alternating speeds aimed to amplify the differences in gait patterns between more distinct speed categories, encouraging the classification models to learn more discriminative features. Remarkably, this approach led to a substantial increase in classification accuracy, reaching approximately 85% for both odd and even speed sets. In addition, the test accuracies closely mirrored the training accuracies, confirming that the models generalized well to unseen data and did not suffer from overfitting.

It is worth noting that the highest classification accuracy was observed at the highest speed (2 m/s), where a very high percentage of gait cycles were correctly classified. This observation suggests that stronger EMG signals and more pointing differences in feature values between speeds contribute to improved classification performance. In essence, when the physiological signals are more distinct and the differences between feature values are more pronounced, the classification models can more effectively discern and categorize the different gait patterns. This phenomenon may explain the enhanced accuracy observed with speed skipping, as this strategy inherently increases the separation between speed categories and amplifies the differences in feature values.

This striking improvement in classification accuracy highlights the importance of considering speed variations in gait analysis. By strategically selecting alternating speeds during training, the classification models were able to learn more distinct and robust representations of gait patterns. This finding has significant implications for the design and control of assistive devices, where accurate and reliable classification of gait speed is crucial for intuitive and effective user interaction.

In comparison to the percentile-based approach, which achieved even higher accuracies of around 88% for training and 90% for testing with a similar speed skipping strategy, the feature-based approach with speed skipping still shows considerable improvement over the baseline feature-based analysis. This suggests that while the percentile-based approach may capture more distinct difference in the data, the feature-based approach with speed skipping offers a valuable alternative with strong performance and potential for real-world applications.

This study underscores the potential of feature engineering and strategic data manipulation techniques to enhance the performance of machine learning models in gait analysis. Future research will explore the application of these techniques in clinical populations and investigate their integration into real-time assistive device control systems.

Chapter 6

General Conclusion, Limitations & Future Work

6.1 Conclusion

This study aimed to classify gait patterns at distinct speeds by examining the relationship between lower body kinematics and EMG activity during treadmill walking. The study involved 25 healthy subjects walking on a treadmill at speeds ranging from 0.8 m/s to 2 m/s. This research included collecting surface EMG signals from 10 key muscles in the lower extremities and kinematic data using Xsens sensors. A feature selection process was employed to identify the most valuable features, and therefore reducing the redundant features and finally using a set of 14 key features.

The study found that a cubic Support Vector Machine (SVM) model achieved accuracy of 78.49% for the right leg and 77.29% for the left leg in classifying gait speeds based on sEMG data. The accuracies were notably higher when the model was trained and tested using only odd-numbered or even-numbered speeds, which suggests a non-monotonic relationship between muscle activation and walking speed.

Reducing the temporal resolution (frame reduction) of the sEMG signals resulted in a gradual decline in classification accuracy, demonstrating the importance of temporal information for accurate speed classification. Analyzing the stance and swing phases separately revealed that both phases play a significant role in speed classification, as evidenced by the similar and relatively high accuracies achieved in each phase. This suggests that both phases contain valuable and distinct information relevant to speed discrimination. Further analysis of individual muscle contributions could reveal specific muscles, such as the biceps femoris and tibilias anterior, that are involved in a particularly important role in the classification due to their high activity and distinct activation patterns. The study also explored the impact of speed variations on classification accuracy using a feature-based approach. A baseline analysis using a feature analysis approach achieved a classification accuracy of approximately 66% for both left and right leg datasets. However, a significant improvement in accuracy was achieved by introducing a novel approach: strategically skipping speeds during training. This strategy involved creating two distinct training datasets, one with data from odd-set speeds and the other with data from even-set speeds. This approach led

to a substantial increase in classification accuracy, reaching approximately 85% for both odd and even speed sets. The highest classification accuracy from all speeds was observed at the highest speed (2 m/s), suggesting that stronger EMG signals and more pronounced differences in feature values between speeds contribute to improved classification performance.

The study concluded that feature engineering and strategic data manipulation techniques, such as speed skipping, can enhance the performance of machine learning models in gait analysis. Future research will explore the application of these techniques in clinical populations and investigate their integration into real-time assistive device control systems.

6.2 Limitations and Future Work

This study provides valuable insights into the classification of gait patterns using EMG and kinematic data, but it also has limitations that can be addressed in future research.

- Limited Sample Size: The study involved 25 healthy subjects in which not a large dataset was acquired, which may not fully represent the broader population. Increasing the sample size and including participants with diverse physical characteristics and gait patterns would enhance the generalizability of the findings.

- Treadmill Walking: The study was conducted using treadmill walking, which differs from overground walking in terms of sensory feedback and biomechanics. It also affects psychologically the participant due to its limited spaces. Future studies should investigate whether the findings translate to overground walking conditions.

- EMG Signal Limitations: The accuracy of EMG signal detection can be affected by factors such as muscle depth and crosstalk from other muscles. For example, the soleus muscle's EMG signal may be influenced by the overlying gastrocnemius muscle, potentially affecting the accuracy of muscle activation amplitude measurements. Future studies could explore advanced EMG techniques or electrode placements to mitigate these limitations.

- Focus on Healthy Subjects: The study focused on healthy subjects, and the findings may not directly apply to individuals with gait abnormalities or pathological conditions. This approach on the healthy subjects wont absolutely reflect the patient population. Future research should investigate the applicability of the methods and findings in clinical populations to assess their potential for diagnostic or therapeutic purposes.

- Cubic SVM Model: The study employed a cubic SVM model, and the performance of other machine learning models could be explored in future research. Additionally, the study focused on feature-based classification, and future work could investigate alternative approaches, such as deep learning models, to further enhance classification accuracy.

- Real-time Applications: The study did not evaluate the feasibility of implementing the classification methods in real-time applications, such as assistive device control. Future research should investigate the computational efficiency and latency of the methods to determine their suitability for real-time implementation.

Addressing these limitations and exploring new research directions will contribute to a deeper understanding of gait analysis and its applications in rehabilitation and assistive technologies.

REFERENCES

Ahkami, B., Ahmed, K., Thesleff, A., Hargrove, L., & Ortiz-Catalan, M. (2023). Electromyography-Based Control of Lower limb prostheses: A Systematic review. *IEEE Transactions on Medical Robotics and Bionics*, *5*(3), 547–562. https://doi.org/10.1109/tmrb.2023.3282325

Alnuaimi, A. F., & Albaldawi, T. H. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, *97*, 00133. https://doi.org/10.1051/bioconf/20249700133

Anatomy, Bony Pelvis and Lower Limb, Gastrocnemius Muscle. (2024, January 1). Retrieved from https://pubmed.ncbi.nlm.nih.gov/30422541/

Angin, S., & Demirbüken, L. (2020a). Ankle and foot complex. In *Elsevier eBooks* (pp. 411–439). https://doi.org/10.1016/b978-0-12-812162-7.00023-0

Angin, S., & Demirbüken, L. (2020b). Ankle and foot complex. In *Elsevier eBooks* (pp. 411–439). https://doi.org/10.1016/b978-0-12-812162-7.00023-0

Bakkum, B. W., & Cramer, G. D. (2013). Muscles That Influence the Spine. In *Elsevier eBooks* (pp. 98–134). https://doi.org/10.1016/b978-0-323-07954-9.00004-9

Baldry, P., & Thompson, J. W. (2005). Pain in the lower limb. In *Elsevier eBooks* (pp. 315–324). https://doi.org/10.1016/b978-044306644-3.50022-x

Bassile, C. C., & Hayes, S. M. (2016). Gait Awareness. In *Elsevier eBooks* (pp. 194–223). https://doi.org/10.1016/b978-0-323-17281-3.00009-5

Begg, R., Palaniswami, M., & Owen, B. (2005). Support Vector Machines for Automated Gait Classification. *IEEE Transactions on Biomedical Engineering*, *52*(5), 828–838. https://doi.org/10.1109/tbme.2005.845241

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age And Ageing*, *26*(1), 15–19. https://doi.org/10.1093/ageing/26.1.15

Bohannon, R. W., & Andrews, A. W. (2011). Normal walking speed: a descriptive meta-analysis. *Physiotherapy*, *97*(3), 182–189. https://doi.org/10.1016/j.physio.2010.12.004

Bordoni, B., & Varacallo, M. (2023, April 17). Anatomy, Bony Pelvis and Lower Limb, Gastrocnemius Muscle. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK532946/

Brockett, C. L., & Chapman, G. J. (2016). Biomechanics of the ankle. *Orthopaedics and Trauma*, *30*(3), 232–238. https://doi.org/10.1016/j.mporth.2016.04.015

Chakraborty, A., & Chattaraj, S. (2023). CAGSI: A Classification Approach towards Gait Speed Identification. *Human-Centric Intelligent Systems*, *4*(1), 161–170. https://doi.org/10.1007/s44230-023-00052-0

Chambers, H. G., & Sutherland, D. H. (2002a). A Practical Guide to Gait Analysis. *Journal of the American Academy of Orthopaedic Surgeons*, *10*(3), 222–231. https://doi.org/10.5435/00124635-200205000-00009

Chambers, H. G., & Sutherland, D. H. (2002b). A Practical Guide to Gait Analysis. *Journal of the American Academy of Orthopaedic Surgeons*, *10*(3), 222–231. https://doi.org/10.5435/00124635-200205000-00009

Chang, W., Huang, W., & Lai, P. (2015). Muscle Activation of Vastus Medialis Oblique and Vastus Lateralis in Sling-Based Exercises in Patients with Patellofemoral Pain Syndrome: A Cross-Over Study. *Evidence-based Complementary and Alternative Medicine*, *2015*, 1–8. https://doi.org/10.1155/2015/740315

Choi, N., & Lee, S. (2015). Discomfort Evaluation of Truck Ingress/Egress Motions Based on Biomechanical Analysis. *Sensors*, *15*(6), 13568–13590. https://doi.org/10.3390/s150613568

Cimolato, A., Driessen, J. J. M., Mattos, L. S., De Momi, E., Laffranchi, M., & De Michieli, L. (2022). EMG-driven control in lower limb prostheses: a topic-based systematic review. *Journal of NeuroEngineering and Rehabilitation*, *19*(1). https://doi.org/10.1186/s12984-022-01019-1

Dhawan, A. (2021). EMG Signal Analysis for Identifying Walking Patterns of Normal Healthy Individuals. *www.academia.edu*. Retrieved from https://www.academia.edu/63212705/EMG_Signal_Analysis_for_Identifying_Walking_Patterns_of_Normal_Healthy_Individuals

Faisal, A. I., Majumder, S., Mondal, T., Cowan, D., Naseh, S., & Deen, M. J. (2019). Monitoring Methods of Human Body Joints: State-of-the-Art and Research Challenges. *Sensors*, *19*(11), 2629. https://doi.org/10.3390/s19112629

Farris, D. J., & Sawicki, G. S. (2012). Human medial gastrocnemius force–velocity behavior shifts with locomotion speed and gait. *Proceedings of the National Academy of Sciences*, *109*(3), 977–982. https://doi.org/10.1073/pnas.1107972109

Gao, X., He, Y., Zhang, M., Diao, X., Jing, X., Ren, B., & Ji, W. (2021). A multiclass classification using one-versus-all approach with the differential partition sampling ensemble. *Engineering Applications of Artificial Intelligence*, *97*, 104034. https://doi.org/10.1016/j.engappai.2020.104034

Gibson, V., & Prieskorn, D. (2007). The Valgus Ankle. *Foot and Ankle Clinics*, *12*(1), 15–27. https://doi.org/10.1016/j.fcl.2006.11.001

Halabi, R., Banna, I. E., Malaeb, R., Halabi, R., & Diab, M. (2019). Novel Approach for Wireless EMG Database Collection: Applied to Muscle Building Workout Routine Optimization. *IEEE*. https://doi.org/10.1109/icabme47164.2019.8940157

Hanson, S., & Jones, A. (2015). Is there evidence that walking groups have health benefits? A systematic review and meta-analysis. *British Journal of Sports Medicine*, *49*(11), 710–715. https://doi.org/10.1136/bjsports-2014-094157

Hebenstreit, F., Leibold, A., Krinner, S., Welsch, G., Lochmann, M., & Eskofier, B. M. (2015a). Effect of walking speed on gait sub phase durations. *Human Movement Science*, *43*, 118–124. https://doi.org/10.1016/j.humov.2015.07.009

Hebenstreit, F., Leibold, A., Krinner, S., Welsch, G., Lochmann, M., & Eskofier, B. M. (2015b). Effect of walking speed on gait sub phase durations. *Human Movement Science*, *43*, 118–124. https://doi.org/10.1016/j.humov.2015.07.009

Hebenstreit, F., Leibold, A., Krinner, S., Welsch, G., Lochmann, M., & Eskofier, B. M. (2015c). Effect of walking speed on gait sub phase durations. *Human*

Movement Science, *43*, 118–124. https://doi.org/10.1016/j.humov.2015.07.009

Hug, F., Vogel, C., Tucker, K., Dorel, S., Deschamps, T., Carpentier, É. L., & Lacourpaille, L. (2019). Individuals have unique muscle activation signatures as revealed during gait and pedaling. *Journal of Applied Physiology*, *127*(4), 1165–1174. https://doi.org/10.1152/japplphysiol.01101.2018

Jacob, S. (2008). Lower limb. In *Elsevier eBooks* (pp. 135–179). https://doi.org/10.1016/b978-0-443-10373-5.50009-9

Kharazi, M., Theodorakis, C., Mersmann, F., Bohm, S., & Arampatzis, A. (2023). Contractile Work of the Soleus and Biarticular Mechanisms of the Gastrocnemii Muscles Increase the Net Ankle Mechanical Work at High Walking Speeds. *Biology*, *12*(6), 872. https://doi.org/10.3390/biology12060872

Kibushi, B. (2023). Muscle coordination patterns in regulation of medial gastrocnemius activation during walking. *Human Movement Science*, *90*, 103116. https://doi.org/10.1016/j.humov.2023.103116

Kim, P., Lee, J., & Shin, C. S. (2021a). Classification of Walking Environments Using Deep Learning Approach Based on Surface EMG Sensors Only. *Sensors*, *21*(12), 4204. https://doi.org/10.3390/s21124204

Kim, P., Lee, J., & Shin, C. S. (2021b). Classification of Walking Environments Using Deep Learning Approach Based on Surface EMG Sensors Only. *Sensors*, *21*(12), 4204. https://doi.org/10.3390/s21124204

Kunju, N. (2016a). EMG Signal Analysis for Identifying Walking Patterns of Normal Healthy Individuals. *Iitm*. Retrieved from https://www.academia.edu/26356792/EMG_Signal_Analysis_for_Identifying_Walking_Patterns_of_Normal_Healthy_Individuals

Kunju, N. (2016b). EMG Signal Analysis for Identifying Walking Patterns of Normal Healthy Individuals. *Iitm*. Retrieved from https://www.academia.edu/26356792/EMG_Signal_Analysis_for_Identifying_Walking_Patterns_of_Normal_Healthy_Individuals

Kunju, N., Kumar, N., Pankaj, D., Dhawan, A., & Kumar, A. (2009). EMG Signal Analysis for Identifying Walking Patterns of Normal Healthy Individuals. *Indian Journal of Biomechanics: Special Issue (NCBM 7-8 March 2009)*. Retrieved from https://core.ac.uk/download/pdf/34212669.pdf

Maharaj, J. N., Cresswell, A. G., & Lichtwark, G. A. (2019). Tibialis anterior tendinous tissue plays a key role in energy absorption during human walking. *Journal of Experimental Biology*. https://doi.org/10.1242/jeb.191247

McKeon, J. M. M., & Hoch, M. C. (2019). The Ankle-Joint Complex: A Kinesiologic Approach to Lateral Ankle Sprains. *Journal of Athletic Training*, *54*(6), 589–602. https://doi.org/10.4085/1062-6050-472-17

MejiaCruz, Y., Franco, J., Hainline, G., Fritz, S., Jiang, Z., Caicedo, J. M., . . . Hirth, V. (2021). Walking Speed Measurement Technology: a Review. *Current Geriatrics Reports*, *10*(1), 32–41. https://doi.org/10.1007/s13670-020-00349-z

Meyer, C., Killeen, T., Easthope, C. S., Curt, A., Bolliger, M., Linnebank, M., . . . Filli, L. (2019). Familiarization with treadmill walking: How much is enough? *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41721-0

Morbidoni, C., Cucchiarelli, A., Fioretti, S., & Di Nardo, F. (2019a). A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. *Electronics*, *8*(8), 894. https://doi.org/10.3390/electronics8080894

Morbidoni, C., Cucchiarelli, A., Fioretti, S., & Di Nardo, F. (2019b). A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. *Electronics*, *8*(8), 894. https://doi.org/10.3390/electronics8080894

Morbidoni, C., Cucchiarelli, A., Fioretti, S., & Di Nardo, F. (2019c). A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. *Electronics*, *8*(8), 894. https://doi.org/10.3390/electronics8080894

Morbidoni, C., Cucchiarelli, A., Fioretti, S., & Di Nardo, F. (2019d). A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking. *Electronics*, *8*(8), 894. https://doi.org/10.3390/electronics8080894

Moreira, L., Figueiredo, J., Fonseca, P., Vilas-Boas, J. P., & Santos, C. P. (2021a). Lower limb kinematic, kinetic, and EMG data from young healthy humans during walking at controlled speeds. *Scientific Data*, *8*(1). https://doi.org/10.1038/s41597-021-00881-3

Moreira, L., Figueiredo, J., Fonseca, P., Vilas-Boas, J. P., & Santos, C. P. (2021b). Lower limb kinematic, kinetic, and EMG data from young healthy humans during walking at controlled speeds. *Scientific Data*, *8*(1). https://doi.org/10.1038/s41597-021-00881-3

Murtagh, E. M., Mair, J. L., Aguiar, E., Tudor-Locke, C., & Murphy, M. H. (2020). Outdoor Walking Speeds of Apparently Healthy Adults: A Systematic Review and Meta-analysis. *Sports Medicine*, *51*(1), 125–141. https://doi.org/10.1007/s40279-020-01351-3

Olewnik, Ł., Zielinska, N., Paulsen, F., Podgórski, M., Haładaj, R., Karauda, P., & Polguj, M. (2020). A proposal for a new classification of soleus muscle morphology. *Annals of Anatomy - Anatomischer Anzeiger*, *232*, 151584. https://doi.org/10.1016/j.aanat.2020.151584

Padulo, J., Rampichini, S., Borrelli, M., Buono, D. M., Doria, C., & Esposito, F. (2023). Gait Variability at Different Walking Speeds. *Journal of Functional Morphology and Kinesiology*, *8*(4), 158. https://doi.org/10.3390/jfmk8040158

Park, H., Han, S., Sung, J., Hwang, S., Youn, I., & Kim, S. (2023a). Classification of gait phases based on a machine learning approach using muscle synergy. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1201935

Park, H., Han, S., Sung, J., Hwang, S., Youn, I., & Kim, S. (2023b). Classification of gait phases based on a machine learning approach using muscle synergy. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1201935

Péter, A., Andersson, E., Hegyi, A., Finni, T., Tarassova, O., Cronin, N., . . . Arndt, A. (2019). Comparing Surface and Fine-Wire Electromyography Activity of Lower Leg Muscles at Different Walking Speeds. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.01283

Phinyomark, A., Khushaba, R. N., Ibáñez-Marcelo, E., Patania, A., Scheme, E., & Petri, G. (2017a). Navigating features: a topologically informed chart of electromyographic features space. *Journal of the Royal Society Interface*, *14*(137), 20170734. https://doi.org/10.1098/rsif.2017.0734

Phinyomark, A., Khushaba, R. N., Ibáñez-Marcelo, E., Patania, A., Scheme, E., & Petri, G. (2017b). Navigating features: a topologically informed chart of electromyographic features space. *Journal of the Royal Society Interface*, *14*(137), 20170734. https://doi.org/10.1098/rsif.2017.0734

Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems With Applications*, *39*(8), 7420–7431. https://doi.org/10.1016/j.eswa.2012.01.102

Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., & Laurillau, Y. (2013). EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems With Applications*, *40*(12), 4832–4840. https://doi.org/10.1016/j.eswa.2013.02.023

Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006a). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, *8*(1), 11–35. https://doi.org/10.1251/bpo115

Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006b). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, *8*(1), 11–35. https://doi.org/10.1251/bpo115

Reeves, J., Starbuck, C., & Nester, C. (2020). EMG gait data from indwelling electrodes is attenuated over time and changes independent of any

experimental effect. *Journal of Electromyography and Kinesiology*, *54*, 102461. https://doi.org/10.1016/j.jelekin.2020.102461

Research on Gait Recognition Based on Lower Limb EMG Signal. (2021a, August 8). Retrieved from https://ieeexplore.ieee.org/document/9512759

Research on Gait Recognition Based on Lower Limb EMG Signal. (2021b, August 8). Retrieved from https://ieeexplore.ieee.org/document/9512759

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6). https://doi.org/10.1007/s42979-021-00815-1

Schmidt, C. M., Szatkowski, J. P., & Riehl, J. T. (2020a). Tibial Plateau Fracture. In *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.92684

Schmidt, C. M., Szatkowski, J. P., & Riehl, J. T. (2020b). Tibial Plateau Fracture. In *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.92684

Semaan, M. B., Wallard, L., Ruiz, V., Gillet, C., Leteneur, S., & Simoneau-Buessinger, E. (2021). Is treadmill walking biomechanically comparable to overground walking? A systematic review. *Gait & Posture*, *92*, 249–257. https://doi.org/10.1016/j.gaitpost.2021.11.009

Soofi, A. A., & Awan, A. (2017, January 5). Classification Techniques in Machine Learning: Applications and Issues. Retrieved from https://set-publisher.com/index.php/jbas/article/view/1715

Stevens, B. M., Nichols, B. R., Doty, H. I., & Korak, J. A. (2022, August 1). Muscle Activation Patterns of the Proximal Medial and Distal Biceps Femoris and Gluteus Maximus Among 6 Hip Extension and Knee Flexion Exercises in

Trained Women. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC9362892/

Toledo-Pérez, D. C., Rodríguez-Reséndiz, J., Gómez-Loenzo, R. A., & Jauregui-Correa, J. C. (2019). Support Vector Machine-Based EMG Signal Classification Techniques: A Review. *Applied Sciences*, *9*(20), 4402. https://doi.org/10.3390/app9204402

Ugbolue, U. C., Robson, C., Donald, E., Speirs, K., Dutheil, F., Baker, J. S., . . . Gu, Y. (2021a). Joint Angle, Range of Motion, Force, and Moment Assessment: Responses of the Lower Limb to Ankle Plantarflexion and Dorsiflexion. *Applied Bionics and Biomechanics*, *2021*, 1–13. https://doi.org/10.1155/2021/1232468

Ugbolue, U. C., Robson, C., Donald, E., Speirs, K., Dutheil, F., Baker, J. S., . . . Gu, Y. (2021b). Joint Angle, Range of Motion, Force, and Moment Assessment: Responses of the Lower Limb to Ankle Plantarflexion and Dorsiflexion. *Applied Bionics and Biomechanics*, *2021*, 1–13. https://doi.org/10.1155/2021/1232468

Vandervoort, A. A., & McComas, A. J. (1983). A comparison of the contractile properties of the human gastrocnemius and soleus muscles. *European Journal of Applied Physiology and Occupational Physiology*, *51*(3), 435–440. https://doi.org/10.1007/bf00429079

Wang, J., Dai, Y., Kang, T., & Si, X. (2021). Research on Gait Recognition Based on Lower Limb EMG Signal. *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*, 212–217. https://doi.org/10.1109/icma52036.2021.9512759

Winner, T. S., Rosenberg, M. C., Berman, G. J., Kesar, T. M., & Ting, L. H. (2024a). Gait signature changes with walking speed are similar among able-bodied young adults despite persistent individual-specific differences. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2024.05.01.591976

Winner, T. S., Rosenberg, M. C., Berman, G. J., Kesar, T. M., & Ting, L. H. (2024b). Gait signature changes with walking speed are similar among able-bodied young adults despite persistent individual-specific differences. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2024.05.01.591976

Yamaguchi, S., Sasho, T., Kato, H., Kuroyanagi, Y., & Banks, S. A. (2009). Ankle and Subtalar Kinematics during Dorsiflexion-Plantarflexion Activities. *Foot & Ankle International*, *30*(4), 361–366. https://doi.org/10.3113/fai.2009.0361

Yao, X., Li, H., & Xiu, C. (2024). Biomechanical Analysis of Injury Risk in Two High-Altitude Landing Positions Using Xsens Inertial Units and EMG Sensors. *Sensors*, *24*(21), 6822. https://doi.org/10.3390/s24216822

Yousefi, J., & Hamilton-Wright, A. (2014). Characterizing EMG data using machine-learning tools. *Computers in Biology and Medicine*, *51*, 1–13. https://doi.org/10.1016/j.compbiomed.2014.04.018