İnsansız Su Üstü Aracı için Pekiştirmeli Öğrenme Temelli Yörünge Planlama

Gülce Yücel Bedir

DOKTORA TEZİ

Endüstri Mühendisliği Anabilim Dalı

Nisan 2025

Trajectory Planning Based on Reinforcement Learning for Unmanned Surface Vehicle

Gülce Yücel Bedir

DOCTORAL DISSERTATION

Department of Industrial Engineering

Nisan 2025

İnsansız Su Üstü Aracı için Pekiştirmeli Öğrenme Temelli Yörünge Planlama

Gülce Yücel Bedir

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Eğitim ve Öğretim Yönetmeliği Uyarınca
Endüstri Mühendisliği Anabilim Dalı
Üretim ve Servis Sistemleri Bilim Dalında
DOKTORA TEZİ
olarak hazırlanmıştır.

Danışman: Prof. Dr. İnci Sarıçiçek
İkinci Danışman: Prof. Dr. Ahmet Yazıcı

Nisan 2025

ONAY

Endüstri Mühendisliği Anabilim Dalı Doktora öğrencisi Gülce Yücel Bedir'in DOKTORA tezi olarak hazırladığı "İnsansız Su Üstü Aracı için Pekiştirmeli Öğrenme Temelli Yörünge Planlama" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oybirliği ile kabul edilmiştir.

Danışman : Prof. Dr. İnci Sarıçiçek

İkinci Danışman : Prof. Dr. Ahmet Yazıcı

Doktora Tez Savunma Jürisi:

Üye: Prof. Dr. İnci Sarıçiçek

Üye: Prof. Dr. Kemal Özkan

Üye: Doç. Dr. Servet Hasgül

Üye: Prof. Dr. Gürkan Öztürk

Üye: Doç. Dr. Emre Çimen

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ……..…….. tarih ve ……..…… sayılı kararıyla onaylanmıştır.

Prof. Dr. ....

Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. İnci Sarıçiçek ve Prof. Dr. Ahmet Yazıcı danışmanlığında hazırlamış olduğum "İnsansız Su Üstü Aracı için Pekiştirmeli Öğrenme Temelli Yörünge Planlama" başlıklı Doktora tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallarına uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 30/04/2025

Gülce Yücel Bedir

ÖZET

Otonom sistemlerin gelişimiyle birlikte, İnsansız Su Üstü Araçlar (Unmanned Surface Vehicles-USV), deniz ortamında güvenli ve verimli görevlerin gerçekleştirilmesine olanak sağlamaktadır. Bu tez çalışmasında, USV'lerin çarpışma risklerini en aza indirerek, Uluslararası Denizde Çarpışmayı Önleme Kurallarına (COLREGs) uygun şekilde hareket etmelerini sağlayacak otonom bir sistem geliştirilmiştir. Amaç, gerçek zamanlı yörünge planlaması ile çarpışmalardan kaçınan, kurallara uygun ve çevresel koşullara duyarlı bir navigasyon sistemi tasarlamaktır. Geliştirilen sistemde, Derin Q Ağı (Deep Q Network-DQN) modeli tabanlı pekiştirmeli öğrenme yöntemi önerilmiştir. Yörünge üretiminde B-Spline eğrileri kullanılarak aracın manevra kabiliyeti modellenmiş, kontrol noktaları aracılığıyla çevreye uyumlu yumuşak eğriler elde edilmiştir. Pekiştirmeli öğrenme tabanlı yörünge planlamada dalga boyu, rüzgâr ve akıntı hızı gibi çevresel etkenler dikkate alınmaktadır. Önerilen yöntem, farklı simülasyon senaryolarında test edilmiş, doğrusal ve açısal hız değerleri ile theta açısının değişimi gözlemlenmiştir. Farklı öğrenme oranları ve sinir ağı yapılarına göre yapılan karşılaştırmalarda, elde edilen Q değerleri üzerinden sistem performansı analiz edilmiştir. Deneysel sonuçlar, önerilen pekiştirmeli öğrenme tabanlı yörünge planlama sisteminin başarılı sonuçlar verdiğini ortaya koymaktadır. Yöntem, yalnızca tek bir eğitim süreci ile benzer senaryolara doğrudan uygulanabilirlik sunarak, geleneksel yöntemlere kıyasla önemli ölçüde esneklik ve hız avantajı sağlamaktadır. Bu özellikler, otonom deniz araçlarıyla gerçekleştirilen görevlerin operasyonel verimliliğini artıracaktır.

Anahtar Kelimeler: İnsansız Su Üstü Araç, Yörünge Planlama, Pekiştirmeli Öğrenme, COLREG Kuralları, Çevresel Etkenler, Çarpışmadan Kaçınma

SUMMARY

The advancement of autonomous systems has enabled unmanned surface vehicles (USVs) to conduct safe and efficient operations in marine environments. In this thesis, an autonomous system is developed to enable USVs to navigate in compliance with the International Regulations for Preventing Collisions at Sea (COLREGs) while minimizing the risk of collisions. The goal is to design a navigation system that performs real-time trajectory planning, avoids collisions, and adapts to environmental conditions. In the developed system, a reinforcement learning method based on the Deep Q-Network (DQN) model is proposed. B-spline curves are used for trajectory generation, modeling the maneuverability of the USV, and producing smooth, adaptive paths through control points. Reinforcement learning-based trajectory planning integrates environmental conditions such as wave height, wind speed, and current velocity. The proposed method was tested in various simulation scenarios, and changes in linear and angular velocity values as well as the theta angle were observed. System performance was evaluated based on Q-values obtained under different learning rates and neural network architectures. Experimental results reveal that the proposed reinforcement learning-based trajectory planning system yields successful outcomes. The system provides direct applicability to similar scenarios with a single training process, offering significant flexibility and speed advantages compared to traditional methods. These features will enhance the operational efficiency of tasks carried out by autonomous maritime vehicles.

Keywords: Unmanned Surface Vehicle, Trajectory Planning, Reinforcement Learning, COLREG Rules, Environmental Factors, Collision Avoidance

TEŞEKKÜR

Doktora eğitimim ve tez çalışmam boyunca emeği geçen, bilgi ve birikimleriyle bana yol gösteren, rehberlik eden değerli danışman hocalarım Sayın Prof. Dr. İnci Sarıçiçek ve Sayın Prof. Dr. Ahmet Yazıcı'ya en içten teşekkürlerimi sunarım. Engin bilgileri, sabırları ve sürekli destekleri olmasaydı bu çalışma mümkün olamazdı.

Tez çalışmam süresince yapıcı katkılarını ve değerli önerilerini esirgemeyen Tez İzleme Komitesi üyeleri Sayın Prof. Dr. Kemal Özkan ve Sayın Doç. Dr. Servet Hasgül hocalarıma da teşekkür ederim.

Bu zorlu süreçte yanımda olan, her daim desteğini hissettiren, moral ve motivasyon kaynağım sevgili eşim Cem Bedir'e; gülümsemesiyle hayatıma neşe ve enerji katan kızım Nil Bedir'e; sevgileri ve desteğiyle bugünlere gelmemi sağlayan canım aileme, annem Havva Yücel, babam İlyas Yücel ve kardeşim Yıldız Yücel Karacan'a en derin teşekkürlerimi sunarım. Varlıkları, sevgileri ve inançları her zaman en büyük güç kaynağım olmuştur.

Son olarak, bu çalışmanın başarıyla tamamlanmasında katkısı olan, beni cesaretlendiren ve sabırla yanımda olan tüm dostlarıma teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET ..... **vi**
SUMMARY ..... **vii**
TEŞEKKÜR ..... **viii**
İÇİNDEKİLER ..... **ix**
ŞEKİLLER DİZİNİ ..... **xi**
ÇİZELGELER DİZİNİ ..... **xiii**
SİMGELER VE KISALTMALAR DİZİNİ ..... **xiv**
1. GİRİŞ VE AMAÇ **1**
2. LİTERATÜR ARAŞTIRMASI **6**
2.1. Uluslararası Çarpışmaları Önleme Kuralları 9
2.2. Yörünge Planlama 11
2.3. Pekiştirmeli Öğrenme 14
2.3.1. Derin pekiştirmeli öğrenme 21
2.3.2. Derin Q ağı 22
3. MATERYAL VE YÖNTEM **25**
3.1. Materyal 25
3.2. Problemin Tanımlanması ve Örnek Çözüm Yaklaşımı 26
3.3. İnsansız Su Üstü Araç için Yörünge Planlama 32
3.3.1. Yörünge planlamada iç faktörler 32
3.3.2. Yörünge planlamada dış faktörler 34
3.4. USV Yörünge Planlama için Pekiştirmeli Öğrenme 36
3.5. Pekiştirmeli Öğrenme Temelli Yörünge Planlama Sistemi Akış Diyagramı 44
4. BULGULAR VE TARTIŞMA **47**
4.1. USV Aracın Kinematik Model Doğrulaması 47
4.2. Pekiştirmeli Öğrenme Hiper Parametreleri 49

İÇİNDEKİLER

Sayfa

4.3. Kafa Kafaya Çarpışma Senaryosu.....52
4.4. Arkadan Gelerek Sollama Senaryosu.....55
4.5. Yol Veren Çarpışma Senaryosu.....57
4.6. Karşıdan Karşıya Geçme Senaryosu.....60
4.7. Q ve Kayıp Değerleri.....63
4.8. Pekiştirmeli Öğrenme Temelli Yörünge Planlama Model Doğrulaması.....79
5. SONUÇ VE ÖNERİLER.....82
KAYNAKLAR DİZİNİ.....83

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

1.1. Operasyonlarının yürütüldüğü filo kontrol merkezi 2
1.2. Farklı yörünge eğrileri arasından seçilen yörünge 4
2.1. Yıllar temelinde İnsansız Su Üstü Araçlarda rotalama ile ilgili yayın sayısı 6
2.2. İnsansız Su Üstü Araçlarında yapılan çalışmalarda kullanılan yöntemler 7
2.3. İnsansız Su Üstü Araçlarda çakışmasız rota planlamanın ele alınma durumu 8
2.4. COLREG'ler altında USV'lerin karşılaşma durumları 10
2.5. Yörünge planlamasında alternatif yörüngelerin gösterimi 12
2.6. RL modeli 15
2.7. DQN modelinin temel bileşenleri ve bilgi akış yapısı 24
3.1. Sabit ve hareketli engellerin olduğu ortamda RL ile yörünge seçimi 28
3.2. 0-200 m arası bölüm bazlı kontrol noktaları ve yörüngeler 31
3.3. USV kinematik denklemi 33
3.4. Pekiştirmeli öğrenmeye dayalı yörünge planlama yapısı 37
3.5. Önerilen DQN modeli 39
3.6. Pekiştirmeli öğrenme temelli yörünge planlama sisteminin akış diyagramı 45
3.7. Sistem kullanıcı arayüzü 46
4.1. Engelsiz ortamda USV hareketi 48
4.2. Dalga etkisinde USV hareketi 48
4.3. Akıntı etkisinde USV hareketi 49

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

4.4. Kafa kafaya çarpışma senaryosunda gözlemlenen olası senaryolar 52
4.5. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri 53
4.6. Sollama senaryosunda gözlemlenen olası senaryolar 55
4.7. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri 56
4.8. Yol Verme senaryosunda gözlemlenen olası senaryolar 58
4.9. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri 59
4.10. Karşıdan karşıya geçen senaryosunda gözlemlenen olası senaryolar 61
4.11. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri 62
4.12. 20000 zaman dilimi eğitilen ajan Q/Kayıp Grafiği 64
4.13. 1. senaryo 0-2000 zaman dilimi 65
4.14. 2. senaryo 2000-4000 zaman dilimi 65
4.15. 3. senaryo 4000-6000 zaman dilimi 66
4.16. 4. senaryo 6000-8000 zaman dilimi 67
4.17. 5. senaryo 8000-10000 zaman dilimi 67
4.18. 6. senaryo 10000-12000 zaman dilimi 68
4.19. 7. senaryo 12000-14000 zaman dilimi 69
4.20. 8. senaryo 14000-16000 zaman dilimi 70
4.21. 9.senaryo 16000-18000 zaman dilimi 71
4.22. 10. senaryo 18000-20000 zaman dilimi 71
4.23. Senaryo sayısına bağlı olarak öğrenme durumu 78
4.24. Engelli ortamda USV hareketi 80
4.25. Kafa kafaya senaryosunda yörünge takibi 80
4.26. Sollama senaryosunda yörünge takibi 81

ÇİZELGELER DİZİNİ

Cizelge **Sayfa**

3.1. Y_0 yörüngesi için 0-25 m aralığında kinematik denklem adım hesaplamaları 29
3.2. Y_0 yörüngesi için 50-75 m aralığında kinematik denklem adım hesaplamaları 29
3.3. 0-25 m ve 50-75 m arasında Y_0 yörüngesinde elde edilen sonuçlar 30
3.4. İç faktörler 34
3.5. Dış faktörler 34
3.6. Pekiştirmeli öğrenme tabanlı yörünge planlama hesabında kullanılan girdiler.....38
4.1. Pekiştirmeli öğrenme hiper parametreleri 50
4.2. Kafa kafaya senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları.....72
4.3. Arkadan sollama senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları 74
4.4. Yol veren çarpışma senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları 75
4.5. Karşıya geçme senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları 76

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler	**Açıklama**
x, y	İki boyutlu kartezyen koordinatları (konum)
θ	Başlangıç açısı (heading angle)
V	Doğrusal hız (forward speed) (m/s)
W	Açısal hız (angular speed) (rad/s)
h	Dalga boyu (metre)
q	Dalga yönü (derece)
V_c	Akıntı hızı (m/s) (Copernicus sisteminden alınan değer)
F	Rüzgâr hızı (m/s)
α	Rüzgâr yönü ile gemi arasındaki açı (radyan)
W_d	Gemi yükü (ton)
V_a	Geminin gerçek hızı (knot)
V_0	Geminin deniz şartlarını göz önüne almadan normal hızı (knot)
V_x, V_y	Boylamsal ve enlemsel akıntı hızları
V_0	Geminin normal şartlar altındaki hızı (m/s)
θ	Başlangıç açısı (heading angle)
αc	Akıntı ile gemi arasındaki açı (radyan)
γ	İskonto faktörü (Discount Factor)
ε	Keşif oranı (Exploration rate)

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
APF	Yapay Potansiyel Alan (Artificial Potential Field)
COLREG	Uluslararası Çarpışmaları Önleme Kuralları (International Regulations for Preventing Collisions at Sea)
DQN	Derin Q Ağı (Deep Q Network)
GA	Genetik Algoritmalar (Genetic Algorithms)
MDP	Markov Karar Süreci (Markov Decision Process)
MSE	Ortalama Kare Hatası (Mean Squared Error)
RELU	Doğrultulmuş Lineer Birim (Rectified Linear Unit)
RL	Pekiştirmeli Öğrenme (Reinforcement Learning)
UAV	İnsansız Hava Aracı (Unmanned Aerial Vehicle)
UGV	İnsansız Kara Aracı (Unmanned Ground Vehicle)

USV	İnsansız Su Üstü Aracı (Unmanned Surface Vehicle)
UUV	İnsansız Su Altı Aracı (Unmanned Underwater Vehicle)
WOS	Bilimsel Araştırma Veritabanı (Web of Science)

1. GİRİŞ VE AMAÇ

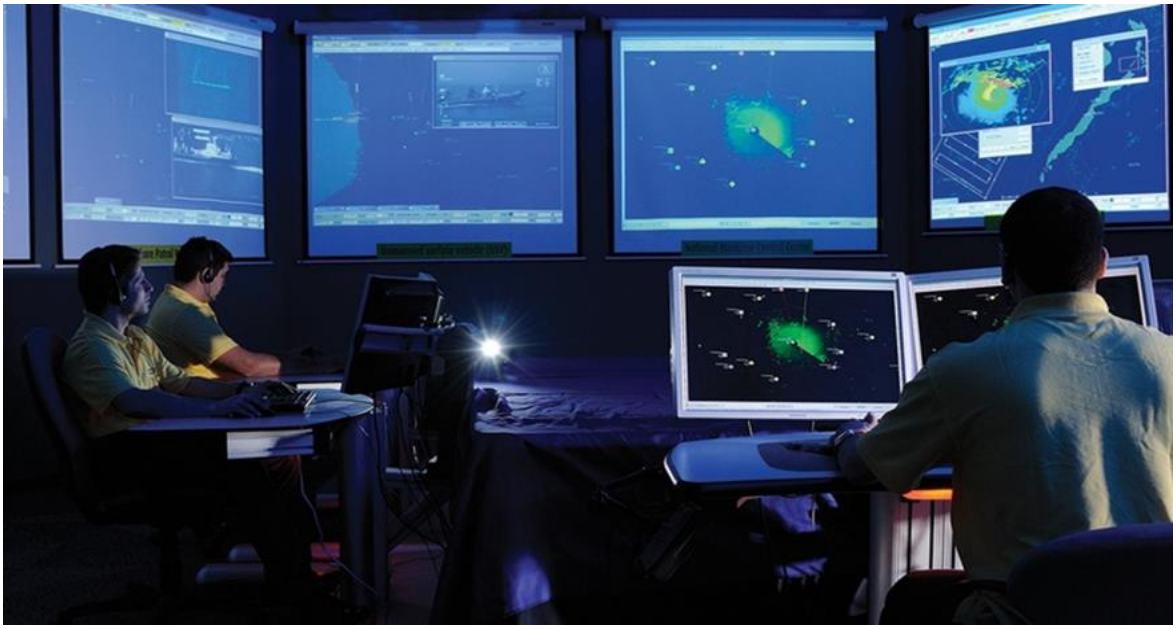
Otonom araç teknolojilerinin gelişimi, öncelikli olarak İnsansız Hava Araçları (Unmanned Aerial Vehicles –UAV) ile başlamış, bu araçlar askeri ve sivil amaçlarla geniş bir kullanım alanı bulmuştur. UAV sistemlerinin başarısı, kara ve deniz ortamlarına yönelik otonom sistem araştırmalarının artmasına öncülük etmiştir (Austin, 2010). 2006 yılından itibaren ise İnsansız Su Üstü Araçları (Unmanned Surface Vehicles – USV), özellikle denizcilik alanında ön plana çıkmış, çeşitli görevleri bağımsız biçimde yerine getirebilen araçlar olarak yaygınlık kazanmıştır (Manley, 2008). Bu süreçte, USV teknolojileri sivil ve askeri uygulamalarda bağımsız görev gerçekleştirebilecek şekilde önemli gelişmeler göstermiştir (Xu, Tang ve Cheng, 2014).

İnsansız araç sistemleri, çoğunlukla belirli bir bölgenin güvenliğini sağlamak veya riskli görevleri insansız olarak yerine getirmek amacıyla tasarlanmış uzaktan yönetilebilen veya otonom kontrole sahip sistemlerdir. Bu tür araçlar kullanım alanlarına göre üç ana gruba ayrılmaktadır. İnsansız Kara Araçları (Unmanned Ground Vehicle-UGV), İnsansız Deniz Araçları (Unmanned Marine Vehicles-UMV) ve İnsansız Hava Araçları (Unmanned Aerial Vehicles-UAV). Denizcilik alanında ise İnsansız Deniz Araçları, İnsansız Su Üstü Araçları (Unmanned Surface Vehicles_USV) ve İnsansız Su Altı Araçları (Unmanned Underwater Vehicles _UUV) olarak ikiye ayrılmaktadır (Kahveci vd., 2017).

USV'lerin temel görevleri arasında su üstü harbi, keşif ve gözetleme faaliyetleri, mayın karşı tedbirleri ve denizaltı savunma harbi gibi askeri operasyonlar yer almaktadır. Denizcilik ortamındaki değişen koşullara göre İnsansız Su Üstü Araçları üzerinde araştırmalar devam etmektedir. USV'lere olan talep, yalnızca askeri alanda değil, aynı zamanda deniz araştırmaları, kıyı gözetimi, insan kurtarma ve deniz kirliliği kontrolü gibi özel sektörde de artmıştır. USV'lerde, suya düşen insan arama, şüpheli deniz araçlarının izlenmesi ve bölgeden çıkarılması, belirlenen bir sahada veya limanda geçişlerin kontrolü gibi görevler bulunmaktadır. Bu görevler kapsamında, denizdeki çevresel koşullara göre hız, manevra kabiliyeti, faydalı yük kapasitesi, yakıt tankı hacmi, yakıt tüketimi gibi teknik özelliklerle donatılmakta, görev ihtiyaçlarına uygun esnek çözümler sunmaktadır. Bu

bağlamda, filo kontrol merkezleri görev önceliklerine ve çevresel koşullara göre uygun USV'leri seçerek görevlendirme yapmaktadır (Park vd., 2017).

USV operasyonlarının yönetimi genellikle filo kontrol merkezleri üzerinden gerçekleştirilmektedir. Bu merkezler, birden fazla USV'nin eş zamanlı olarak takip edildiği ve görevlerinin koordine edildiği yerlerdir. Şekil 1.1'de, bir USV filosunun operasyonlarının yürütüldüğü örnek bir kontrol merkezi yer almaktadır.



Şekil 1.1. Operasyonlarının yürütüldüğü filo kontrol merkezi

Filo kontrol merkezlerinde görev atamaları, çevresel durum analizi ve USV'lerin anlık takibi gibi işlemler gerçekleştirilmektedir.

Otonom araçlara olan ilgi ve ihtiyacın artmasıyla birlikte, USV'lerde yapay zekâ temelli yaklaşımların kullanımı da hız kazanmıştır. USV'ler için güvenli navigasyonun sağlanmasında en kritik sorunlardan biri çarpışma önleme problemidir. Bu kapsamda geliştirilen Otonom Çarpışma Önleme Algoritmaları, çevresel etkenlere duyarlı ve uluslararası denizcilik kurallarına uygun kararlar alabilen otonom sistemler geliştirilmesini hedeflemektedir.

USV'ler, deniz ortamında faaliyet gösterdikleri için; dalga boyu, rüzgâr ve akıntı hızı gibi dış etkenlerden önemli ölçüde etkilenmektedir. Bu koşullar altında, sadece engellerden kaçınmak yeterli değildir. Aynı zamanda uluslararası denizcilik kuralları doğrultusunda da hareket etmeleri gerekir. Bu durum, USV'lerin karar alma süreçlerini daha karmaşık ve çok boyutlu hale getirmektedir. Araç yalnızca engellerden kaçınmakla kalmaz, çevredeki diğer araçların hız ve yönlerini, mevcut çevresel şartları ve görev hedefini aynı anda dikkate almak zorundadır. Bu nedenle, USV'lerin akıllı ve esnek sistemlerle donatılması, özellikle yoğun trafiğe sahip veya belirsizlik içeren deniz koşullarında oldukça kritik hale gelmiştir.

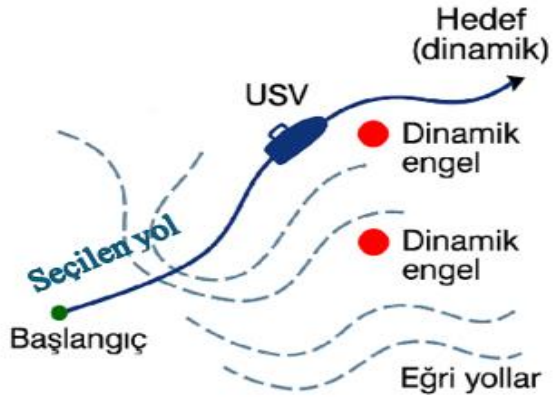
Literatürde, USV'lerin yörünge planlaması için birçok farklı yöntem kullanılmıştır (Zhou vd., 2020). Bu yöntemlerin çoğu, en kısa yolu bulmaya ve statik engellerden kaçınmaya odaklanır. Yaygın olarak kullanılan bazı yöntemler şunlardır:

- Genetik Algoritmalar (Genetic Algorithms): En uygun çözümü, farklı çözüm kombinasyonları arasından seçerek bulur (Li ve Yang 2020).
- A* Algoritması (A* Algorithm): Harita üzerinde, başlangıç noktasından hedefe en kısa yolu bulmak için kullanılan klasik bir arama algoritmasıdır (Zhuang vd., 2020).
- Yapay Potansiyel Alan (Artificial Potential Field): Hedef noktayı çekici bir kuvvet gibi modelleyerek aracı hedefe yönlendirirken, engellerden itici kuvvetlerle uzaklaşmasını sağlar (Wang vd., 2022).
- Karınca Kolonisi Optimizasyonu (Ant Colony Optimization): Gerçek karıncaların en kısa yolu feromon izleriyle bulma davranışından esinlenir (Yang vd., 2019).

Bu yöntemlerin çoğu deniz ortamının dinamik yapısını yeterince dikkate almaz. Yani, değişen çevre koşullarına uyum sağlama ve deneyimden öğrenme gibi özellikleri sınırlıdır.

Bu çalışmada, USV'lerin manevra kabiliyeti ve çevresel koşullar dikkate alınarak çarpışmalardan kaçınmasını sağlayacak Pekiştirmeli Öğrenme (Reinforcement Learning-RL) tabanlı yörünge planlama önerilmiştir. RL ajanı, ödül-ceza yapısına dayalı olarak Uluslararası Çarpışmaları Önleme Kurallarına (COLREG) uygun, çarpışmalardan uzak ve

hedef odaklı yörüngeleri seçmeyi öğrenmektedir. Şekil 1.2 önerilen sistemin işleyişini genel olarak göstermektedir.



Şekil 1.2. Farklı yörünge eğrileri arasından seçilen yörünge

Bu şekilde açık mavi çizgiler alternatif yolları, kalın mavi çizgi ise RL ajanı tarafından seçilen güvenli ve uygun yörüngeyi temsil etmektedir. Kırmızı noktalar, çevredeki dinamik engelleri ve potansiyel çarpışma tehditlerini göstermektedir. Bir USV'nin başlangıç noktasından bir hedefe ulaşması için her adımda farklı yörünge eğrileri arasında seçim yapması gerektiği bir senaryo sunulmuştur. RL ajanı, COLREG kuralları ve çevresel etkenlere duyarlı bir ödül fonksiyonu aracılığıyla, çarpışmadan kaçınmayı öğrenerek en uygun yörüngeyi belirlemektedir.

USV'ye farklı kontrol noktaları üzerinden çok sayıda B-Spline eğrisi sunulmakta ve RL yöntemi bu eğrilerden birini seçmektedir. Böylece hem esnek yörüngeler elde edilmekte hem de çarpışmalardan kaçınılırken kurallara uygun kararlar alınabilmektedir.

Pekiştirmeli öğrenme, son yıllarda otonom sistemlerde öne çıkan güçlü bir yapay zekâ yöntemidir. Bu yöntem, bir ajanın çevresiyle etkileşime girerek deneme-yanılma yoluyla öğrenmesini sağlar. Ajan, yaptığı her eylemden sonra bir ödül ya da ceza alır ve zamanla en uygun davranış biçimini öğrenir. Statik (sabit) ve dinamik (hareketli) engellerin bulunduğu karmaşık ortamlarda RL çok etkili sonuçlar verebilmektedir. Bu nedenle, çevresel değişkenlerin yoğun olduğu USV sistemlerinde RL tabanlı çözümler büyük önem taşımaktadır.

Bu çalışmanın literatüre katkılarından biri, USV'nin özgün kinematik yapılarına uygun şekilde, yörünge planlamasında RL algoritmalarının kullanılması ve bu algoritmaların B-Spline eğrileriyle entegre edilmesidir. Böylece araç, hem çevresel koşullara duyarlı esnek yörüngeler oluşturmakta hem de COLREG kurallarına uygun kararlar alabilmektedir.

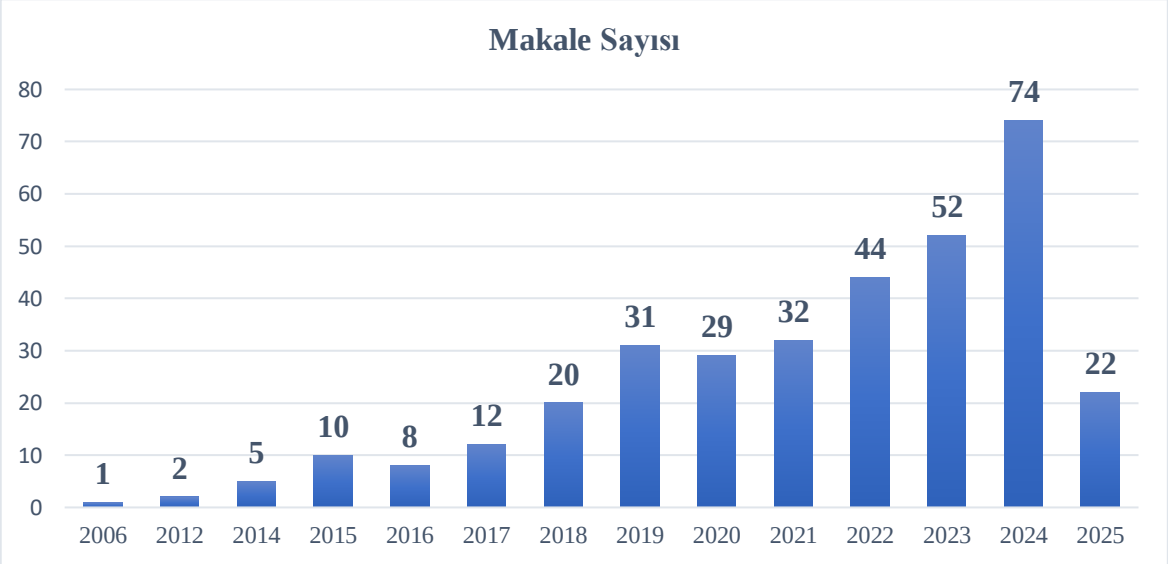
Önerilen yapının önemli bir avantajlardan biri, ödül-ceza fonksiyonlarının sabit değil çevresel durumlara göre dinamik biçimde tasarlanmış olmasıdır. Literatürde genellikle sabit veya sezgisel ödül yapıları tercih edilirken, bu çalışmada ödüller hem USV'nin güvenliği hem de COLREG kurallarına uygunluk temelinde yapılandırılmıştır (Bergman vd., 2020). Bu durum sistemin öğrenme sürecinde daha anlamlı ve verimli geri bildirim almasını sağlamaktadır.

Literatürde genellikle ilgili bileşenlerin her biri (kinematik modelleme, çevresel etkenler, COLREG kuralları, pekiştirmeli öğrenme, yörünge planlama) ayrı çalışmalarda ele alınmıştır. Ancak, USV'nin hem kinematik denkleminden gelen iç parametreler hem dış ortam parametreleri bir araya getirilerek gerçek hayata uyum sağlayan bir çalışmaya ihtiyaç vardır. Bu çalışma ise tüm bu bileşenleri entegre ederek disiplinler arası bir yaklaşım sunmakta ve gerçek uygulamalara geçişte önemli bir adım olarak öne çıkmaktadır.

Çalışmanın geri kalanı şu şekilde yapılandırılmıştır: Bölüm 2'de literatür araştırması sunulmaktadır. Bu bölümde, İnsansız Su Üstü Araçları, COLREG'ler, yörünge planlama ve pekiştirmeli öğrenme ile ilgili önceki çalışmaların bir özeti yapılmıştır. Bölüm 3'te materyal, problem tanımı, örnek çözüm yaklaşımı, İnsansız Su Üstü Araçlarda yörünge planlama, yörünge planlamada ve pekiştirmeli öğrenmede ele alınan parametreler ve akış diyagramı açıklanmaktadır. Bölüm 4'te, Pekiştirmeli Öğrenme Algoritması ile ödül-ceza fonksiyonları, grafikler, çözüm önerileri, model doğrulama sunulmaktadır. Son bölümde ise, çalışmanın sonuçları yer almakta olup, elde edilen bulgular ve öneriler tartışılmaktadır.

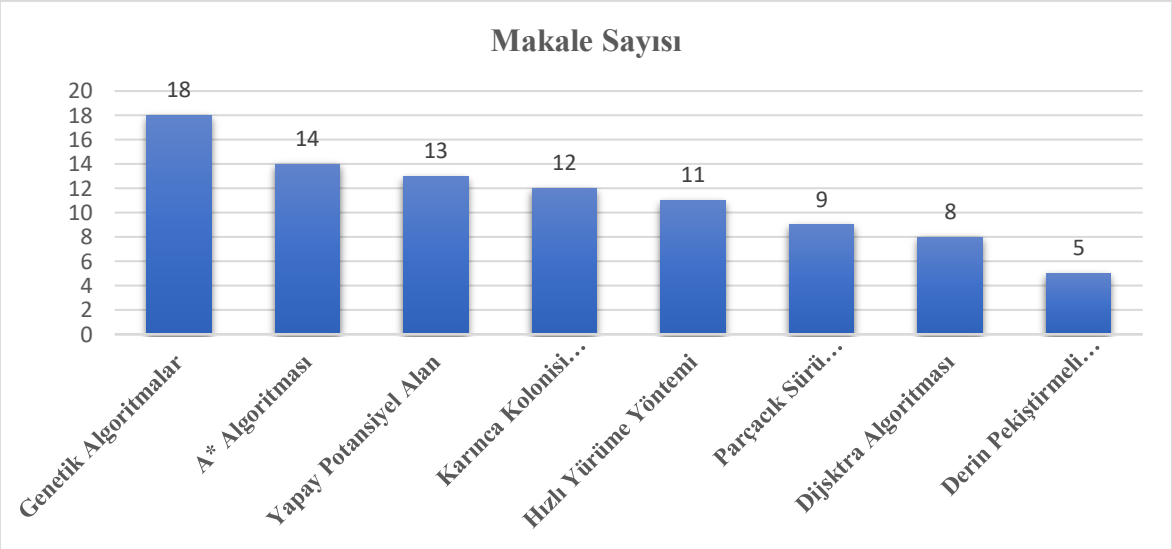
2. LİTERATÜR ARAŞTIRMASI

Yapılan literatür araştırması sonucunda, "İnsansız Su Üstü Araçları (Unmanned Surface Vehicles_USV)" ve ilgili "yol planlama, görev atama ve rotalama" konularında yapılan akademik çalışmaların son yıllarda belirgin bir artış gösterdiği tespit edilmiştir. Web of Science (WOS) veri tabanındaki 2006-2025 yılları arasını kapsayan aramalar, özellikle 2006 yılından sonra bu alanda gerçekleştirilen araştırmalarda dikkate değer bir artış olduğunu ortaya koymaktadır. Şekil 2.1'de sunulan istatistiksel veriler, USV odaklı çalışmaların yıllara göre artan eğilimini açıkça yansıtmaktadır.



Şekil 2.1. Yıllar temelinde İnsansız Su Üstü Araçlarda rotalama ile ilgili yayın sayısı

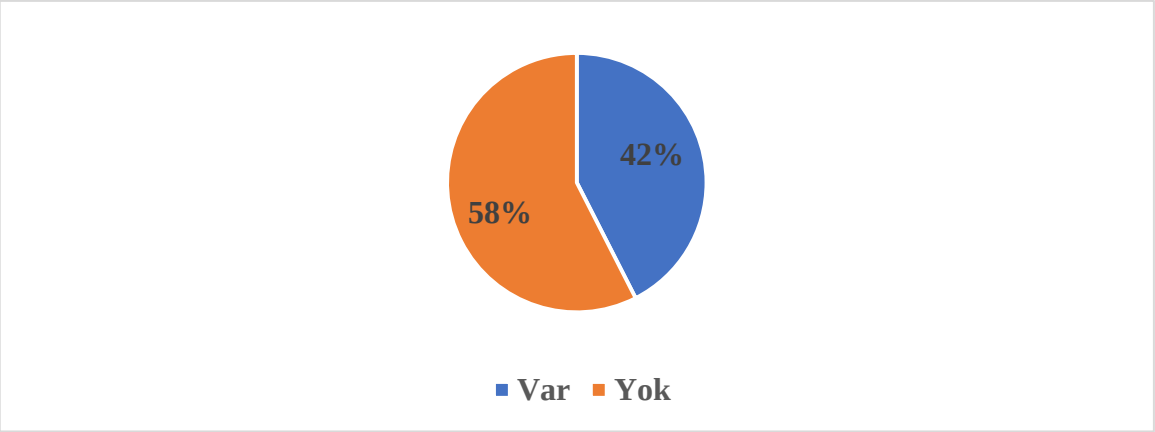
Şekil 2.1'de, özellikle 2006 yılından itibaren USV'ler ile ilgili rotalama ve yol planlama çalışmalarında belirgin bir artış gözlemlenmektedir. USV uygulamalarında yol planlama ve görev atama problemlerinde en sık kullanılan yöntemler Şekil 2.2'de gösterilmiştir.



Şekil 2.2. İnsansız Su Üstü Araçlarında yapılan çalışmalarda kullanılan yöntemler

Literatür incelemelerinde en sık karşılaşılan algoritmalar arasında Genetik Algoritmalar, A* Algoritması, Hızlı Yürüme Yöntemi (Fast Marching Method), Yapay Potansiyel Alan Yöntemi ve Dijkstra Algoritması yer almaktadır (Zhou vd., 2020). Şekil 2.2'deki veriler, USV'lere yönelik rota planlama ve görev atama problemlerinde hangi yöntemlerin ön plana çıktığını göstermektedir.

Meta-sezgisel yöntemlerin tercih edilme nedenlerinin başında, gerçek dünya problemleri için kısa sürede uygun çözümler sunabilme yetenekleri gelmektedir. Bu tür yöntemler, özellikle karmaşık ve değişken çevresel koşullar altında etkili çözümler sağlayabilmektedir. Çakışmasız rota planlamanın ele alınma durumu Şekil 2.3'te gösterilmiştir.



Şekil 2.3. İnsansız Su Üstü Araçlarda çakışmasız rota planlamanın ele alınma durumu

Literatür araştırmasının sonuçlarına göre, Şekil 2.3'te USV'ler üzerine yapılan çalışmaların yaklaşık %42'sinde çakışmasız rota planlama (collision-free path planning) konusuna odaklanmaktadır (Zhou vd., 2020). Bu çalışmaların çoğunda, dinamik ve statik engellerden kaçınmaya yönelik stratejiler geliştirilmiş, özellikle meta-sezgisel algoritmaların gerçek dünya problemlerine yakınsayan çözümler sunma yeteneği nedeniyle tercih edildiği gözlemlenmiştir.

Literatürdeki çalışmaların büyük çoğunluğu, hedef noktaya ulaşmak için en kısa mesafeyi belirlemeye odaklanmaktadır. Bu yöntemlerdeki dezavantajlar yerel minimuma yakalanması, gerçek hayat problemlerine uygulama zorluğu ve çözüm sürelerinin uzun olmasıdır.

Bu çalışmanın odak noktası olan güvenli navigasyon ihtiyacı, hedefe yaklaşırken belirli güvenlik limitlerinin korunmasını gerektirmektedir. Bu bağlamda, mevcut çözümlerin çoğunun USV'ler için gerekli olan hassas ve güvenli planlamayı sağlayamadığı görülmektedir. Ayrıca, Su Üstü Araçlarının maruz kaldığı dalga, rüzgâr ve akıntı gibi çevresel etkilerin de yörünge planlamasında göz önünde bulundurulması büyük önem taşımaktadır. Bu faktörler dikkate alınmadan yapılan planlamalar, özellikle dinamik ortamlarda faaliyet gösteren USV'ler açısından hedefe ulaşmada önemli sapmalara yol açabilmektedir.

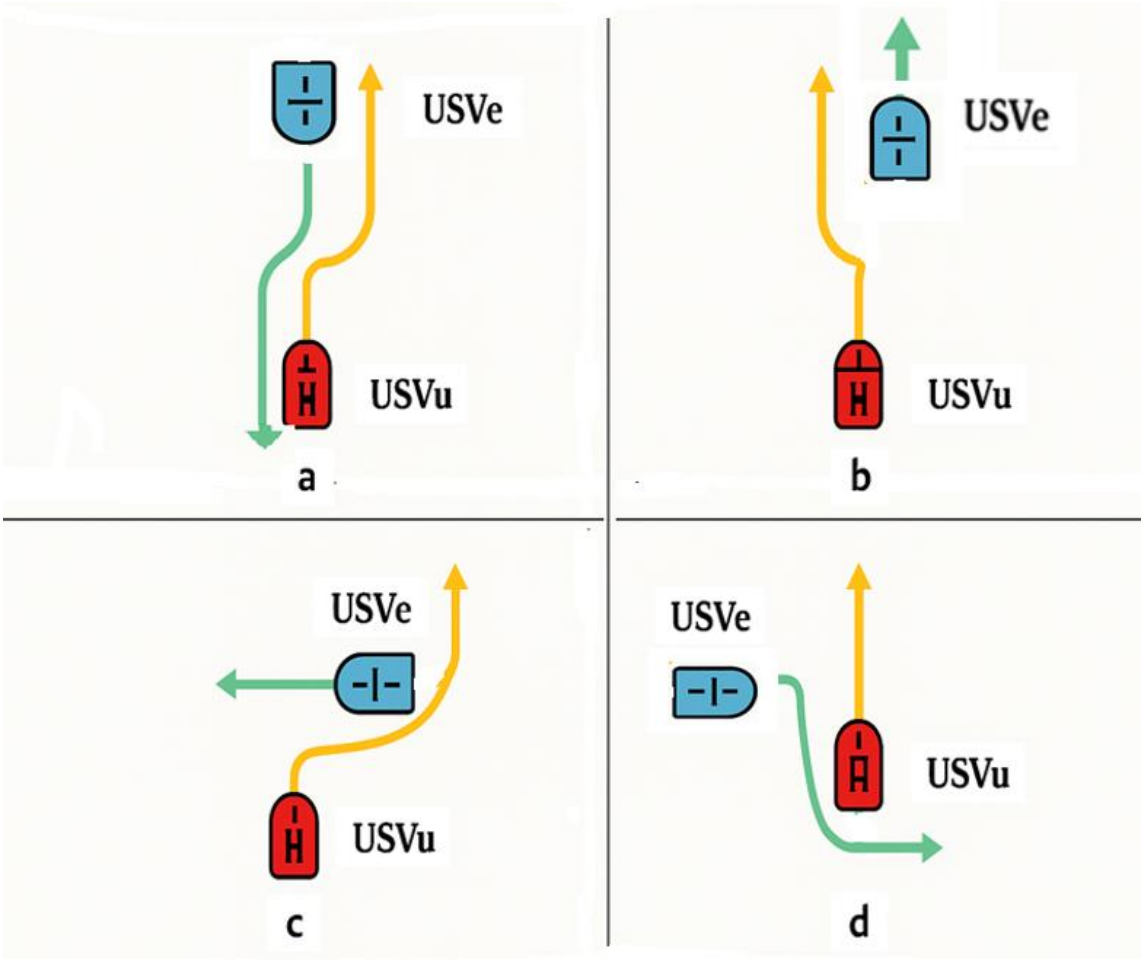
Literatürdeki birçok sezgisel ve meta-sezgisel yöntem, çevresel koşulları planlama sürecine entegre edemediği için bu ihtiyaçlara yeterince cevap verememektedir. Bu nedenle,

bu çalışmada çevresel değişkenleri doğrudan dikkate alarak pekiştirmeli öğrenme yaklaşımı önerilmiştir. Pekiştirmeli öğrenme, çevresel faktörlere uyum sağlama ve belirsizlik içeren ortamlarda karar verme yeteneği sayesinde, dinamik sistemlerde yol planlaması yapabilen güçlü bir yapay zekâ yöntemidir. Bu özellikleriyle, çalışmada ele alınan farklı çevresel koşullar, çarpışma önleme kuralları, araç kinematiği vb. karakteristikler dikkate alınarak yörünge planlama problemin çözümüne yönelik en uygun yöntem olarak öne çıkmaktadır.

2.1. Uluslararası Çarpışmaları Önleme Kuralları

Uluslararası Çarpışmaları Önleme Kuralları (International Regulations for Preventing Collisions at Sea -COLREG), USV'lerin karşılaşmalar sırasında güvenli bir şekilde hareket etmelerini sağlamak için kritik bir rol oynamaktadır. COLREG'lerin karşılaşmalarda USV davranışını doğru bir şekilde kısıtlayabilmesi, algoritma tasarımı açısından büyük bir öneme sahiptir. USV'nin çarpışmadan kaçınma hareketini gerçekleştirmesi gerektiği durumlar, algoritmanın karşılaşan diğer USV'nin konum ve hareket bilgilerini doğru bir şekilde algılamasını gerektirir.

Literatür incelendiğinde COLREG kuralları arkadan yaklaşma, kafa kafaya karşılaşma, çapraz geçiş ve yol verme durumu olmak üzere genellikle dört temel senaryo üzerinden sınıflandırılmaktadır. COLREG kapsamında, USV'lerin çeşitli karşılaşma durumlarına ilişkin senaryolar Şekil 2.4'te sunulmaktadır. Bu senaryolar, İnsansız Su Üstü Araçlarının diğer deniz araçlarıyla karşılaştıklarında hangi yönde, ne tür bir manevra gerçekleştirmesi gerektiğini tanımlayan kurallardır. Yapılan çalışmaların bir kısmında bu kurallar öğrenme algoritmalarına entegre edilmiş ödül fonksiyonları aracılığıyla modellenirken, diğer bir kısmında ise sabit kurallar mantığıyla sistem davranışı yönlendirilmiştir (Xie vd., 2021; Tamura vd., 2019).



Şekil 2.4. COLREG'ler altında USV'lerin karşılaşma durumları

Şekil 2.4'te, USV arasındaki karşılaşma durumlarını ve bu durumlarda uygulanacak çarpışmadan kaçınma davranışlarını göstermektedir. COLREG kurallarına göre, her bir karşılaşma durumu farklı bir kaçınma davranışı gerektirir.

Kafa kafaya çarpışma: Şekil 2.4(a), iki USV'nin zıt yönler veya zıt yönlere yakın bir şekilde seyretmesi sonucunda ortaya çıkan kafa kafaya karşılaşmasını göstermektedir. Bu tip karşılaşma durumunda, her iki USV'nin de çarpışmadan kaçınmak için manevra yapması gerekmektedir. COLREG kurallarına göre, her iki araç da yön değiştirmeli ve birbirlerine yaklaşmalarını engellemek için uygun manevra yapmalıdır.

Arkadan gelerek sollama: Şekil 2.4(b) durumunda bir USVe'nin arkasından gelen diğer bir USVu'nun öndeki USVe'yi geçmek amacıyla sağından veya solundan geçmesi gerektiği durumu gösterilmektedir. Bu tip karşılaşmalarda, arkadaki USVu'nun güvenli bir şekilde geçiş yapabilmesi için gerekli önlemleri alması gerekir. COLREG'lere göre, geçiş yapacak

olan USVu'nun, öndeki USVe'yi güvenli bir şekilde geçmesi ve gerektiğinde hızını azaltması veya yönünü değiştirmesi beklenir.

Yol veren çarpışma: Şekil 2.4(c)'deki durumda, önünden geçen USVe engelinden kaçınmak için, USVu'nun sancak tarafına dönerek çapraz bir yol izleyip, arkasından geçmesi gerektiği durum gösterilmektedir. COLREG kurallarına göre, USVu'nun yön değiştirmesi ve başka bir aracın yolunu geçmesine izin vermesi gerekebilir. Bu tür bir kaçınma manevrası, güvenli bir yörüngeyi sağlamak amacıyla yapılır ve diğer araçların hareketine göre dinamik olarak yönlendirilir.

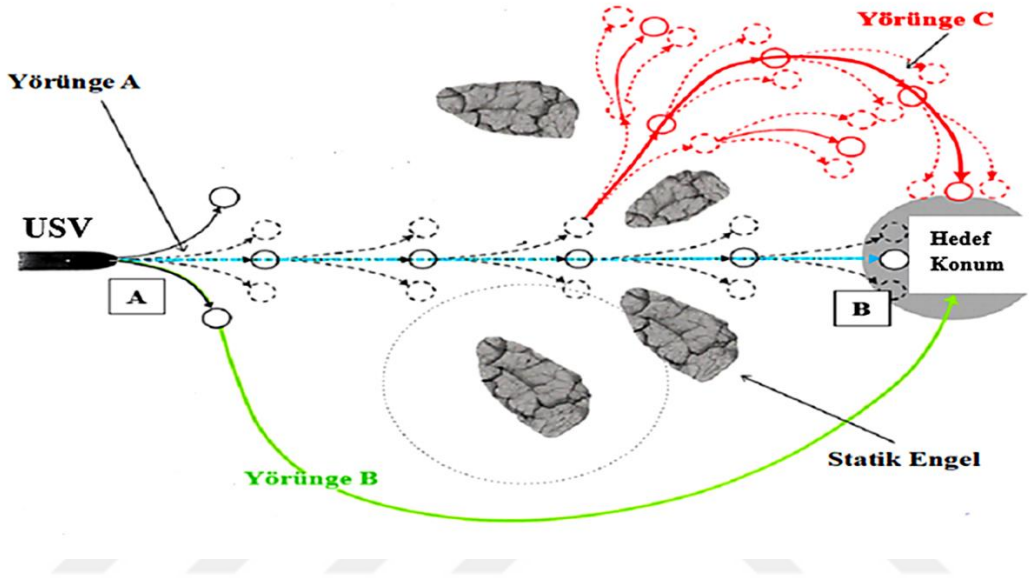
Karşıdan karşıya geçme: Şekil 2.4(d)'de karşıdan karşıya geçen USV karşılaşma durumu gösterilmiştir. Bu durumda bir USVu'nun, karşısındaki USVe ile karşılaşma durumunda kaçınma sorumluluğu yoktur. COLREG'lere göre, bu tür durumlarda kaçınma sorumluluğu, yalnızca diğer aracın manevra yapma yükümlülüğü altında olan tarafın sorumluluğundadır. Dolayısıyla, bu durumda her iki araç da kendi yörüngesinde seyretmeye devam edebilir.

Bu davranışlar, aracın güvenli seyrinin sürdürülmesi ve etkili bir navigasyonun sağlanması açısından kritik bir rol oynamaktadır (Fan ve Sun, 2022). Bu senaryolar ortaya çıktığında USV'nin çarpışma olmayacak şekilde yörünge planlaması önemlidir.

2.2. Yörünge Planlama

İnsansız Su Üstü Araçları için çarpışma önlemeye yönelik yörünge planlama algoritmaları üzerine yapılan araştırmalar son yıllarda önemli bir gelişim göstermiştir. Yörünge planlama (Trajectory Planning), bir aracın belirli bir başlangıç noktasından hedefe doğru aracın kinematiğini ve başlangıç yönelimini dikkate alarak güvenli hareket planı oluşturmasıdır. Bu süreç, genellikle engellerden kaçınma, enerji tüketimini minimize etme, hız ve zaman kısıtlamaları gibi çeşitli faktörler altında bir yol tasarımı yapılmasını gerektirir (Zhou vd., 2020). Yörünge planlamanın temel amacı, İnsansız Su Üstü Aracının hareket kabiliyetini ve çevresel koşulları dikkate alarak, dinamik ortamlarda güvenli ve verimli bir şekilde yol almasını sağlamaktır. Bu amaç doğrultusunda, USV'nin kinematik modeli ve hareket dinamikleri detaylı bir şekilde tanımlanmakta; hız, yön, dönüş yarıçapı gibi fiziksel özellikleri dikkate alınmaktadır.

Yörünge planlama, endüstriyel robotlar, İnsansız Hava Araçları, İnsansız Su Üstü Araçları ve benzeri sistemlerde güvenli bir hareket sağlamak amacıyla kullanılır. Şekil 2.5 yörünge planlamada farklı yörüngelerin değerlendirilmesini, engellerden kaçınma stratejilerini ve belirsizlik altında üretilen alternatif yörüngeleri örneklemektedir (Zhou vd., 2020).



Şekil 2.5. Yörünge planlamasında alternatif yörüngelerin gösterimi

Şekil 2.5'te A noktasından B noktasına hareket etmekte olan USV için rota üzerindeki farklı yörünge seçeneklerini göstermektedir. Aracın başlangıç noktasından hedef konuma ulaşmak için izleyebileceği üç farklı yörünge yer almaktadır. Yörünge A, engellerden geniş bir açıyla kaçınarak güvenli ancak daha uzundur. Yörünge B, daha kısa bir yol olmasına rağmen, çevredeki statik engellere yakın geçiş yaptığı için riskli bir alternatiftir. Yörünge C ise pekiştirmeli öğrenme yöntemi ile elde edilmiştir. Bu yörünge üzerinde yer alan düğüm noktalarından çıkan kırmızı noktalı eğriler, algoritmanın her adımda çevresel belirsizliklere bağlı olarak üretebileceği alternatif yörüngeleri temsil etmektedir.

Yörünge planlama sürecinde, yalnızca hedefe ulaşmak değil, aynı zamanda hareketin güvenli ve belirli kısıtlamalar altında gerçekleşmesi hedeflenmektedir. Bu nedenle hem çevresel engeller hem de belirsizlik içeren karar noktaları dikkate alınmaktadır.

Yörünge planlama ve esnek yol üretimi için etkili yöntemlerden biri de B-Spline Algoritmasıdır. Bu yöntem, aracın hareketini yumuşak eğrilerle temsil ederek hem manevra kabiliyetini hem de çevresel uyumu artırmaktadır. B-Spline (Basis Spline) Algoritması, kesikli noktalar arasındaki yolun pürüzsüz bir şekilde tasarlanması için yaygın olarak kullanılan matematiksel bir yöntemdir (de Boor, 1978). B-Spline eğrileri, belirli kontrol noktalarından geçerek bu noktalar arasında esnek ve pürüzsüz geçişler sağlayan bir yol oluşturur. B-Spline yöntemi, kontrol noktalarının belirli bir sırayla ve uygun aralıklarla yerleştirilmesiyle, eğrinin istenilen geometrik ve kinematik özellikleri taşıması sağlanabilir. Bu özellikler arasında eğrinin pürüzsüz olması, aracın belirli bir hızda ve yumuşak geçişlerle hareket edebilmesi yer almaktadır. Böylece B-Spline eğrileri, aracın kesintisiz ve güvenli bir şekilde seyretmesine olanak tanır. Ayrıca algoritmanın sunduğu esneklik, özellikle engellerin bulunduğu dinamik ortamlarda yörünge planlaması yaparken önemli avantajlar sunmaktadır. Yörünge planlama sırasında, aracın engellerden kaçınması kritik bir gerekliliktir. B-Spline eğrileri, engellerin etrafından geçmek için uygun yollar tasarlamada kullanılabilir. Ancak, bu tür bir yörünge planlama yalnızca statik engeller için değil, dinamik engeller (diğer araçlar veya hareketli nesneler) için de kullanılmalıdır. Dinamik yörünge planlaması, aracın çevresindeki değişen engelleri göz önünde bulundurarak sürekli bir yörüngenin güncellenmesini gerektirir. Bu süreç, araçların çevreye uyum sağlamasını ve güvenli bir şekilde hedeflerine ulaşmasını mümkün kılar. Derin öğrenme teknikleri, bu tür dinamik engellerin tanımlanması ve en iyi rotanın seçimini hızlandırılması için kullanılabilir (Li ve Yu, 2019).

Literatür incelendiğinde, USV için yörünge planlama çalışmaları için farklı algoritmaların kullanımı görülmektedir. Bu çalışmalar genellikle çarpışma önleme, hareketli engellerden kaçınma ve çevresel belirsizliklere uyum sağlama hedefleri doğrultusunda şekillenmiştir. Pekiştirmeli Öğrenme Algoritmaları bu alandaki uygulamaları giderek artmakta olup, çarpışmasız rota üretimi ve dinamik ortama adaptasyon konularında etkin çözümler sundukları görülmektedir. Çalışma (Zhu vd., 2018), simülasyon ortamlarında geliştirdikleri RL algoritması ile çarpışmadan kaçınma başarımını test etmiş ve bu yöntemlerin güvenli yol planlamasında etkili olduğunu göstermiştir. Yörünge planlamayı bir aracın başlangıç noktasından hedefe gitmek için izleyeceği yolun planlaması ve bu yol üzerinde engellerden kaçınma amacıyla kullanmışlardır. Literatürde önerilen diğer bir çalışmada engellerden kaçınmak ve güvenli rotayı oluşturmak için çeşitli modeller

geliştirilmiştir (Xia vd., 2020). Bu çalışmada, Gauss Karışım Modeli (Gaussian Mixture Model) ve Gauss Süreci Regresyonu (Gaussian Process Regresion) gibi yöntemlerle, hareketli engellerin yol üzerindeki etkileri tahmin edilmiştir ve belirsizlikler dikkate alınarak güvenli rotalar tasarlanmıştır. Diğer bir çalışmada çoklu engellerle başa çıkmak ve güvenli yörünge planlamak amacıyla geliştirdikleri dinamik yörünge planlama algoritması ile önemli bir katkı sağlamıştır (Zhang vd., 2023). Bu çalışmada, yörünge değişikliklerini daha etkin biçimde düzenlemek amacıyla B-Spline tabanlı bir model kullanılmış; ayrıca, hareketli engellerin bulunduğu bir ortamda yaklaşan tehlikeleri öngörebilmek için belirsizlik yönetimine dayalı bir çözüm geliştirilmişlerdir.

2.3. Pekiştirmeli Öğrenme

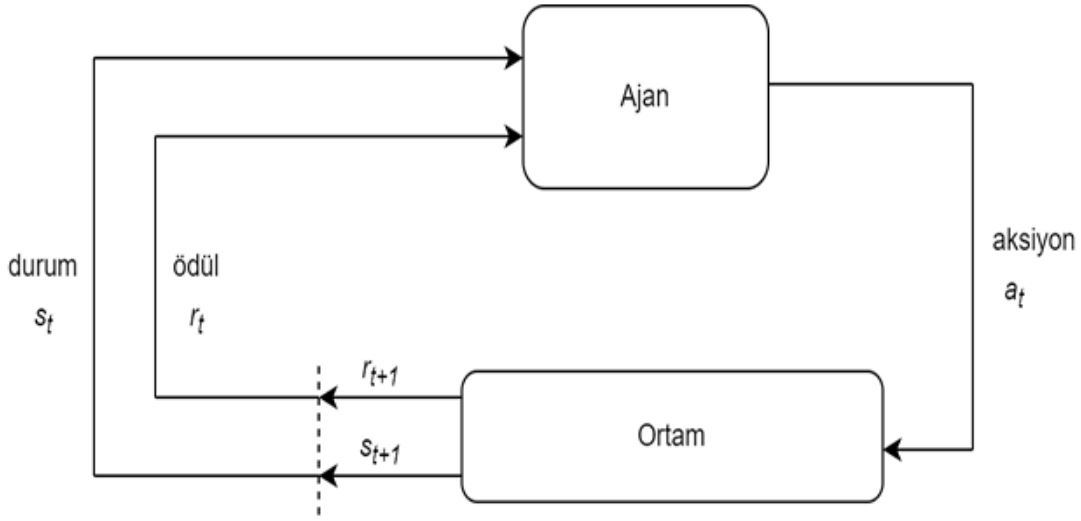
USV ile ilgili çarpışma önleme ve yol planlama algoritmalarına yönelik son yıllarda yapılan araştırmalar, özellikle Pekiştirmeli Öğrenme (Reinforcement Learning-RL) algoritmasının kullanımıyla önemli bir ilerleme göstermiştir. Bu çalışmalar, otonom navigasyon ve güvenli yol planlaması için yeni algoritmalar geliştirilmesine olanak sağlamaktadır.

Pekiştirmeli Öğrenme makine öğrenmesinin önemli bir dalıdır ve özellikle karar verme süreçlerinde, ajanın çevresiyle etkileşimi yoluyla öğrenmesini sağlar. RL'nin temel amacı, ajanın bir dizi eylemde bulunarak en yüksek uzun vadeli ödülü elde edebileceği bir politika geliştirmesidir. Bu politika, ajanın belirli bir durumda hangi hareketi yapması gerektiğine dair bir harita sunar. RL, özellikle karmaşık ve dinamik ardışık karar verme problemleri söz konusu olduğunda etkili bir yöntem olarak tercih edilmektedir.

Pekiştirmeli öğrenme, çeşitli endüstrilerdeki uygulamalarda giderek daha fazla tercih edilmektedir. Ajanın her durum için hangi kararı alacağı önceden belirlenmez, bunun yerine ajan çevreyle etkileşimde bulunarak ve eylemleri deneyerek hangi kararların daha yüksek ödüller getirdiğini öğrenir. Bu öğrenme süreci ajanın çevresiyle sürekli bir etkileşim içinde olması gerektiğinden, deneyimsel öğrenme olarak da adlandırılabilir.

Özellikle ortamın tam olarak modellenemediği veya yüksek karmaşıklığa sahip olduğu durumlarda etkili sonuçlar üretmektedir. Bu yönüyle pekiştirmeli öğrenme, denetimli

ve denetimsiz öğrenmeden farklı olarak, doğrudan örnek verilerden değil, ortamla etkileşimden öğrenilen deneyimlere dayanır. Şekil 2.6'da ajan ve ortam arasındaki bu etkileşim süreci şematik olarak gösterilmektedir (Sutton ve Barto, 2018).



Şekil 2.6. RL modeli

Şekil 2.6'da r_t ajanın ortamdan almış olduğu ödülü, a_t ajanın ortama gönderdiği eylemleri ifade etmektedir. Ajanın ortam içerisinde bulunduğu durumlar ise s_t ile ifade edilmektedir. Ajan herhangi bir t anında, s_t durumunda bulunurken a_t eylemini gerçekleştirerek r_{t+1} ödülünü almakta ve buna bağlı olarak s_{t+1} durumuna geçmektedir.

RL'nin temel hedefi, eylemlerin uzun vadeli sonuçlarını göz önünde bulundurarak beklenen kümülatif ödülü maksimize edecek bir politika öğrenmektir (Denklem 2.1). Denklem 2.1'deki γ indirim faktörünü R_{t+k+1} ise gelecekteki anlık ödülleri temsil eder. İndirim faktörü gelecekteki ödüllerin bugünkü ödül üzerindeki ağırlığını belirleyen bir parametredir.

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{\{t+k+1\}} \qquad (2.1)$$

Pekiştirmeli öğrenmenin dinamik çevre koşullarını dikkate alabilme yeteneği, elektrikli araçların rota optimizasyonu gibi uygulamalarda da etkin bir şekilde kullanılmasına olanak sağlamaktadır (Aslan Yıldız vd., 2024; Aslan Yıldız vd., 2025).

Pekiştirmeli Öğrenme, birkaç ana bileşenden oluşur ve bu bileşenler arasında etkileşim sağlanarak öğrenme süreci gerçekleşir. Bu bileşenler aşağıda sıralanmıştır:

- **Ajan (Agent):** Çevresiyle etkileşimde bulunan ve ödüller aracılığıyla geri bildirim alarak öğrenen varlıktır. Ajan, çevresindeki durumu gözlemler, eylemler seçer ve bu eylemlerle ilgili ödüllerle davranışını günceller.
- **Ortam (Environment):** Ajanın etkileşimde bulunduğu dış sistemdir. Çevre, ajana mevcut durumunu iletir ve ajanın gerçekleştirdiği eylemlere göre tepki verir. Çevre, aynı zamanda ajanın ödülünü belirler.
- **Durum (State):** Ajanın çevresindeki durumu ifade eder. Durum, ajanın çevresini algıladığı şekilde bir temsilidir ve ajanın bir sonraki eylemini belirlemek için kritik bir öneme sahiptir.
- **Eylem (Action):** Ajanın belirli bir durumda alacağı karardır. Eylemler, ajanın çevresindeki durumu değiştirmek ve ödülleri maksimize etmek amacıyla yaptığı hareketlerdir.
- **Ödül (Reward):** Çevreden alınan geri bildirim sinyalidir. Ajan, bir eylem gerçekleştirdikten sonra ödül alır ve bu ödül, eylemin kalitesini gösterir. Ajan, bu ödülleri dikkate alarak gelecekteki davranışlarını şekillendirir.

Pekiştirmeli Öğrenme Algoritmaları genellikle yenilemeli bir süreç şeklinde işler. Ajan, çevreyi keşfeder, ödüller toplar ve gözlemlerine dayanarak politikasını günceller. Bu öğrenme süreci, ajan için daha fazla bilgi toplandıkça, daha etkili ve optimal kararlar alınmasına olanak tanır.

Pekiştirmeli öğrenmenin avantajları, özellikle çevreyle etkileşim yoluyla öğrenme ve belirsizlikle başa çıkma konusundaki esnekliğinden kaynaklanmaktadır. Deneyimlerden

öğrenme, RL'nin değişen ortamlara hızla uyum sağlama yeteneğini artırır ve ajanın uzun vadeli hedefler doğrultusunda öğrenmesini sağlar. Bu özellik, pekiştirmeli öğrenmeyi dinamik ve belirsiz ortamlarda ideal bir çözüm haline getirir. Robotik, otonom sistemler oyun teorisi ve öneri sistemleri gibi çok çeşitli uygulama alanlarında RL başarıyla kullanılmaktadır (Kaiser vd., 2019).

RL'nin üstün yanları yanında zayıf yönleri de vardır. İlk olarak, RL algoritmaları genellikle büyük miktarda veri ve keşif süresi gerektirdiğinden, uygulama süreci pahalı ve zaman alıcı olabilir. Özellikle yüksek boyutlu çevrelerde, eğitim süreci hesaplama açısından oldukça pahalı hale gelebilir. Ayrıca, pekiştirmeli öğrenme algoritmalarının başarısı genellikle hiper parametre ayarlamalarına bağlıdır ve bu ayarların doğru şekilde yapılması zorlu olabilir. Bu durum algoritmaların performansını etkileyen önemli bir faktördür.

Pekiştirmeli öğrenme, aynı zamanda ajanın keşif ve sömürü arasındaki dengeyi yönetmesini gerektirir. Bu dengeyi doğru kurmak, ajanın öğrenme sürecini etkili kılmak için kritik öneme sahiptir. Keşif, ajanın yeni stratejiler öğrenme çabasıdır, ancak keşif aşaması verimsiz olabilir. Diğer taraftan, sömürü aşaması ajanın mevcut bilgisini kullanarak ödülleri maksimize etmeyi hedefler, ancak bu da yenilikçi stratejilerin öğrenilmesini engelleyebilir. Bu iki süreç arasında sağlanacak doğru denge, RL algoritmalarının etkinliğini belirleyen önemli bir unsurdur. Keşif stratejileri, pekiştirmeli öğrenmede vurgulanması gereken bir diğer önemli konu olup, aynı zamanda en zorlu alanlardan biridir. Keşif, ajanın yeni kararlar deneyerek daha iyi politikalar arayışıdır. Literatürde bu keşif amacıyla Epsilon-Greedy (Epsilon-Greedy), SoftMax Keşfi (SoftMax Exploration), Üst Güven Aralığı (UCB_Upper Confidence Bound), Thompson Örneklemesi (Thompson Sampling) ve Bayesyen Optimizasyonu (Bayesian Optimization) gibi farklı stratejiler kullanılmaktadır. Diğer makine öğrenme yöntemlerinde olduğu gibi, iyi çözümler etrafında arama yapma ve yeni keşifler yapma arasında bir denge kurulması gerekmektedir. Pekiştirmeli öğrenme algoritmalarında da benzer dengelemeler mevcuttur.

Pekiştirmeli öğrenme, Markov Karar Süreçleri (Markov Decision Process-MDP) ile doğrudan ilişkili bir alandır. MDP'leri çözmek için keskin ve tam bir çözüm sağlarken, çevreye ait geçiş olasılıklarını kullanır. Ancak bu geçiş olasılıkları çoğu zaman bilinmediği için, pekiştirmeli öğrenme, çevrenin modelinin mevcut olmadığı durumlarda MDP'leri

çözmenin bir yolu olarak ortaya çıkar. RL, özellikle çevre ile doğrudan etkileşim yoluyla öğrenme gerçekleştiren modelden bağımsız (model-free) algoritmalar sunarak, bilinmeyen çevre dinamikleri ile başa çıkılmasını mümkün kılar. Bu bağlamda, Q-öğrenme (Q-learning) ve Politika Gradyanı (Policy Gradient) gibi RL algoritmaları, model gerektirmeyen yaklaşımlar sunar.

Pekiştirmeli öğrenme, geleneksel denetimli veya denetimsiz öğrenme yöntemlerinin yetersiz kaldığı ardışık karar verme problemleri üzerinde etkili bir şekilde kullanılabilmektedir. Bu yöntemler, özellikle çevre ile etkileşimin sürekli olduğu ve kararların birbiriyle ilişkili olduğu durumlarda faydalıdır. RL algoritmaları, temel prensiplerine ve öğrenme yaklaşımlarına göre farklı kategorilerde sınıflandırılabilir. Çoğunlukla model tabanlı (model-based) ve modelden bağımsız (model-free) yöntemler olarak iki ana gruba ayrılırlar.

Model tabanlı RL algoritmaları, çevre dinamiklerinin bir modelini öğrenmeye yönelik çalışır ve bu algoritmalar genellikle çevre dinamiklerinin bilindiği veya tahmin edilebildiği durumlarda kullanılır. Diğer taraftan, modelden bağımsız RL algoritmaları, doğrudan deneyim yoluyla bir politika veya değer fonksiyonu öğrenirler ve genellikle çevre dinamiklerinin belirsiz olduğu veya çok karmaşık olduğu durumlarda tercih edilir. Bu algoritmalar, doğrudan çevreyle etkileşim yoluyla deneyimsel öğrenme gerçekleştirirler.

Modelden bağımsız RL algoritmaları, genel olarak iki ana kategoriye ayrılmaktadır: değer tabanlı ve politika tabanlı algoritmalar. Değer tabanlı algoritmalar, belirli bir durum-eylem çiftine ilişkin değer fonksiyonu öğrenirken, politika tabanlı algoritmalar doğrudan eylem seçim politikasını öğrenirler. Değer tabanlı algoritmalar arasında yer alan Q-öğrenme bu tür algoritmaların en yaygın örneklerinden biridir (Watkins ve Dayan, 1992). Bir ajanın optimal durum-eylem değer fonksiyonunu (Q fonksiyonu) öğrenmesini amaçlar. Bu algoritma, verilen bir durum için en uygun eylemi, belirli bir politika çerçevesinde ya da rastgele seçerek, bu durum-eylem çiftinin beklenen ödül değerine göre belirler.

Bir diğer önemli değer tabanlı algoritma ise, Derin Q Ağı (Deep Q Network-DQN)'dir. DQN, Q öğrenme ile derin sinir ağlarını birleştirerek, yüksek boyutlu durum alanlarında daha etkili çözümler üretir. Bu tür algoritmalar, özellikle yüksek-dönüşümlü

çevrelerde önemli bir avantaj sağlar. Ayrıca, Çift Derin Q Ağı (Double Deep Q-Network-DDQN) Algoritmasındaki aşırı tahmin problemine çözüm getiren bir modifikasyon sunar. DDQN, değer tahminlerini iyileştirerek daha stabil öğrenme süreçleri sağlar (Van Hasselt vd., 2015).

Değer tabanlı RL algoritmaları, aynı zamanda on-policy ve off-policy olmak üzere iki alt gruba ayrılabilir. On-policy yöntemler, ajanın çevreyle doğrudan etkileşimde bulunarak, mevcut deneyimlerden öğrenmesini sağlar. Buna karşılık, off-policy yöntemler, geçmiş deneyimlerden faydalanarak politikanın güncellenmesini sağlar. Bu tür bir yaklaşım, deneyim tekrar mekanizması gibi yöntemlerle örneklenebilir. Özellikle DQN Algoritmasında bu yöntem, ajanın daha etkili bir şekilde öğrenmesine olanak tanır (Sutton ve Barto, 2018).

Pekiştirmeli öğrenme ajanlarının eğitimi genellikle simüle edilmiş ortamlarda yapılır. Bu yaklaşım daha hızlı ve daha güvenlidir, ancak ajanların gerçek dünyaya transferi, ince ayar ve güvenlik önlemleri gerektirir.

Literatür incelendiğinde; pekiştirmeli öğrenme yöntemlerini kullanarak USV üzerinde yol planlama algoritmaları geliştirilmiştir. DQN Algoritması kullanılarak, USV'nin navigasyonunun güvenli ve etkin bir şekilde gerçekleştirilmesi sağlanmıştır (Zhou vd., 2019). Genetik algoritmalar (Genetic Algorithms-GA) pekiştirmeli öğrenme kombinasyonunu kullanarak USV'lerin yol planlaması ve çarpışma önleme problemini çözmeye çalışmışlardır (Li ve Yang 2020). Genetik algoritmalar, rotaların başlangıç noktasından hedefe daha verimli bir şekilde oluşturulmasına yardımcı olurken, pekiştirmeli öğrenme yöntemleri ise gerçek zamanlı karar verme süreçlerinde etkin çözümler sunmaktadır. Bu tür hibrit sistemler, dinamik deniz koşullarında etkili çözümler sunabilmektedir. Pekiştirmeli öğrenme ile karmaşık deniz ortamlarında çarpışmasız navigasyon önerisi geliştirmiştir. Çalışmalarını simüle edilmiş bir platformda doğrulayan yazarlar, başlangıç noktasından hedef noktasına ulaşırken dinamik engellerden kaçınmayı ve güvenli rotaların belirlenmesini hedeflemişlerdir. Bu tür navigasyon algoritmaları hem statik hem de dinamik engelleri dikkate alarak gelişmiş güvenlik sağlamakta kullanılır (Yan vd., 2021). USV'ler için çok ajanlı pekiştirmeli öğrenme (multi-agent reinforcement learning) tekniklerini kullanarak, grup halindeki USV'lerin koordinasyonu üzerinde

çalışmalar yapmışlardır. Çalışmada, birden fazla USV'nin aynı anda güvenli bir şekilde hareket etmesini sağlayan bir algoritma geliştirilmiştir. Bu tür uygulamalar, özellikle keşif ve güvenlik görevlerinde çok sayıda aracın koordineli bir şekilde çalışmasını sağlar (Roh vd., 2021). DQN Algoritmasının yavaş ve kararsız yakınsama problemini ele almış ve bu problemi aşmak için deneyim tekrarı tekniği kullanmıştır. Sinir ağının eğitim hızını hızlandırmak için algoritma tasarımında çeşitli iyileştirmeler yapmış ve DQN Algoritmasının hiper parametrelerinin model performansı üzerindeki etkilerini incelemişlerdir. Bu tür geliştirmeler, otonom sistemlerin eğitim süreçlerini hızlandırarak daha verimli çözümler elde edilmesini sağlamaktadır (Zhu vd., 2021). Gemi etki alanına ve en yakın yaklaşma noktasına dayalı bir çarpışma riski hesaplama yöntemi önermiştir. Bu yöntem, COLREG'leri göz önünde bulundurarak güvenli navigasyon stratejileri geliştirmeyi hedeflemiştir. Çarpışma riski hesaplaması, USV'nin diğer gemilerle olası çarpışma riskini tahmin etmek için kullanılır (Li ve Wu 2021). COLREG'lere uygun bir çarpışma önleme algoritması geliştirmiştir. Çalışmada, yapay potansiyel alan (Artificial Potential Field-APF) algoritması kullanılarak, DQN Algoritmasının eylem alanı ve ödül fonksiyonu geliştirilmiştir. Bu sayede, USV'nin manevra kabiliyetleri ve COLREG'lerin gereksinimleri arasında denge sağlanmıştır (Xu vd., 2022). COLREG'leri dikkate alarak dinamik bir çarpışma önleme algoritması tasarlamışlardır. Çalışmalarında bir risk değerlendirme modeli geliştirilmiş ve ödül fonksiyonları bu modelin gereksinimlerine göre şekillendirilmiştir. Bu model, COLREG'lere uyumu sağlamanın yanı sıra, gerçek zamanlı çarpışma önleme stratejileri için de önemli bir yaklaşım sunmaktadır (Fan vd., 2022). USV'lerin manevra kabiliyetini dikkate alarak, araçların özelliklerine yönelik kategori bazlı keşif yöntemi ile çarpışma önleme algoritmaları geliştirmiştir. Bu yaklaşım, USV'nin farklı kategorilerdeki manevra kabiliyetlerine göre uygun ödül stratejilerinin belirlenmesine olanak tanır (Wen vd., 2023). DQN Algoritmasından kaynaklanan yakınsama problemini çözmek için sinir ağının hızlandırılması gerektiğini savunmuşlardır. Bu tür araştırmalar, özellikle karmaşık deniz koşullarında USV'lerin etkinliği ve güvenliği konusunda önemli katkılar sağlamaktadır (Chun ve Roh 2021). Pekiştirmeli öğrenme ve COLREG'leri kullanarak bir iki aşamalı öğrenme uygulamışlardır. İlk aşamada USV'nin kaçınma yörüngeleri kaydedilmiş, ikinci aşamada ise bu yörüngeler üzerinden yeniden bir eğitim gerçekleştirilmiştir. Bu iki aşamalı öğrenme hem daha hızlı öğrenme süreçleri hem de daha güvenli navigasyon stratejileri geliştirilmesine olanak tanımaktadır (Wen vd., 2023). Derin Pekiştirmeli Öğrenme (DRL) tabanlı bir rota takibi ve çarpışma önleme sistemi çalışmaları mevcuttur (Larsen vd., 2021).

Son yıllarda, çoklu USV formasyon çarpışma önleme için DRL ve COLREG'leri birleştirmiştir (Wang vd., 2022). COLREG'lere uygun bir pekiştirmeli öğrenme yaklaşımıyla USV manevra ve çarpışma önleme problemlerini ele almıştır (Sonntag vd., 2025). Ancak bu çalışmaların hiçbiri, USV'nin manevra kabiliyetini esnek biçimde modellemek için pekiştirme öğrenme temelli yörünge planlama yaklaşımı sunmamaktadır.

Son yıllarda karmaşık ortamlarda hareket edebilmek ve dinamik çevreler ile hareketli engellerle başa çıkabilmek için daha gelişmiş çözüm yöntemlerine olan ilgi artmıştır. Bu doğrultuda, özellikle İnsansız Su Üstü Araçları için yol planlaması ve çarpışma önleme alanındaki araştırmalar büyük bir hızla ilerlemiştir. İlk başlarda, sabit engellerden kaçınma temalı çalışmalar ön planda iken, son yıllarda karmaşık deniz ortamlarında hareketli ve dinamik engellerin dikkate alınması gerekliliği üzerinde durulmuştur. Denizcilik ortamındaki rüzgârlar, dalgalar gibi dış etkenlerin yanı sıra, diğer gemiler ve deniz trafiği gibi dinamik engellerin USV'lerin navigasyon stratejileri üzerinde önemli etkiler yarattığı anlaşılmıştır. Bu nedenle, dinamik engellerin tanımlanması ve bunlardan etkili bir şekilde kaçınılması son derece kritik hale gelmiştir (Xu vd., 2022). Çarpışma önleme ve engellerden kaçınma stratejileri, USV'lerin otonom navigasyonunda temel bir gereklilik olarak öne çıkmaktadır. Pekiştirmeli öğrenme ve Markov Karar Süreci gibi algoritmalar, dinamik ve statik engellerden kaçınmayı öğrenebilmek için son yıllarda yaygın şekilde kullanılmıştır (Wen vd., 2023). İlk çalışmalarda, engeller genellikle sabit ve statik olarak modellemişken, son yıllarda deniz ortamındaki dinamik engeller, örneğin diğer gemiler, dalgalanma ve rüzgâr koşulları gibi etmenlerle birlikte ele alınmaya başlanmıştır.

2.3.1. Derin pekiştirmeli öğrenme

Derin Pekiştirmeli Öğrenme (Deep Reinforcement Learning_DRL), Pekiştirmeli Öğrenme yaklaşımının bir alt disiplindir. Bu yaklaşımın temel amacı, yüksek boyutlu ve karmaşık problem alanlarında etkili çözümler üretmektir. DRL, özellikle robotik kontrol, video oyunları, doğal dil işleme ve otonom sürüş gibi uygulamalarda dikkate değer başarılar göstermiştir. Bu başarılardan ötürü, son yıllarda geniş çapta ilgi görmüş ve popülerlik kazanmıştır.

Derin Sinir Ağları (Deep Neural Networks-DNN), RL ve DRL arasındaki farkı belirleyen temel faktördür. Derin Q-Ağları (DQN) veya Derin Politika Ağları, değer fonksiyonlarını veya politikaları tahmin etmek için fonksiyon yakınlaştırıcıları olarak kullanılır. Bu özellik, DRL'yi gerçek dünyadaki birçok probleme uygulamak için uygun hale getirir, çünkü sinir ağları yüksek boyutlu ve sürekli durum alanlarını işleyebilir. Ancak, RL algoritmalarının yapay sinir ağları ile birleştirilmesi bazı sorunları da beraberinde getirmiştir, bunlar arasında sıralı verilerin korelasyonu, sapma, kararsızlık ve aşırı tahmin gibi problemler bulunmaktadır.

USV'lerde rotalama ve çarpışma önleme problemlerine yönelik olarak DRL yöntemlerinin etkin biçimde kullanıldığı görülmektedir. Literatüre baktığımızda DQN modeli ile USV'lerin güvenli rotalama yapmasını sağlamış ve çarpışma risklerini önemli ölçüde azaltmıştır (Zhou vd., 2019). Genetik Algoritmalarla DQN'i birleştirerek dinamik çevrelerde adaptif rotalama çözümleri geliştirmiştir (Li ve Yang, 2020). Çoklu USV'lerin iş birliği içinde hareket etmesini sağlayan çok ajanlı DRL modelleri geliştirmiştir (Roh vd, 2021). Deneyim tekrarı ve ağ yapısındaki düzenlemeler, DQN'in yakınsama sürecini hızlandırmada etkili olmuştur (Zhu vd., 2021). DRL tabanlı sistemleri COLREGs denizcilik kurallarına entegre ederek daha yasal ve güvenli navigasyon sağlamışlardır (Xu vd., 2022).

2.3.2. Derin Q ağı

Derin Q Ağı (DQN), yüksek boyutlu durum uzaylarında temsil öğrenme yoluyla bilgi çıkarımı yaparak, değer fonksiyonunu Derin Sinir Ağlarıyla yaklaşık olarak öğrenmeyi amaçlayan bir yaklaşımdır. DQN, Q-fonksiyonunu (Denklem 2.2) kullanır ve tüm durum-eylem çiftleri için Q değerlerini tablo şeklinde saklamak yerine, bir sinir ağı aracılığıyla Q değerlerini tahmin eder. Bu yaklaşımda, durum-eylem çiftleriyle ilgili gerekli bilgiler sinir ağının ağırlıklarında depolanır.

$$Q\ (s,a) \leftarrow R(s,a) + \gamma . max_{a'} Q(s',a') \qquad (2.2)$$

Denklem 2.2'de verilen fonksiyon, bir (s,a) çifti için Q değerinin hesaplanmasında kullanılır. Q değeri, sistemin s durumunda a eylemi seçilmesinin başarısını ölçen bir değerdir. Q değeri belirli bir eylemi bir durumda seçmenin ve sonrasında belirli bir politikaya

uymanın beklenen birikimli gelecekteki ödüllerini temsil eden bir fonksiyondur. Burada $R\ (s,a),$ anlık ödülü, γ gelecekteki ödüllerin önemini belirten indirim faktörünü ve s' ise s durumunda a eylemiyle ulaşılan bir sonraki durumu ifade eder $max_{a'}Q(s',a')$ ise bir sonraki s' durumunda mümkün olan tüm eylemler için Q değerleri arasında en yüksek olanı temsil eder (Watkins vd., 1992).

DQN'de, ajanın amacı zaman içinde beklenen birikimli ödülü maksimize edecek bir politika öğrenmektir ve Q değeri, karar süreçlerinde temel rol oynar. Q değeri fonksiyonu, ajanın öğrenme sürecinde farklı durumlarda hangi eylemlerin seçileceğine karar vermesinde yardımcı olur. Ajan, çevreyi keşfettikçe ve ödüller şeklinde geri bildirim aldıkça, Q-değerlerini güncelleyerek zamanla karar verme süreçlerini iyileştirir.

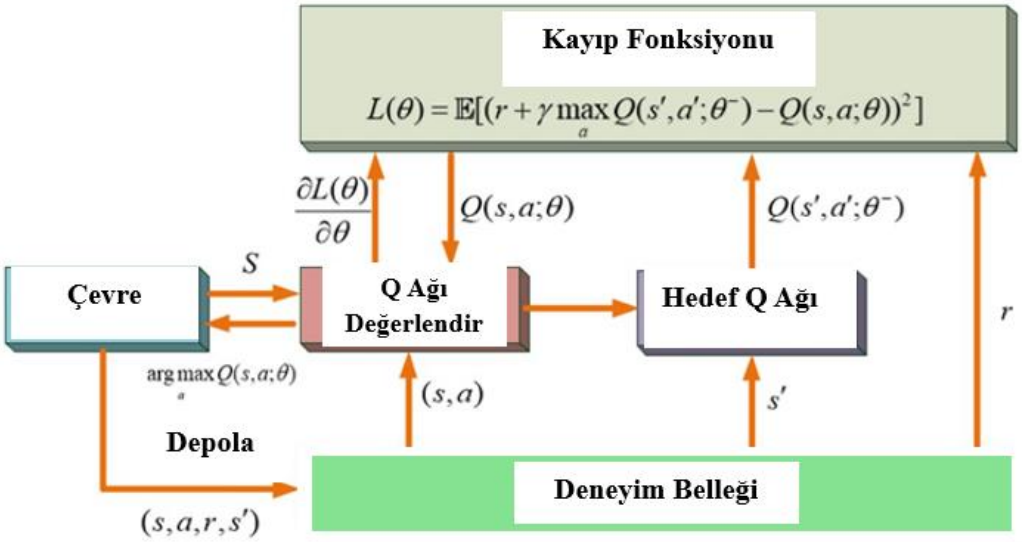
DQN modelinde, tahminlerin kararsızlığını azaltmak için iki ayrı Q ağı kullanılır. Tahmin Q ağı (prediction network) ve hedef Q ağı (target network). Hedef Q ağı, güncellenme sıklığı düşük olan sabit parametrelere sahiptir ve bu sayede tahmin edilen Q değerlerinin daha kararlı olmasını sağlar (Mao vd., 2018). Bu yapıdaki güncel Q fonksiyonu (Denklem 2.3) ile ifade edilirken, modelin öğrenme süreci kayıp fonksiyonu (Denklem 2.4) ile tanımlanır. Burada, parametreler θ, tahmin ve hedef Q ağları arasındaki değer farkını minimize edecek şekilde güncellenir.

$$Q\ (s,a;\theta) = R(s,a) + \gamma\ .max_{a'}Q(s,a';\theta') \qquad (2.3)$$

$$L(\theta) = \mathbb{E}_{(s,a,r,s')}[(R(s,a) + \gamma\ .max_{a'}Q(s',a';\theta'') - Q\ (s,a;\theta))^2] \qquad (2.4)$$

DQN modelinde $\mathbb{E}$, öğrenmenin temelini oluşturan ortalama kare hata kaybının bir beklenen değeri olarak değerlendirilir. Modelin dört temel adımı vardır ve her bir adım, bir iterasyon boyunca uygulanır. İlk olarak, sistem durumu başlatılır, ardından epsilon-greedy yaklaşımı kullanılarak rastgele ya da modelin tahmin ettiği bir eylem seçilir. Bu eyleme karşılık gelen bir adım ödülü $R\ (s,a)$, ve bir sonraki durum s' belirlenir. Sonrasında, (s,a,r,s') bir bellek yapısında, deneyim tekrar mekanizması aracılığıyla depolanır. Her iterasyonun sonunda, tahmin ağı, tekrar hafızasından rastgele seçilen örnekler ile bir set

boyutuna kadar eğitilir ve belirli aralıklarla θ güncellenir. Şekil 2.7'de DQN modelinin temel bileşenleri ve bilgi akış yapısı görülmektedir.



Şekil 2.7. DQN modelinin temel bileşenleri ve bilgi akış yapısı

Şekil 2.7'de bir ajan ile çevre arasındaki etkileşim süreci ve öğrenmenin nasıl gerçekleştiği görselleştirilmiştir (Fan vd., 2022). Çevre, mevcut durumu ajana sunar. Ajan, değerlendirme ağı kullanarak bu duruma uygun bir eylem üretir. Bu eylem, çevrede gerçekleştirilir ve sonucunda bir ödül ve yeni bir durum elde edilir. Bu deneyim, tekrarlama belleği (replay buffer) adı verilen hafızaya kaydedilir. Bu bellekte biriken örnekler, ağın eğitimi için rastgele seçilerek kullanılır. Değerlendirme ağı, Q değerlerini tahmin ederken, hedef ağ ise daha sabit tahminlerle öğrenmeyi istikrarlı hale getirir. Kayıp fonksiyonu (Loss Function), hedef Q değeri ile tahmin Q değeri arasındaki farkı minimize etmeye çalışır.

Son yıllarda, bu algoritma denizcilik ve otonom sistemlerde rotalama için de başarıyla uygulanmıştır. Literatürü incelediğimizde DQN modelini kullanarak USV'lerin güvenli bir şekilde çarpışmalardan kaçınmasını sağlanmıştır (Zhou vd., 2019). DQN'in yakınsama süresini iyileştirmiş ve daha hızlı karar alma süreçleri geliştirilmiştir (Zhu vd., 2021). Engelli ortamlarda USV'lerin yönlendirilmesi için bir kontrol sistemi geliştirilmiştir (Xu vd., 2018). DQN tabanlı bir model ile gemilerin yön kontrolü ve rota takibi için bir sistem önerilmiştir (Sivaraman vd., 2022).

3. MATERYAL VE YÖNTEM

Bu bölümde USV için geliştirilen pekiştirmeli öğrenme tabanlı yörünge planlama sistemi sunulmaktadır. Öncelikle, çalışma kapsamında materyal verilmekte olup daha sonra ele alınan problemin tanımı yapılmakta ve denizcilik alanında otonom sistemlerin karşılaştığı temel zorluklar açıklanmaktadır. Yörünge planlama sürecinde dikkate alınan içsel (kinematik) ve dışsal (çevresel) parametreler detaylandırılmıştır. Kullanılan Pekiştirmeli Öğrenme Algoritması, bu algoritmanın giriş parametreleri, hiper parametre ayarlarının seçimi ve çok bileşenli ödül fonksiyonu sunulmuş; öğrenme sürecinde kullanılan sistem mimarisi, yazılım altyapısı ve deneysel test ortamı tanıtılmıştır. Son olarak, önerilen DQN modelinin sistemsel akışı, kullanıcı arayüzü ve gerçek zamanlı karar verme yapısı diyagramlarla desteklenerek açıklanmıştır. Bu yapı sayesinde, USV'nin dinamik ve belirsizlik içeren deniz koşullarında güvenli, kurallara uygun ve çevresel faktörlere duyarlı şekilde hareket edebilmesi hedeflenmiştir.

3.1. Materyal

Bu çalışmada, fiziksel deney ortamı yerine Python tabanlı bir simülasyon altyapısı kullanılarak, USV sisteminin çevresel koşullar altındaki hareket kabiliyeti modellenmiştir. Çalışmada kullanılan materyaller, simülasyon ortamının kurulumu, algoritmaların uygulanması ve sonuçların analizine yönelik yazılım araçlarından oluşmaktadır.

Python 3.9: Tüm simülasyonlar, algoritma entegrasyonları ve veri analizleri bu sürüm kullanılarak gerçekleştirilmiştir.

Pygame: USV hareketlerinin görsel olarak izlenebilmesi ve simülasyon ortamının modellenmesi amacıyla tercih edilmiştir.

NumPy & Matplotlib: Kayıp fonksiyonu (Loss Function) ve Q-değerlerinin analiz edilmesi ve grafiksel olarak sunulması bu kütüphaneler aracılığıyla yapılmıştır.

Geomdl (NURBS-Python): Çoklu yörünge üretimi için B-Spline algoritmalarının uygulanmasında kullanılmıştır.

CSV Log Dosyaları: Eğitim sürecinde modelden elde edilen Q değerleri, kayıp değerleri ve seçilen aksiyonlar düzenli olarak CSV dosyalarına kaydedilerek, sonradan analiz ve doğrulama için kullanılmıştır.

Ortam Verisi: Gerçek zamanlı çevresel veriler yerine, Copernicus Deniz ve Hava Ortam Servisi'nden alınan tipik değerler kullanılarak rüzgâr hızı, akıntı hızı, dalga yüksekliği gibi etkenler simülasyona dâhil edilmiştir.

Tüm bu bileşenler, önerilen pekiştirmeli öğrenme tabanlı yörünge planlama ve çarpışma önleme sisteminin çalışabilirliğini test etmek ve farklı senaryolarda sistemin performansını değerlendirmek amacıyla entegre edilmiştir.

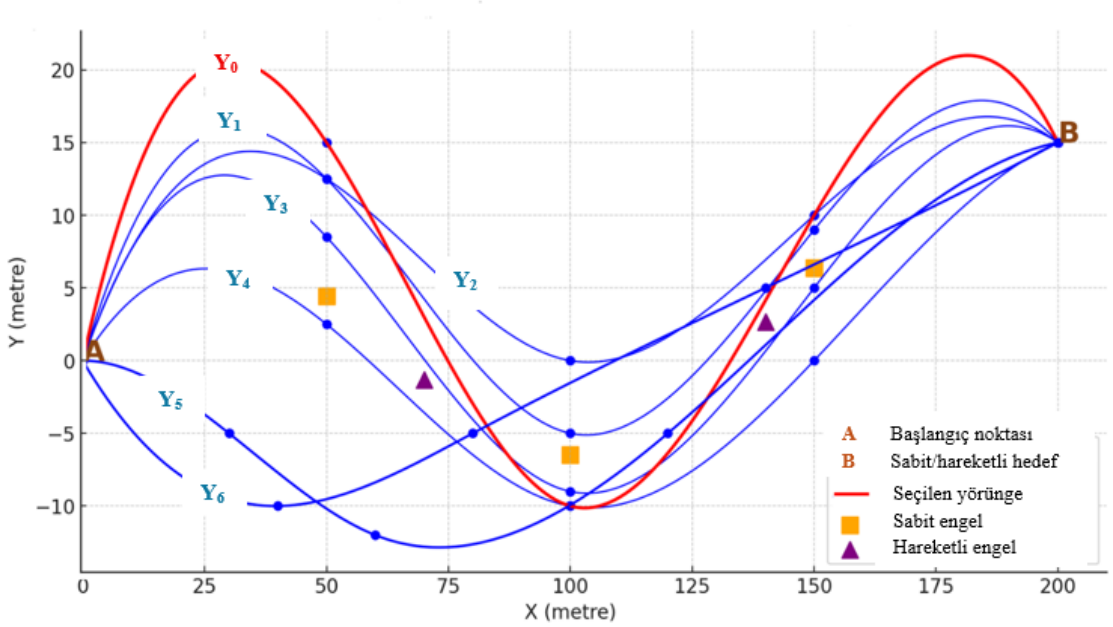
3.2. Problemin Tanımlanması ve Örnek Çözüm Yaklaşımı

Son yıllarda, otonom sistemlerin denizcilik alanındaki uygulamaları hızla artmış ve İnsansız Su Üstü Araçları hem sivil hem de askeri görevlerde önemli roller üstlenmeye başlamıştır. USV'ler; keşif-gözetleme, çevresel izleme, denizaltı tespiti, deniz haritalama, liman güvenliği gibi görevlerde etkin bir şekilde kullanılmaktadır. Bu görevlerin başarıyla yerine getirilebilmesi için, USV'lerin güvenli, verimli ve çevresel faktörlere uyumlu şekilde navigasyon gerçekleştirmesi önemlidir. Ancak açık denizlerde veya karmaşık deniz trafiğinin bulunduğu alanlarda USV'lerin güvenli şekilde hareket etmesi, birçok zorluğu da beraberinde getirmektedir. Bu zorlukların başında, çarpışma riski, çevresel belirsizlikler (rüzgâr hızı, dalga, akıntı vb.), dinamik engellerin varlığı ve COLREG kurallarına uyum zorunluluğu gelmektedir. Bu nedenle, otonom USV'lerin yalnızca engellerden kaçınmakla kalmayıp, aynı zamanda COLREG'lere uygun şekilde manevra yapabilmesi ve diğer deniz araçlarıyla güvenli mesafeyi koruyarak hareket etmesi gerekmektedir. Bu çalışmada amaç, bir başlangıç noktasından hareket eden USV'nin ortamdaki dinamik engellerden sakınarak verilen bitiş noktasına ulaşmasıdır.

Çalışmada ele alınan varsayımlar:

- Hedef aracın hızı sabit ve doğrusal olarak ele alınmıştır.
- USV'nin hareketi, iki boyutlu (2D) bir düzlem üzerinde modellenmiştir. Yalnızca konum (x, y) ve açı (θ) dikkate alınmış, yükseklik/z-boyutundaki değişimler ve üç boyutlu etkileşimler çalışma kapsamı dışında tutulmuştur.
- USV'nin hareketi, hız ve yönelim değişimlerini içeren bir kinematik modelle hesaplanmaktadır. Araç dinamikleri (kütle, atalet, sürtünme, vb.) bu çalışmanın kapsamı dışında bırakılmıştır.
- Kontrol noktaları sabit aralıklarla (25 m) belirlenmiştir ve her adımda bu noktalar üzerinden alternatif B-Spline yörüngeleri üretilmektedir.
- USV'nin hızı Copernicus Deniz ve Hava Ortam Servisi'nden alınan dalga yönü ve yüksekliği, rüzgâr hızı ve yönü, akıntı hızı bilgileri kullanılarak hesaplanmıştır.
- USV ile diğer deniz araçları veya engeller arasında her zaman minimum güvenlik mesafesi korunmalıdır. Bu mesafe temsili değer olarak d (20 m) yaklaşma mesafesi ile sınırlandırılmıştır.
- USV, hareketli hedefe belirli bir tolerans bölgesi (r) içerisinde ulaşmalıdır. Hedef noktasal değil, yarıçap üzerinden tanımlanmıştır.

Hedeflenen problem özelinde başlangıç noktasından hedefe doğru izlenen ve sabit ve hareketli engellerin bulunduğu bir ortamda seçilen yörünge için örnek gösterim Şekil 3.1'de verilmiştir.



Şekil 3.1. Sabit ve hareketli engellerin olduğu ortamda RL ile yörünge seçimi

Şekil 3.1'de, önerilen sistemde A başlangıç noktasından B noktasına ulaşmak amacıyla çoklu B-Spline yörünge seçeneklerinin oluşturulma sürecini ve bu yörüngelerin pekiştirmeli öğrenme tabanlı karar mekanizmasıyla seçimini göstermektedir. Her bir yörünge, belirli mesafe aralıklarıyla yerleştirilmiş kontrol noktaları yardımıyla tanımlanmakta ve farklı manevra alternatifleri üretmektedir. Mavi renkli yörüngeler, engellerle çakışan ve çarpışma riski taşıyan alternatif yolları temsil ederken, kırmızı yörünge (Y_0), RL algoritması tarafından ödül fonksiyonlarına göre seçilmiş, engellerden kaçınan en uygun ve güvenli yolu göstermektedir. Sabit (turuncu kare) ve hareketli (mor üçgen) engeller, navigasyon alanında belirsizlik ve dinamik riskleri ifade eder. Önerilen sistem ile, yörünge planlamasında B-Spline eğrileri ile esnek manevra yaptırmakta, RL tabanlı karar yapısı sayesinde ise COLREG kurallarına uygun, çarpışmasız ve hedef odaklı bir yörünge belirlenmektedir. Bu yaklaşım, otonom deniz araçlarının karmaşık ve dinamik deniz ortamlarında güvenli seyir yeteneğini artırmayı hedeflemektedir.

Örnek senaryoda Y_0 yörüngesi için A-B arasında 200 metrelik mesafede her 25 metrede bir oluşturulan kontrol noktalarında engelle karşılaşmama durumu ilk 3 adım için Çizelge 3.1'de verilmiştir. $0 - 25\ m$ aralığında, $V = 2\ m/s$, $W = 0.10\text{rad/s}$ seçilmiştir. Engelle karşılaşma durumu ilk 3 adım için Çizelge 3.2'de verilmiştir. $50 - 75\ m$ aralığında

$V = 2\ m/s,$ $50 - 60\ m$: W $=$ 0.10 rad/s, $60 - 70$ m: W $=$ 0.20 rad/s $70 -$ 75 m: W $=$ 0.05 rad/s seçilmiştir.

Çizelge 3.1. Y_0 yörüngesi için 0-25 m aralığında kinematik denklem hesaplamaları

X	**Y**	**Q**	**Açıklama**
$x_1 = 0 + 2.0 \times \cos(0) \times 0.5 = 1.0$	$y_1 = 0 + 2.0 \times sin(0) \times 0.5 = 0$	$\theta_1 = 0 + 0.10 \times 0.5 = 0.05$	Yukarı doğru sapma başlar.
$x_2 = 1.0 + 2.0 \times cos(0.05) \times 0.5 \approx 1.99$	$y_2 = 0 + 2.0 \times sin(0.05) \times 0.5 \approx 0.050$	$\theta_2 = 0.05 + 0.10 \times 0.5 = 0.10$	Yukarı sapmada dönüş derinleşir.
$x_3 = 1.99 + 2.0 \times cos(0.10) \times 0.5 \approx 2.994$	$y_3 = 0.050 + 2.0 \times sin(0.10) \times 0.5 \approx 0.150$	$\theta_3 = 0.10 + 0.10 \times 0.5 = 0.15$	Dönüş derinleşir.

Çizelge 3.1'de $0 - 25\ m$ ve Çizelge 3.2'de $50 - 75\ m$ aralığında ilk 3 adım için kinematik denklemler yardımıyla konum ve açı hesabı yapılarak eğri oluşturulmaya başlanmıştır. Eğrinin tamamlanması için benzer şekilde yapılan hesaplamaların sonuçları Çizelge 3.3'te verilmiştir.

Çizelge 3.2. Y_0 yörüngesi için 50-75 m aralığında kinematik denklem hesaplamaları

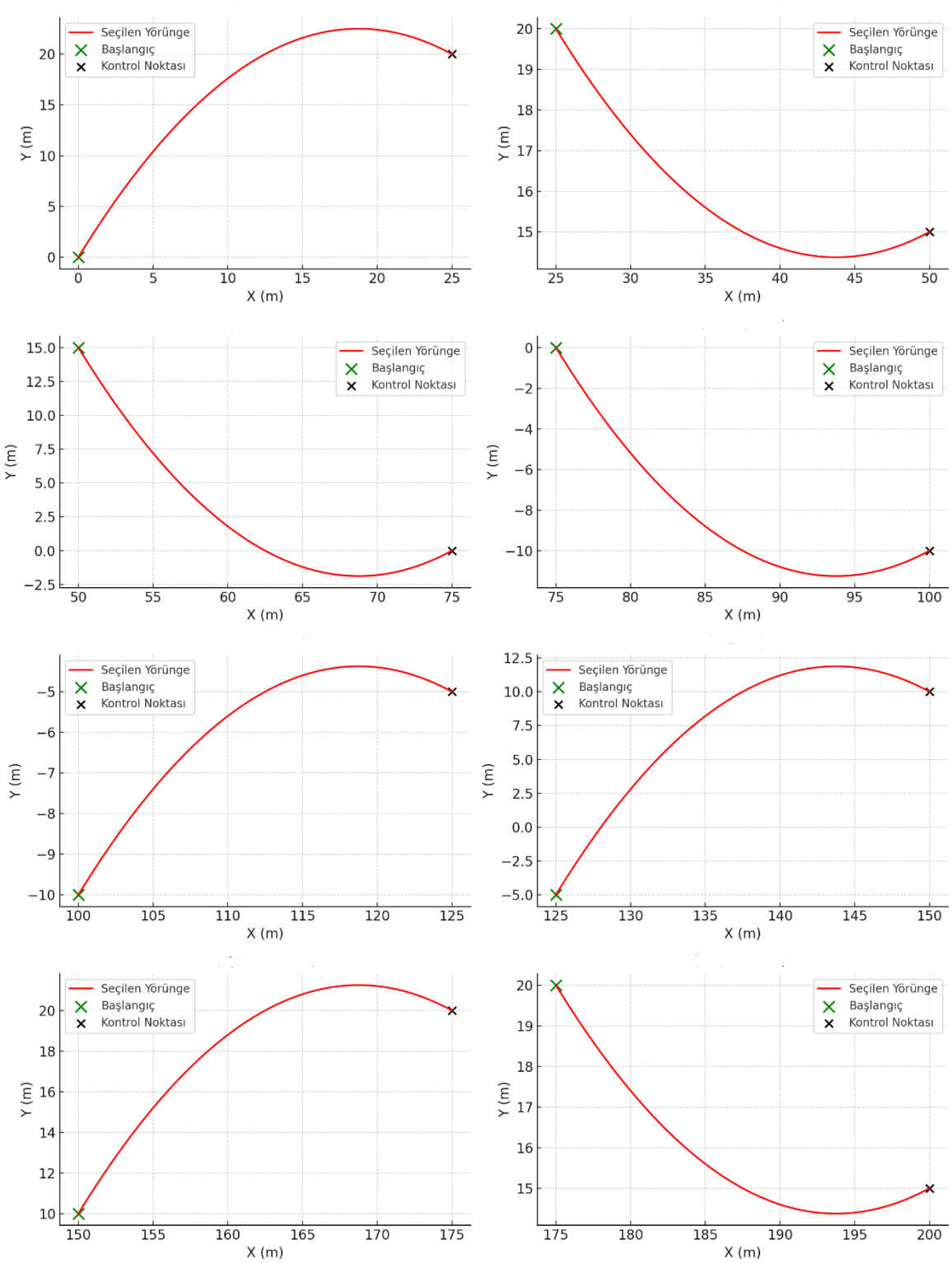
X	**Y**	**Q**	**Açıklama**
$x_1 = 37.951 + 2.0 \times cos(0.000) \times 0.5 = 38.951$	$y_1 = (-27.381) + 2.0 \times sin(0.000) \times 0.5 = (-27.381)$	$\theta_1 = 0.000 + 0.10 \times 0.5 = 0.050$	Yukarı doğru sapma başlar.
$x_2 = 38.951 + 2.0 \times cos(0.050) \times 0.5 \approx 39.950$	$y_2 = (-27.381) + 2.0 \times sin(0.050) \times 0.5 \approx (-27.331)$	$\theta_2 = 0.050 + 0.10 \times 0.5 = 0.100$	Engelden sapma sebebiyle dönüş derinleşir.
$x_3 = 39.950 + 2.0 \times cos(0.100) \times 0.5 \approx 40.945$	$y_3 = (-27.331) + 2.0 \times sin(0.100) \times 0.5 \approx (-27.231)$	$\theta_3 = 0.100 + 0.10 \times 0.5 = 0.150$	Dönüş derinleşir.

Çizelge 3.3. 0-25 m ve 50-75 m arasında Y_0 yörüngesinde elde edilen sonuçlar

Adım	X (m)	Y (m)	θ (rad)
0	0.000	0.000	0.000
1	1.000	0.000	0.050
2	1.999	0.050	0.100
3	2.994	0.150	0.150
4	3.983	0.299	0.200
5	4.963	0.498	0.250
6	5.932	0.745	0.300
7	6.887	1.041	0.350
8	7.826	1.384	0.400
9	8.747	1.773	0.450
10	9.648	2.208	0.500
11	10.525	2.688	0.550
12	11.378	3.210	0.600
13	12.203	3.775	0.650
14	12.999	4.380	0.700
15	13.764	5.024	0.750
16	14.496	5.706	0.800
17	15.192	6.423	0.850
18	15.852	7.175	0.900
19	16.474	7.958	0.950
20	17.056	8.771	1.000
21	17.596	9.613	1.050
22	18.094	10.48	1.100
23	18.547	11.371	1.150
24	18.956	12.284	1.200
25	19.318	13.216	1.250

Adım	X (m)	Y (m)	θ (rad)
50	37.951	-27.381	0.000
51	38.951	-27.381	0.050
52	39.950	-27.331	0.100
53	40.945	-27.231	0.150
54	41.934	-27.082	0.200
55	42.914	-26.883	0.250
56	43.883	-26.636	0.300
57	44.838	-26.340	0.350
58	45.777	-25.997	0.400
59	46.698	-25.608	0.450
60	47.599	-25.173	0.500
61	48.476	-24.693	0.550
62	49.329	-24.171	0.600
63	50.154	-23.606	0.650
64	50.950	-23.001	0.700
65	51.715	-22.357	0.750
66	52.447	-21.675	0.800
67	53.143	-20.958	0.850
68	53.803	-20.206	0.900
69	54.425	-19.423	0.950
70	55.007	-18.610	1.000
71	55.547	-17.768	1.050
72	56.045	-16.901	1.100
73	56.498	-16.010	1.150
74	56.907	-15.097	1.200
75	57.269	-14.165	1.250

Çizelge 3.3'te USV'nin $0-25\,m$ ve $50-75\,m$ arasında kinematik denklemlere göre X, Y ve θ'nin aldığı değerler gösterilmektedir. Bu hesaplama ile $0-25\,m$ aralığında USV yukarı doğru sapmaya başlar. $50-75\,m$ aralığında engel olduğu için aşağı doğru değişken açısal hızla manevra yaparak engelden kaçmaktadır. $0-200\,m$ aralığında Y_0 yörüngesine ait bölüm (epoch) bazlı kontrol noktaları ve yörüngeler Şekil 3.2'de gösterilmiştir.



Şekil 3.2. 0-200 m arası bölüm bazlı kontrol noktaları ve yörüngeler

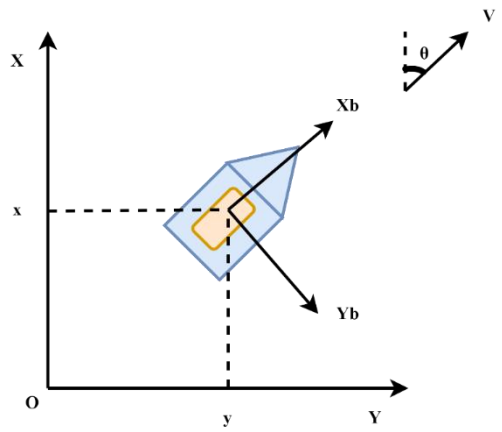
Şekil 3.2'de, Y_0 yörüngesine ait A noktasından B noktasına kadar uzanan toplam 200 metrelik yolun, her biri 25 metrelik sekiz ardışık bölümden oluşan Pekiştirmeli Öğrenme ile seçilen yörünge yapısı sunulmaktadır. Her bölümde, araç sabit doğrusal hız ($V = 2\,m/s$) ile hareket etmekte, bölüm boyunca açısal hız değerleri kullanılarak belirli manevralar yapılmaktadır. Kırmızı eğriler her bir bölümün izlediği kinematik yolu temsil ederken, yeşil X işareti bölümün başlangıç noktasını, siyah nokta ise bir sonraki bölüm için kritik kontrol noktasını göstermektedir. Bu yapı sayesinde, aracın karmaşık deniz ortamında çevresel engellerden güvenli şekilde kaçınarak yönünü kademeli olarak ayarlaması sağlanmakta, COLREG kurallarına uygun bir yol oluşturulmaktadır. Her bölümün bitiş kontrol noktası bir sonraki bölümün başlangıç koşulunu sağlar; bu sayede yörünge sürekliliği ve yönelim tutarlılığı korunur. Tüm bölümler bir araya getirildiğinde, A noktasından B noktasına uzanan birleşik ve çevresel kısıtlamalara uygun bir yörünge elde edilir.

3.3. İnsansız Su Üstü Araç için Yörünge Planlama

Bu bölümde yörünge planlama sürecinde, aracın fiziksel özelliklerini tanımlayan içsel parametreler (örneğin hız, yönelim açısı, konum gibi kinematik değişkenler) ile çevresel koşulları belirleyen dışsal etkenler (dalga hızı, akıntı hızı, rüzgâr hızı gibi) birlikte ele alınmalıdır. Bu parametrelerin uyumlu şekilde modellenmesi, B-Spline temelli esnek yörüngelerin oluşturulmasını mümkün kılmaktadır. Yörünge planlama sürecinde kullanılan bu parametreler sayesinde, USV'lerin belirsiz ve yoğun deniz trafiğine sahip ortamlarda dahi güvenli bir şekilde seyir gerçekleştirebilmesi hedeflenmiştir ve yörünge planlamayı etkileyen iç ve dış faktörler ayrıntılı olarak ele alınmıştır.

3.3.1. Yörünge planlamada iç faktörler

USV kinematik denklemi başlangıç açısı (heading angle), iki boyutlu konum vektörü (x,y), doğrusal hız (forward speed) ve açısal hızdan (angular speed) oluşmaktadır (Yan ve Jiang,2021). Şekil 3.3'te bir USV'nin hareketi sırasındaki konum vektörü ve ilgili parametreler gösterilmektedir.



Şekil 3.3. USV kinematik denklemi

Şekil 3.3'te USV'nin kinematik modeli Denklem 3.1'de verilmektedir.

$$\dot{P}=\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\theta}\end{bmatrix}=\begin{bmatrix}cos\,\theta & 0\\ sin\,\theta & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}V\\ W\end{bmatrix} \tag{3.1}$$

Denklem 3.1'de ifade edilen kinematik model, USV'nin konum ve yönelim açısının zamana bağlı olarak nasıl değiştiğini göstermektedir. Denklemde yer alan $\dot{x}$, $\dot{y}$ ve $\dot{\theta}$ terimleri, sırasıyla x ve y konum bileşenlerinin ve yön açısının zamana göre değişim hızlarını ifade eder. Bu modelde V, USV'nin ileri yöndeki doğrusal hızı; W aracın açısal hızı, θ ise aracın anlık yönelim açısıdır. Bu yapı, özellikle simülasyon ortamında her adımda konum ve yön bilgilerinin güncellenmesine olanak tanıyarak, Pekiştirmeli Öğrenme Algoritmasında toplam ödül/ceza hesaplamalarının yürütülmesine katkı sunar.

USV'nin iç faktörleri aracın çevreyle olan etkileşimini tanımlayan temel kinematik parametrelerdir. Bu parametreler, B-Spline tabanlı yörünge planlama sürecinde, yörüngenin aracın fiziksel sınırlamalarına ve manevra kabiliyetine uygun olarak oluşturulmasını sağlar. Böylece hem izlenebilir hem de denizcilik kurallarına uyumlu güvenli bir yörünge elde edilir. USV kinematik denklemindeki iç ve dış faktörler aracın manevra kabiliyetini etkilemektedir. Çizelge 3.4'te iç faktörler yer almaktadır.

Çizelge 3.4. İç faktörler

Tanım	**Sembol**	**Birim**
İki boyutlu kartezyen koordinat	X	m
İki boyutlu kartezyen koordinat	Y	m
Yönelim Açısı	θ	rad (degree formatında)
Doğrusal Hız	V	m/s (konum değişimi)
Açısal Hız	W	rad/s (teta değişimi)
Araç Yükü	W_d	ton

Çizelge 3.4'te aracın kinematik denklemindeki konum, başlangıç açısı, doğrusal hız, açısal hız ve araç yükü görülmektedir. Örneklem frekansı $10\ hertz$ olacak şekilde kesikli zamana çevrilmiştir. $10\ Hertz$ algoritmanın her saniyede 10 kez çalıştırıldığını ve güncellendiğini göstermektedir. Yani sistem 0.1 saniyelik aralıklar güncellenmektedir. Bu durum daha hassas kontrol, daha pürüzsüz yörünge elde etmek için kullanılmıştır.

3.3.2. Yörünge planlamada dış faktörler

Çevresel etkenler, USV'nin manevra kabiliyeti üzerinde doğrudan etkiye sahip olup, yörünge planlamasında dikkate alınması gereken dinamik değişkenlerdir. Rüzgâr hızı, akıntı ve dalga gibi faktörler, aracın izlediği yörüngede sapmalara neden olabileceğinden, bu etkilere karşı dayanıklı yörüngelerin oluşturulması gereklidir. Gemi yükü gibi değişkenler ise bu dış kuvvetlere karşı aracın tepkisini değiştirdiği için, planlanan yörüngenin doğruluğunu ve güvenliğini doğrudan etkiler. Çizelge 3.5'te aracı etkileyebilecek dış parametre bilgileri yer almaktadır.

Çizelge 3.5. Dış faktörler

Dış Parametre Girdi	**Sembol**	**Birim**
Dalga Yönü	q	Derece
Akıntı Hızı	V_c	m/s
Dalga Boyu	h	m
Rüzgâr Hızı	F	m/s
Rüzgâr Yönü	α	-

Yörünge planlamada dış faktörler Çizelge 3.5'te belirtildiği gibi USV'nin hareketini etkileyen dalga yönü, akıntı hızı, dalga boyu, rüzgâr hızı gibi parametrelerden oluşmaktadır. (Feng, 1992; İnan ve Baba, 2021). Bu faktörler dikkate alınarak geminin gerçek hızı hesaplanmasında Denklem 3.2 kullanılmaktadır.

$$V_a = V_0 - ((1.08 \cdot h - 0.126 \cdot q \cdot h + 2.77 \cdot 10^{-3} \cdot F \cdot cos(\alpha)) \cdot (1 - 2.33 \cdot 10^{-7} \cdot W_d \cdot V_0) \quad (3.3)$$

Gemi hızına akıntının etkisinin hesaplanabilmesi için Denklem 3.3, 3.4 ve 3.5 kullanılmıştır. Denklem 3.3 ve 3.4 kullanılarak Copernicus'tan alınan çevresel etkenlerden arındırılmış hız bileşenleri V_x ve V_y elde edilir (Lin ve Fang, 2013). Boylam ve enlem yönündeki bileşen hızlar kullanılarak USV'nin deniz etkileri göz önüne alınmadan, teorik olarak ulaşabileceği hız (V_0) hesaplanır. Bu değer, USV'nin ideal koşullardaki hızıdır. Son olarak Denklem 3.2 ile USV'nin dalga, rüzgâr ve yük etkisindeki gerçek hızı hesaplanmış olur.

$$V_x = V_c \cdot cos(\alpha_c) - V_{cx} \quad (3.3)$$

$$V_y = V_c \cdot sin(\alpha_c) - V_{cy} \quad (3.4)$$

$$V_0 = \sqrt{V_x^2 + V_y^2} \quad (3.5)$$

Bu denklemlerde;

V_x: Boylamsal akıntının geminin hızına etkisi hesaplandıktan sonra düzeltilmiş hız bileşeni

V_y: Enlemsel akıntının geminin hızına etkisi hesaplandıktan sonra düzeltilmiş hız bileşeni

α_c: Geminin seyir yönü ile akıntı arasındaki açı

V_c: Copernicus sisteminden alınan akıntı hızı

V_{cx}, V_{cy}: Akıntının etkisi çıkarıldıktan sonra, USV'nin x (boylam) ve y (enlem) yönündeki net hız bileşenleri

V_a: Geminin gerçek hızı (m/s)

V_0: Geminin deniz şartlarını göz önüne almaksızın normal hızı (m/s)

h: Dalga boyu (metre)

q: Dalga ile gemi rotası arasındaki açı (radyan)

α: Rüzgâr ile gemi rotası arasındaki açı (radyan)

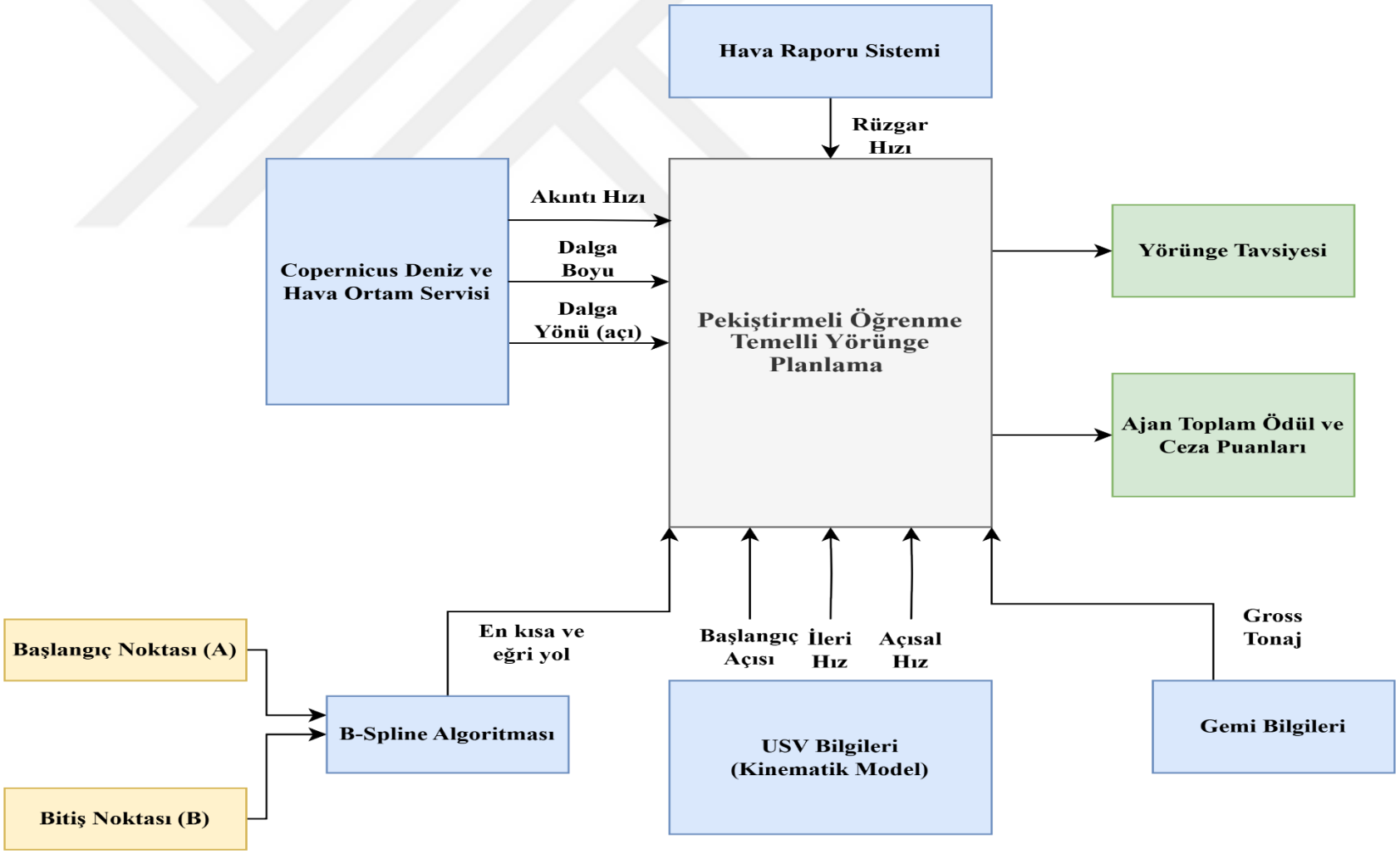
F: Rüzgâr hızı (m/s)

W_d: Aracın yükü (ton) şeklindedir.

3.4. USV Yörünge Planlama için Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, çevresel belirsizliklerin yoğun olduğu deniz ortamlarında USV'lerin güvenli ve kurallara uygun biçimde hareket etmesini sağlayan güçlü bir yaklaşımdır. Bu bölümde, kullanılan Pekiştirmeli Öğrenme Algoritmasının giriş parametreleri, çok bileşenli ödül fonksiyonu ve sistem mimarisi sunulmuştur. Ayrıca, farklı hiper parametre kombinasyonlarının performansa etkisi analiz edilerek en uygun değerler belirlenmiştir.

Bu çalışmada önerilen, hava ve deniz koşullarına duyarlı DQN Algoritmasına dayalı yörünge planlama modeli Şekil 3.4'te sunulmaktadır. Şekil 3.4'te önerilen model, USV'nin çevresel koşulları dikkate alarak verilen A noktasından B noktasına hareket sırasında ortamdaki engeller ve COLREG kurallarını dikkate alarak belli kontrol noktalarında uygun yörüngeleri belirleme için tasarlanmıştır. Bu modeldeki farklı bileşenlerin açıklamaları aşağıda verilmektedir.



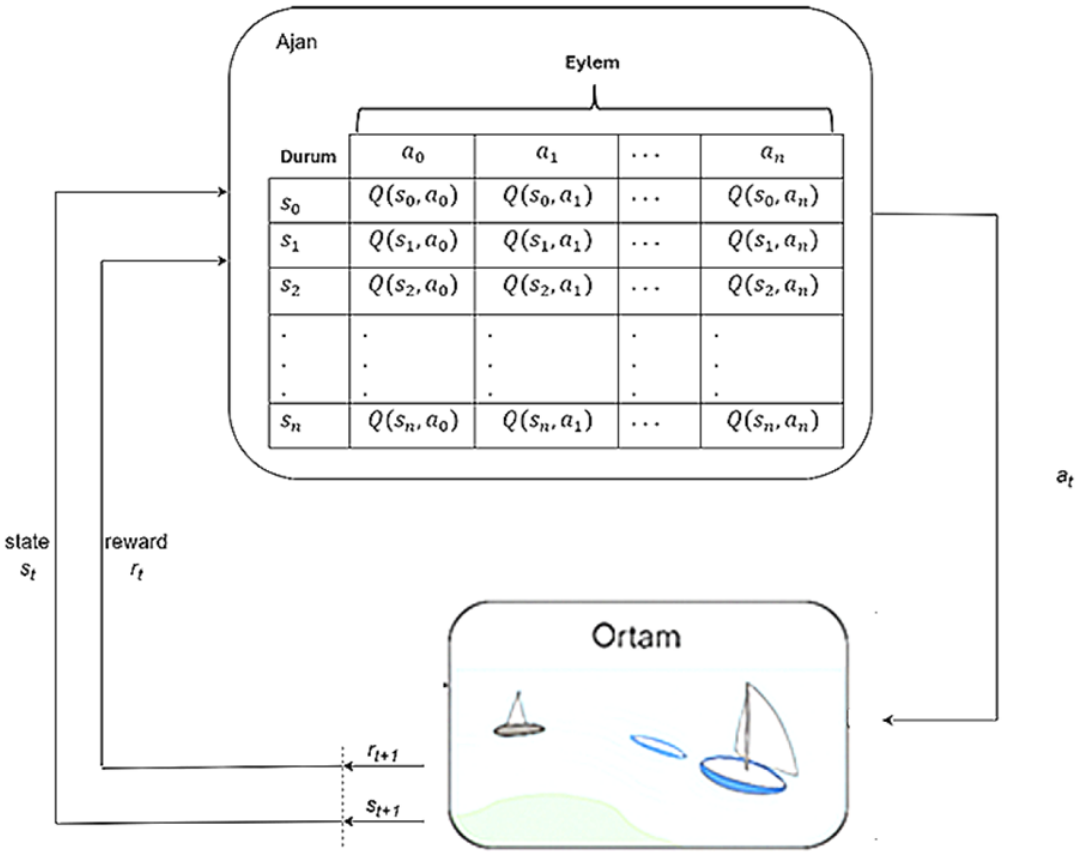
Şekil 3.4. Pekiştirmeli öğrenmeye dayalı yörünge planlama yapısı

USV'nin başlangıç ve varış noktaları USV'nin hareket edeceği güzergâhın sınırlarını tanımlar. Bu iki nokta arasında B-Spline algoritması ile çoklu yörünge seçenekleri oluşturulmaktadır. USV'nin manevra ve hareket kabiliyetini belirleyen kinematik model parametreleri (başlangıç yönü, doğrusal hız, açısal hız) bu aşamada algoritmaya entegre edilir. Gemiye ait statik bilgiler (gross tonaj) hesaplamada dikkate alınır. Çevresel koşullar ise Copernicus Deniz ve Hava Ortam Servisi ile hava raporu sisteminden sağlanan verilere dayanmaktadır. Pekiştirmeli Öğrenme Algoritmasında kullanılan tüm parametreler Çizelge 3.6'da yer almaktadır. Tüm bu parametreler, sistemin merkezi bileşeni olan USV Pekiştirmeli Öğrenme Temelli Yörünge Planlama Modülü tarafından işlenir. Bu modülde, pekiştirmeli öğrenme algoritması devreye girerek her olası yörünge için ödül ve ceza puanları hesaplanır.

Çizelge 3.6. Pekiştirmeli öğrenme tabanlı yörünge planlama hesabında kullanılan girdiler

Parametre	**Açıklama**	**Değer**
d	Yaklaşma Mesafesi	20 m
k_{colreg}	Katsayı	300
U_{cri}	Çarpışma Riski	1
R_{goal}	Hedefe ulaşma katsayı	1000
$R_{çarpışma}$	Çarpışma katsayı	-4000
θ	Başlangıç Açısı (Heading Angle)	rad (degree formatında)
V	Doğrusal Hız (Forward Speed)	m/s (konum değişimi)
W	Açısal Hız (Angular Speed)	rad/s (teta değişimi)
q	Dalga Yönü	Derece
V_c	Akıntı Hızı	m/s
h	Dalga Boyu	m
F	Rüzgâr Hızı	m/s
α	Rüzgâr Yönü	-
W_d	Geminin yükü	ton

Çizelge 3.6 parametreleri kullanılarak yörünge öneren pekiştirmeli öğrenme modeli Şekil 3.5'te gösterilmiştir. Şekil 3.5'te B-Spline tabanlı farklı yörünge seçenekleri arasından DQN Algoritması kullanılarak yapılan seçim süreci görselleştirilmiştir. Her bir durum için, ajan geçerli olan B-Spline yörüngeleri arasından seçim yapar. Ödül fonksiyonu; çarpışma riski, hedefe yakınlık ve COLREG kurallarına uyum gibi faktörleri değerlendirerek Q-değerlerini güncellemektedir.



Şekil 3.5. Önerilen DQN modeli

Şekil 3.5, bu çalışmada kullanılan Derin Q Öğrenme (DQN) algoritmasının karar verme sürecini ve ortamla olan etkileşimini göstermektedir.

State (Durum): USV'nin anlık konumu, yönelimi, hız bilgileri ve çevresel koşullardan (rüzgâr, dalga, akıntı gibi) oluşan gözlemlerdir.

$\boldsymbol{s_0}$: Başlangıç Durumu (A Noktası)

Bu durum, USV'nin göreve başladığı ilk pozisyonu temsil eder. Çalışma kapsamında kullanılan her senaryoda A noktası sabittir ve USV bu noktada hedef yöne doğru hazır beklemektedir. Tüm sensör verileri (konum, yönelim, hız, çevresel koşullar) başlangıç değerindedir. USV'nin çevreyi ilk gözlemlediği ve yörünge alternatiflerini değerlendirmeye başladığı durumdur. Algoritma bu noktada mevcut durum bilgisine göre B-Spline yörünge seçenekleri arasından ilk eylemini seçer.

$s_1 \dots s_{n-1}$: USV'nin Seyir Halindeki Durumları

Bu durumlar, USV'nin belirli bir B-Spline eğrisi üzerinde ilerlerken geçtiği ardışık gözlem anlarını ifade eder. Yani her s_i, yeni bir zaman adımındaki çevresel ve kinetik gözlemlerdir.

- Konum (x, y): USV, yörünge boyunca hareket ettikçe yeni koordinatlara ulaşır.
- Yönelim Açısı (θ): Her segmente geçişte yönelim güncellenir.
- Hız Bilgisi (V, W): Doğrusal ve açısal hız, yörüngeye göre ayarlanır.
- Çevresel Koşullar

Her s_i durumundan sonra, yeni bir eylem (yeni bir B-Spline alternatifi) seçilerek yörüngeye devam edilir. Bu ara durumlar, hedefe ulaşma sürecinde Q değerlerinin güncellendiği kritik öğrenme adımlarıdır.

s_n: Hedefe Ulaşma Durumu (B Noktası)

Bu durum, USV'nin başarıyla hedef noktasına (B noktası) ulaştığı son durumu temsil eder. USV'nin görevi tamamladığı ve karar sürecinin sonlandığı durumdur. Bu durum genellikle pozitif bir ödülle (R_{goal}) değerlendirilir. Eğer USV çarpışma yaşamadan ve COLREG kurallarına uygun şekilde bu duruma ulaşmışsa, en yüksek ödül değerlerinden biri atanır.

Action (Eylem): USV'nin her bir zaman adımında mevcut durumdan (s_t) itibaren takip edeceği B-Spline yörünge alternatiflerinden birinin seçilmesi anlamına gelmektedir. Her bir

eylem, sistemin sonraki adımda uygulayacağı yörüngeyi belirleyerek USV'nin hareket yönünü, çarpışmalardan kaçınma stratejisini ve hedefe doğru ilerleyişini doğrudan etkilemektedir. Ajan çevreden gözlemlediği mevcut durum s_t'ye karşılık olarak bir eylem a_t seçer. Bu süreçte, DQN algoritması kullanılarak hem önceki deneyimlerden öğrenilmiş bilgi hem de yeni alternatif yolların keşfi birlikte değerlendirilir. Ajan, her bir durum için tanımlanan çoklu B-Spline yörünge alternatifleri arasından seçim yapar. Bu seçim, epsilon-greedy stratejisine göre gerçekleştirilir. Bu yöntemle ajan en yüksek Q değerine sahip eylemi seçerek sömürü yapar. Belli bir olasılıkla rastgele bir eylem seçerek keşif sürecini sürdürür. Her bir durum-eylem çifti için gelecekteki toplam ödül beklentisi $Q(s,a)$ değeri ile temsil edilir. DQN algoritmasında bu Q değerleri doğrudan bir tablo üzerinden değil, sinir ağı modeli ile yaklaşık olarak öğrenilir. Bu tahmin süreci, Denklem 2.2 ile gösterilen Q değer güncellemesi üzerine kuruludur. Bu yapıya göre, ajan her yeni durum s_t'de tahmin ağını ($Q(st, at; \theta)$) kullanarak eylemini seçer, ardından yeni duruma geçer ödül alır ve hedef ağ üzerinden yeni Q değerini günceller. Bu süreç, Denklem 2.3 ile verilen hedef Q değerine göre elde edilir. Ajan, bu hedef değeri tahmin ağı ile tahmin edilen değerle karşılaştırarak hata fonksiyonunu tanımlar. Bu hata fonksiyonu, ağın eğitiminde kullanılan kayıp fonksiyonudur ve Denklem 2.4 ile verilir. Bu kayıp, tahmin edilen Q değeri ile hedef Q değeri arasındaki farkın karesidir. Ağın ağırlıkları bu kaybı minimize edecek şekilde güncellenir. Böylece ajan, her yeni örnekle birlikte daha doğru eylem seçimleri yapmayı öğrenir.

Reward (Ödül): Ödül fonksiyonu, USV'nin istenilen hedeflere ulaşmasını sağlamak ve aynı zamanda güvenli ve kurallara uygun şekilde hareket etmesini teşvik etmek amacıyla tasarlanır. Hedefe ulaşma, çarpışmadan kaçınma ve COLREG ile uyum sağlama gibi davranışları öğrenmesini sağlayan 3.6 eşitliğinde belirtilen ödül fonksiyonu dört ana bileşenden oluşmaktadır.

$$R_a = R_{hedef} + R_{hedefmesafe} + R_{çarpışma} + R_{COLREGs} \qquad (3.6)$$

R_{hedef}; USV'nin belirlenen hedefe başarıyla ulaşmasını teşvik eder. USV, hedefe yeterince yaklaştığında bu ödülü alır. Amaç, aracın hedef konuma etkili şekilde yönelmesini sağlamaktır.

$R_{hedefmesafe}$; USV'nin hedefe adım adım yaklaşma sürecini ödüllendirir. Hedefe her yaklaşıldığında pozitif ödül verilir, uzaklaşılırsa ceza uygulanabilir. Böylece USV sadece sonuca değil, sürece de odaklanır.

$R_{çarpışma}$; USV'nin çevresindeki diğer araçlar veya nesnelerle çarpışmaktan kaçınmasını sağlamak için kullanılır. Eğer çarpışma riski varsa ceza uygulanır, güvenli mesafede kalındığında ödül alır.

$R_{COLREGs}$; COLREG, denizdeki çarpışmaları önlemek için belirlenen kurallardır. Bu ödül, USV'nin bu kurallara uygun hareket etmesini sağlar. Eğer USV güvenli ve kuralına uygun manevra yaparsa ödül alır; kurallara aykırı bir hareket yaparsa ceza uygulanır.

- **Hedefe Ulaşma Ödülü**

Hedefe ulaşma ödülü (Rhedef); USV'nin hedefe yaklaşma davranışını teşvik etmek için kullanılır. Bu ödül, USV'nin hedefe yaklaşması ve hedefe ulaşması ile ilişkilidir.

r_{hedef}: Hedefe olan mesafe

k_{hedef}: Hedef ödülünün ölçek faktörü

r: Hedefin çevresindeki küçük bir alan (hedef çevresinde bir "yakınlaşma bölgesi") olmak üzere R_{hedef} 3.7 eşitliğindeki denklem ile hesaplanmaktadır.

$$R_{hedef} = k_{hedef} * r_{hedef}, \sqrt{(x_u - x_t)^2 + (y_u - y_t)^2} < 2r \qquad (3.7)$$

Denklem 3.7'de x_u ve y_u, USV'nin mevcut konumunu, x_t ve y_t ise hedefin konumunu ifade etmektedir. Eğer USV, hedefe $2r$'den daha yakın bir mesafeye geldiyse, yani hedefe ulaştıysa veya çok yaklaştıysa, pozitif bir ödül veriliyor.

- **Bağıl Mesafe Ödülü**

Bağıl Mesafe Ödülü ($R_{hedefmesafe}$); USV'nin hedefe doğru ilerlemesini teşvik eder. USV'nin hedef ile bağıl mesafelerin uzaklığı (hedefe doğru gidiş, hedefe doğru

ilerleme) aşamalı şekilde yapılmasıdır. Hedefe her yakınlaşıldığında ödül artar, hedefe uzaklaşıldığında ise ceza verilebilir. Bu genellikle, hedefe doğru adım adım ilerlemek için kullanılan ödül fonksiyonlarıyla sağlanır.

- **Çarpışmadan Kaçınma Ödülü**

Çarpışmadan Kaçınma Ödülü ($R_{çarpışma}$); USV'nin çarpışmalardan kaçınmasını sağlamak için tasarlanmış bir ödüldür. Burada USV'nin diğer objelerle (örneğin, başka bir USV) çarpışmasını engellemesi beklenir. Çarpışmadan kaçındığında ödül alırken, çarpışma riski oluşturduğunda ceza verilir.

$r_{çarpışma}$: Çarpışma olasılığı ile ilgili mesafe

$k_{çarpışma}$: Çarpışmadan kaçınma ödülünün ölçek faktörü

x_0, y_0: Çarpışma riski oluşturan nesnenin konumu (başka bir USV veya engel)

olmak üzere;

$$R_{çarpışma} = k_{çarpışma} * r_{çarpışma}, \sqrt{(x_u - x_0)^2 + (y_u - y_0)^2} < 2r \qquad (3.8)$$

Eğer USV, bir başka araca çok fazla yaklaşırsa (mesafe $2r$'den küçükse), sistem bu durumu çarpışma riski olarak algılar ve negatif bir ödül (ceza) uygular. Bu, ajanın güvenli mesafeyi koruyarak hareket etmesini teşvik eder.

- **COLREG Kurallarına Uygunluk Ödülü**

USV'nin COLREG kurallarına uymasını teşvik etmek için kullanılan ödüldür. COLREG, denizde çarpışmaları önlemek için belirli davranışları tanımlar. Bu ödül, çarpışma riski oluşturabilecek davranışları cezalandırır ve doğru, güvenli manevraları ödüllendirir.

k_{COLREG}: COLREG ödülünün ölçek faktörü

x_0, y_0: Çarpışma riski oluşturan nesnenin konumu (başka bir USV veya engel)

R: Çevresel güvenlik yarıçapı (yaklaşma mesafesi bu mesafe risk için kabul edilir)

olmak üzere;

$$\sqrt{(x_u - x_0)^2 + (y_u - y_0)^2} << R \ \ ise \ R_{COLREG} = k_{COLREG * Çarpışma\ riski} \qquad (3.9)$$

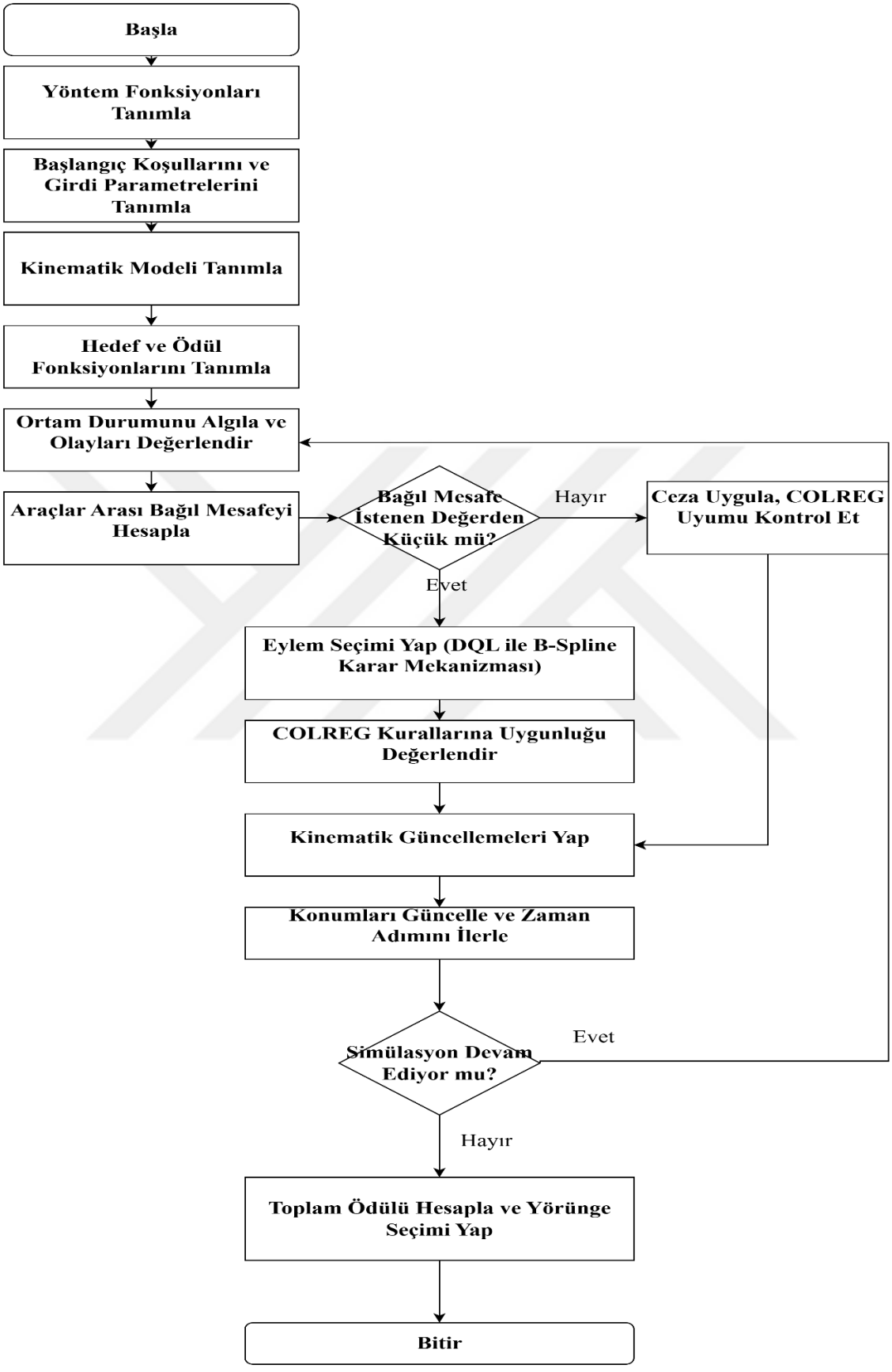
$$\sqrt{(x_u - x_0)^2 + (y_u - y_0)^2} \geq R \ ise \ R_{COLREG} = 0 \qquad (3.10)$$

Denklem 3.9'a göre USV başka bir nesne ya da USV'ye çok yakınsa (yani iki nesne arasındaki mesafe R'den küçükse), potansiyel bir çarpışma riski olduğu kabul edilir. Bu durumda, USV'nin COLREG ile uyumlu davranıp davranmadığı kontrol edilir. Eğer uygun şekilde manevra yapıyorsa, pozitif ödül verilir. Eğer kural dışı, tehlikeli bir manevra yapıyorsa, ceza (negatif ödül) uygulanır. Denklem 3.10'a göre USV ile diğer nesne ya da USV arasında güvenli bir mesafe olduğunu ifade eder (mesafe R'ye eşit veya daha büyük). Bu durumda çarpışma riski yoktur ve bu nedenle COLREG kurallarına ilişkin bir ödül ya da ceza uygulanmaz.

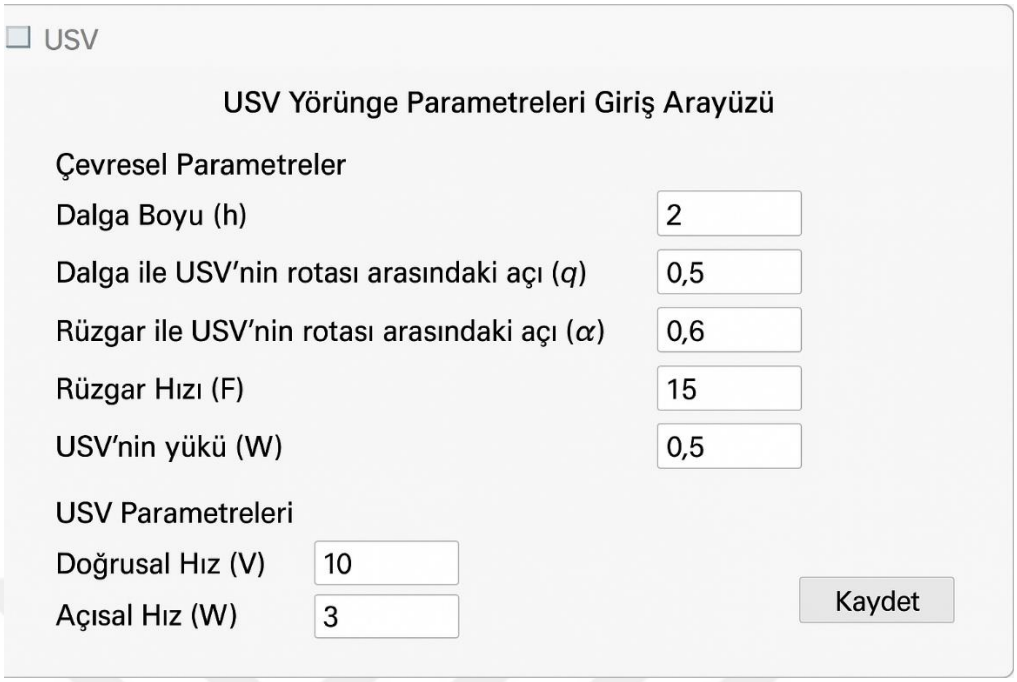
3.5. Pekiştirmeli Öğrenme Temelli Yörünge Planlama Sistemi Akış Diyagramı

Model blok diyagramı 3.4'te verilen sistem için Python 3.5'te kodlanan programın akış şeması Şekil 3.6'da verilmiştir. Şekil 3.6'da önerilen pekiştirmeli öğrenme temelli yörünge planlama sistemine ait akış diyagramı verilmektedir. Bu diyagramda öncelikle sistemin temel bileşenleri olan fonksiyonlar, kinematik denklemler ve bağıl mesafe hesaplamaları tanımlanmakta ve algoritmalara ilişkin işlem adımları açıklanmaktadır. Sistemin kullanıcı etkileşimini artırmak amacıyla, giriş parametrelerinin tanımlanacağı Şekil 3.7'de verilmektedir.

Şekil 3.7'de oluşturulan arayüz için hazırlanan girdiler gösterilmektedir. Bu şekilde aracın başlangıç konumu ve başlangıç açısı sabit (0,0,0) kabul edilmektedir. Bu arayüz üzerinden başlangıç doğrusal ve açısal hızlar tanımlanmaktadır.



Şekil 3.6. Pekiştirmeli öğrenme temelli yörünge planlama sisteminin akış diyagramı



Şekil 3.7. Sistem kullanıcı arayüzü

Akış diyagramı Şekil 3.4'te verilen model alt bileşenlerinin etkileşimlerini tanımlamaktadır. Simülasyon süreci, yönteme ait temel tanımlamaların yapılmasıyla başlar. İlk olarak kullanılacak algoritmalara ilişkin yöntem fonksiyonları tanımlanır. Ardından, USV'nin başlangıç konumu, yönelimi, hızı gibi başlangıç koşulları ile çevresel faktörleri içeren giriş parametreleri belirlenir. Daha sonra aracın hareketini matematiksel olarak modellemek amacıyla kinematik model tanımlanır. Bu temel yapıdan sonra, USV'nin ulaşması gereken hedef noktalar ile bu hedefe doğru ilerlerken alacağı ödül ve ceza yapıları tanımlanır. Ortamdan elde edilen dalga, rüzgâr, akıntı gibi dış etkenler algılanır ve sistemin mevcut durumu değerlendirilir. Araçlar arası bağıl mesafe hesaplanarak çarpışma riski değerlendirmesi yapılır. Hesaplanan mesafe, önceden tanımlanmış güvenlik eşik değerinden küçükse, sistem tarafından ceza uygulanır ve eylemin COLREG kurallarına uygunluğu kontrol edilir. Ardından, DQN algoritması ile durum bilgisine göre B-Spline tabanlı yörüngeler arasından en uygun eylem seçilir. Seçilen eylem, yeniden COLREG kurallarına göre değerlendirilir. Eylem sonucunda USV'nin kinematik bilgileri güncellenir ve simülasyon zamanı bir adım ilerletilerek konum bilgileri yenilenir. Simülasyonun sona erip ermediği kontrol edilir; süreç devam ediyorsa döngü tekrarlanır. Simülasyon tamamlandığında, toplam ödül hesaplanır ve elde edilen sonuçlara göre en uygun yörünge seçilir. Tüm bu işlemler tamamlandığında simülasyon süreci sona erer.

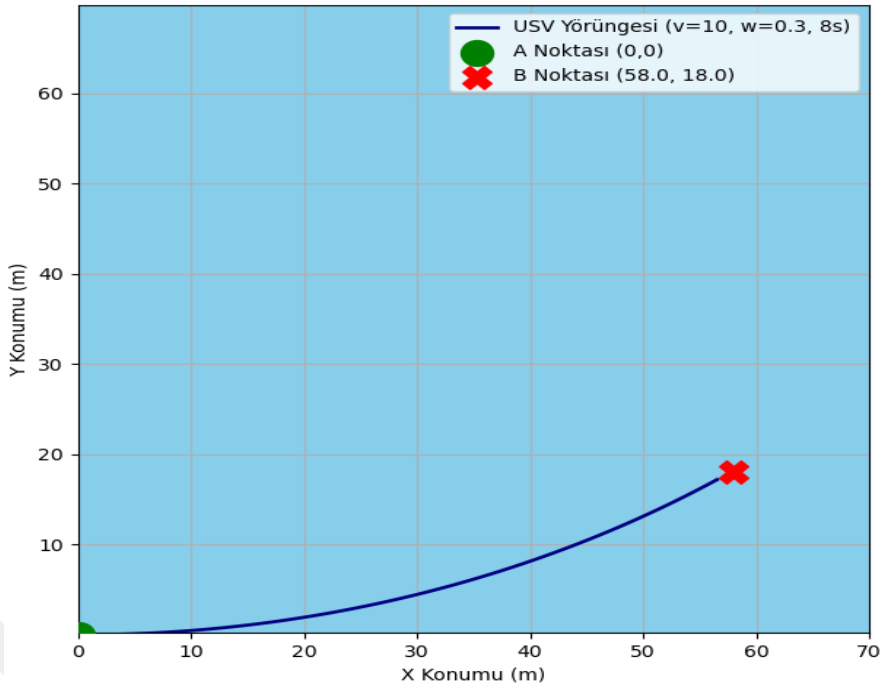
4. BULGULAR VE TARTIŞMA

Bu bölümde geliştirilen pekiştirmeli öğrenme algoritmasının farklı senaryolar altında gösterdiği performans değerlendirilmiş ve sonuçlar detaylı bir şekilde analiz edilmiştir. Eğitim süreci boyunca elde edilen ödül, kayıp fonksiyonu, Q-değerleri gibi temel göstergelerle birlikte sistemin COLREG kurallarına uygunluk, çarpışma önleme başarısı ve çevresel etkenlere karşı tepkisi değerlendirilmiştir.

4.1. USV Aracın Kinematik Model Doğrulaması

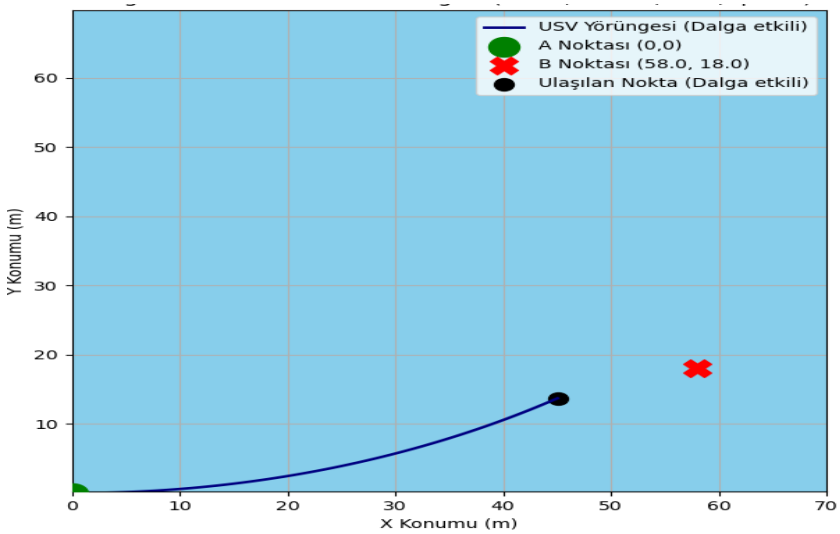
Doğrulama (Verification), geliştirilen sistemin önceden belirlenmiş gereksinimlere ve tasarım kriterlerine uygun olarak doğru biçimde inşa edilip edilmediğini değerlendirme sürecidir (Yazıcı vd., 2021). Bu süreçte, sistemin teknik ve fonksiyonel gereksinimleri sağlayıp sağlamadığı çeşitli analizler ve testlerle incelenir. Amaç, sistemin teorik modellemelere ve yazılım spesifikasyonlarına uygun şekilde çalışıp çalışmadığını belirlemektir. Doğrulama süreçleri, özellikle otonom sistemler ve otomatik kontrol yapıları gibi kritik uygulamalarda güvenli, doğru ve güvenilir bir operasyon sağlamak açısından büyük öneme sahiptir.

Kinematik model doğrulama aşamasında, USV'nin hareketi iç ve dış parametrelere dayalı olarak modellenmiştir. Böylece sistemin yalnızca aracın yapısal ve kinematik özelliklerine bağlı olarak doğru çalışıp çalışmadığı test edilmiştir. İlk olarak akıntı ve dalganın olmadığı bir ortamda kinematik modele göre hareket eden USV Şekil 4.1'de gösterilmiştir.



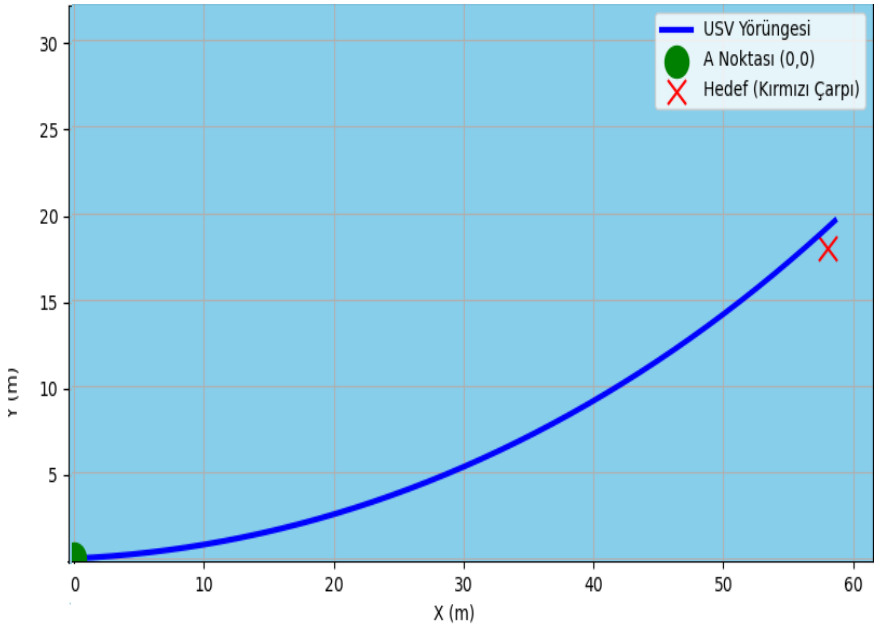
Şekil 4.1. Engelsiz ortamda USV hareketi

Şekil 4.1'de, USV sabit doğrusal hız ($V = 10\,m/s$) ve sabit açısal hız ($W = 0.1\,rad/s$) altında engelsiz ortamda çizdiği yörüngeyi göstermektedir. A Noktası (0,0) başlangıç konumudur ve yeşil daire ile gösterilmiştir. Dalga etkisinde USV'nin izlediği hareketi gösteren şekil 4.2'de verilmiştir.



Şekil 4.2. Dalga etkisinde USV hareketi

Şekil 4.2'de yörünge, dalganın USV üzerindeki etkisini hesaba katmaktadır. Burada ($h = 2$, $q = 0.5$) için yapılan simülasyonda aynı doğrusal ve açısal hız değerleri olmasına rağmen B noktasına ulaşamamıştır. Akıntı etkisinde USV'nin izlediği hareketi gösteren şekil 4.3'te verilmiştir.



Şekil 4.3. Akıntı etkisinde USV hareketi

Şekil 4.3'te, USV'nin A noktasından B noktasına olan seyri sırasında deniz ortamındaki akıntı etkisi dikkate alınmıştır. USV'nin akıntı etkisi ($V_c = 0.5\,m/s, V = 10\,m/s,\ W = 0.1\,rad/s,\ \alpha_c$=0.5 rad) altında çizdiği yörüngeyi göstermektedir. USV hedef noktadan bir miktar sapma göstermiştir.

4.2. Pekiştirmeli Öğrenme Hiper Parametreleri

Kullanılan pekiştirmeli öğrenme algoritmasının başarımını artırmak ve modelin eğitim sürecini kararlı hale getirmek amacıyla, hiper parametrelerin belirlenmesinde hem literatürden elde edilen bilgilerden hem de deneysel analizlerden yararlanılmıştır. Öncelikle, öğrenme oranı (learning rate), indirim faktörü (discount factor), keşif oranı (epsilon), deneyim arabelleği boyutu (replay buffer size) ve parti boyutu (batch size) gibi temel hiper parametreler için literatürde sıklıkla tercih edilen değer aralıkları incelenmiştir. Örneğin, DQN tabanlı çalışmaların büyük çoğunluğunda öğrenme oranı olarak 0.0001 ila 0.01

aralığında, gamma değerinin ise 0.95–0.99 aralığında seçildiği görülmüştür. Benzer şekilde, epsilon değeri için genellikle 1'den başlayarak eğitim süreci boyunca 0.01 seviyesine kadar lineer azalma tercih edilmektedir (Mnih vd., 2015; Lillicrap vd., 2016).

Çalışmada ağ mimarisi olarak 2 katmandan oluşan bir yapı tercih edilmiştir. Her bir katmanda 512 nöron kullanılmıştır. Bu seçim hem modelin öğrenme kapasitesini artırmakta hem de aşırı parametre yükünden kaçınarak hesaplama maliyetlerini yönetilebilir düzeyde tutmaktadır. Literatürde benzer boyuttaki durum-eylem uzaylarında genellikle 2–3 katmanlı ve her katmanda 256–1024 arası nöron içeren yapılar önerilmektedir (Mnih vd., 2015; Lillicrap vd., 2016).

Yapılan ön testlerde daha küçük yapılar (örneğin 128 veya 256 nöron) kullanıldığında modelin öğrenme kapasitesinin sınırlı kaldığı; daha büyük yapılar (1024 nöron veya daha fazla) tercih edildiğinde ise eğitim süresinin belirgin şekilde uzadığı ve modelin aşırı öğrenmeye eğilim gösterdiği gözlemlenmiştir. Bu nedenle literatür incelemelerinde 512 nöronluk 2 katmanlı yapı, performans ve verimlilik açısından dengeli bir seçim olarak öne çıkmıştır. Öğrenme sürecinin verimliliği, doğruluğu ve kararlılığı üzerinde kritik etkiye sahip olan hiper parametreleri Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Pekiştirmeli öğrenme hiper parametreleri

Parametre	**Değer**
Öğrenme Hızı (Learning Rate) (α)	0.001
İndirim faktörü (Discount Factor) (γ)	0.99
İlk Keşif (Initial Exploration) (ε)	1
Son Keşif (Final Exploration)	0.01
Arabellek boyutu (Replay Memory Size)	1000
Parti boyutu (Batch Size)	2
Simülasyon süresi (Simulation Time)	200

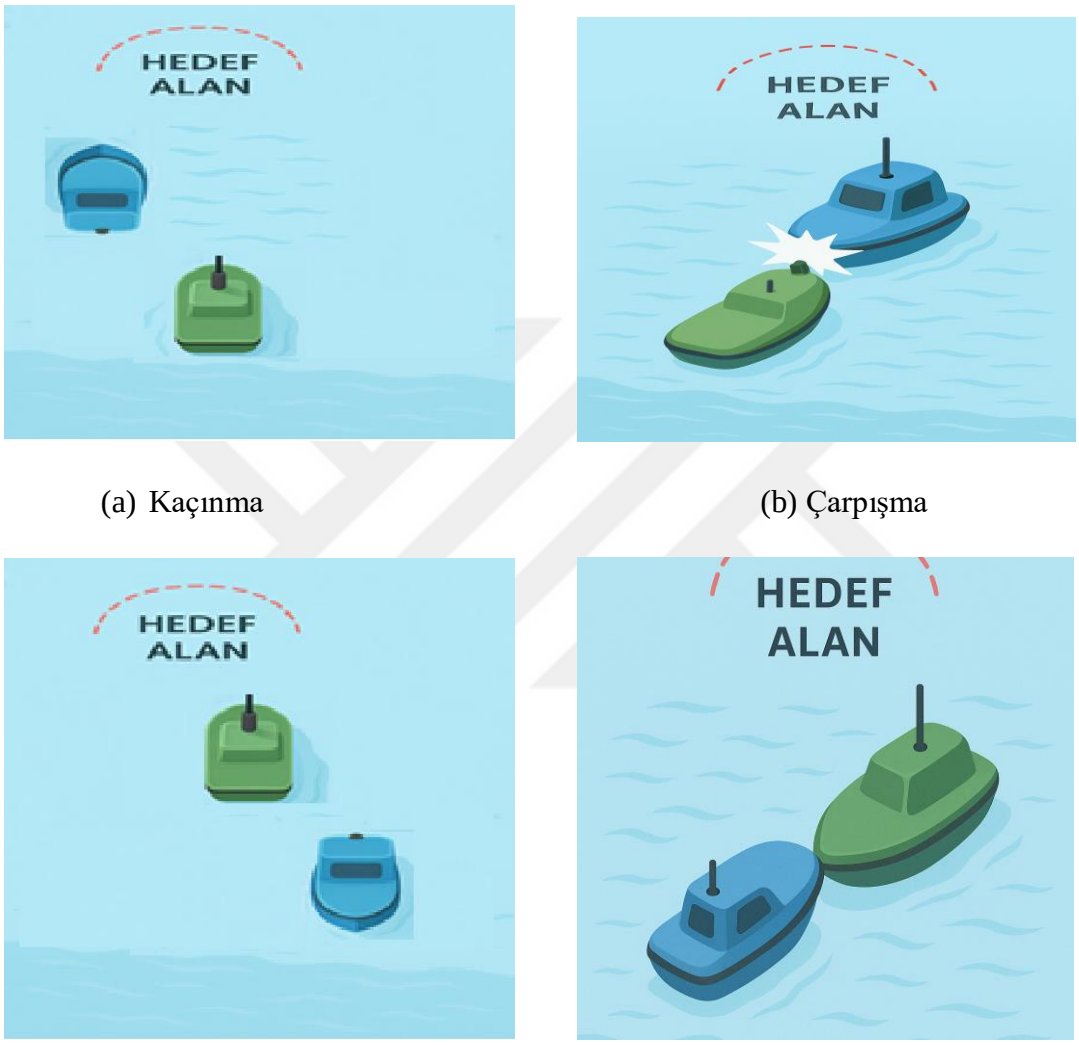
Çizelge 4.1'de bu çalışmada kullanılan hiper parametreler pekiştirmeli öğrenme algoritmasının başarıyla eğitilebilmesi için belirlenmiştir.

- **Öğrenme oranı (Learning Rate):** 0.001 olarak belirlenmiştir. Bu düşük değer, modelin çevresel geri bildirimlere karşı küçük adımlarla güncellenmesini sağlayarak daha stabil bir yakınsama süreci oluşturur.
- **İndirim faktörü (Discount Factor, γ):** 0.99 olarak seçilmiştir. Bu değer, ajanın uzun vadeli ödülleri dikkate alarak geleceğe yönelik stratejiler geliştirmesine olanak tanır.
- **Keşif oranı (Epsilon):** Başlangıçta ε=1 olarak belirlenmiş ve eğitim süreci boyunca doğrusal bir şekilde ε=0.01 seviyesine düşürülmüştür. Bu yaklaşım, ajanın ilk aşamalarda çevreyi daha geniş biçimde keşfetmesini, sonrasında ise öğrendiği bilgileri kullanarak daha kararlı eylemler seçmesini sağlar.
- **Deneyim arabelleği boyutu (Replay Memory Size):** 1000 olarak belirlenmiştir. Bu yapı, ajanın geçmiş deneyimlerini saklamasını ve bunlardan rastgele örnekler seçerek öğrenme sürecini gerçekleştirmesini sağlamaktadır. Bu mekanizma, öğrenme sürecinin daha dengeli ve genellenebilir olmasına katkıda bulunmaktadır.
- **Parti boyutu (Batch Size):** 2 olarak seçilmiştir. Bu düşük değer, modelin her adımda sınırlı sayıda deneyim üzerinden güncellenmesini sağlar. Küçük parti boyutu sayesinde öğrenme daha dinamik olur; ancak bu durum aynı zamanda daha fazla gürültüye yol açabilmektedir.
- **Simülasyon süresi:** Ajanın çevresiyle etkileşime girdiği toplam adım sayısı 200'dür. Bu süre boyunca ajan, çevreden bilgi toplayarak öğrenme sürecini gerçekleştirir.

Algoritmanın eğitimi ve test süreci, bir donanım ve yazılım alt yapısında çalışılmıştır. Deneysel çalışmalar Windows 10 İşletim Sistemi, Intel Xeon Gold 5218 CPU, 64 GB RAM donanım bileşenleri üzerinde ve Python 3.9 programlama dili kullanarak gerçekleştirilmiştir. Simülasyon ve model eğitimi için Anaconda Spyder geliştirme ortamı ile TensorFlow 2.12 derin öğrenme kütüphanesi tercih edilmiştir. Eğitim sürecinde her biri 200 adımlık 100 farklı senaryo üzerinden toplam 20000 durum-eylem çifti kullanılmıştır. Modelin eğitimi yaklaşık 3 saat sürmüş ve bu süreçte her 200 adımda bir performans metrikleri kaydedilerek sistemin öğrenme durumu izlenmiştir.

4.3. Kafa Kafaya Çarpışma Senaryosu

Kafa kafaya çarpışma senaryosunda karşılıklı zıt yönlerden gelen iki USV'nin hareketleri incelenmiştir. Gözlemlere ait görseller Şekil 4.4'te gösterilmiştir.

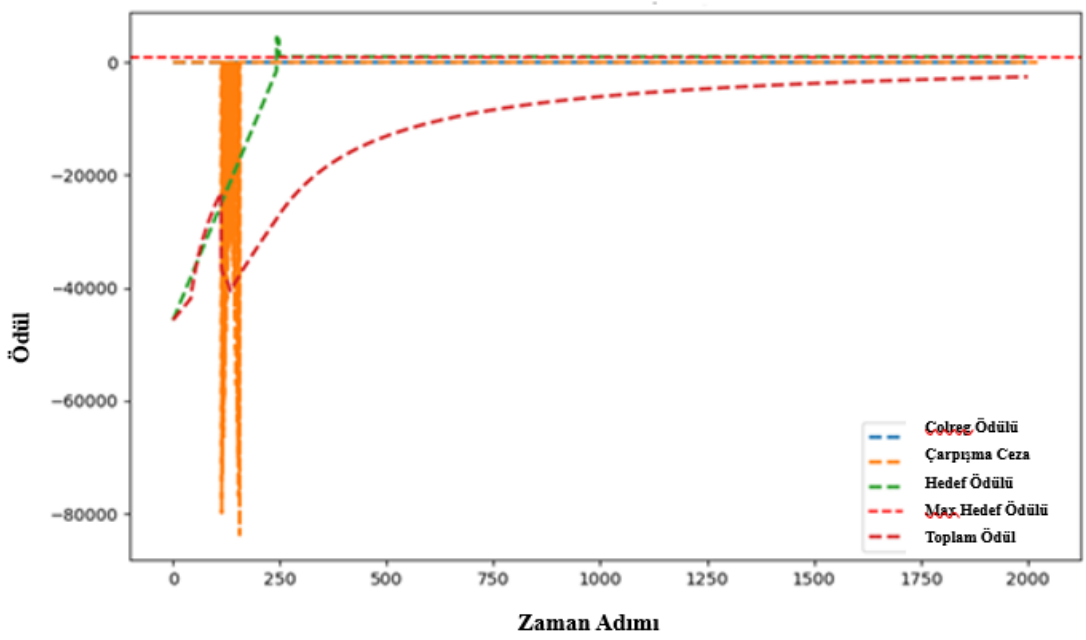


(a) Kaçınma (b) Çarpışma

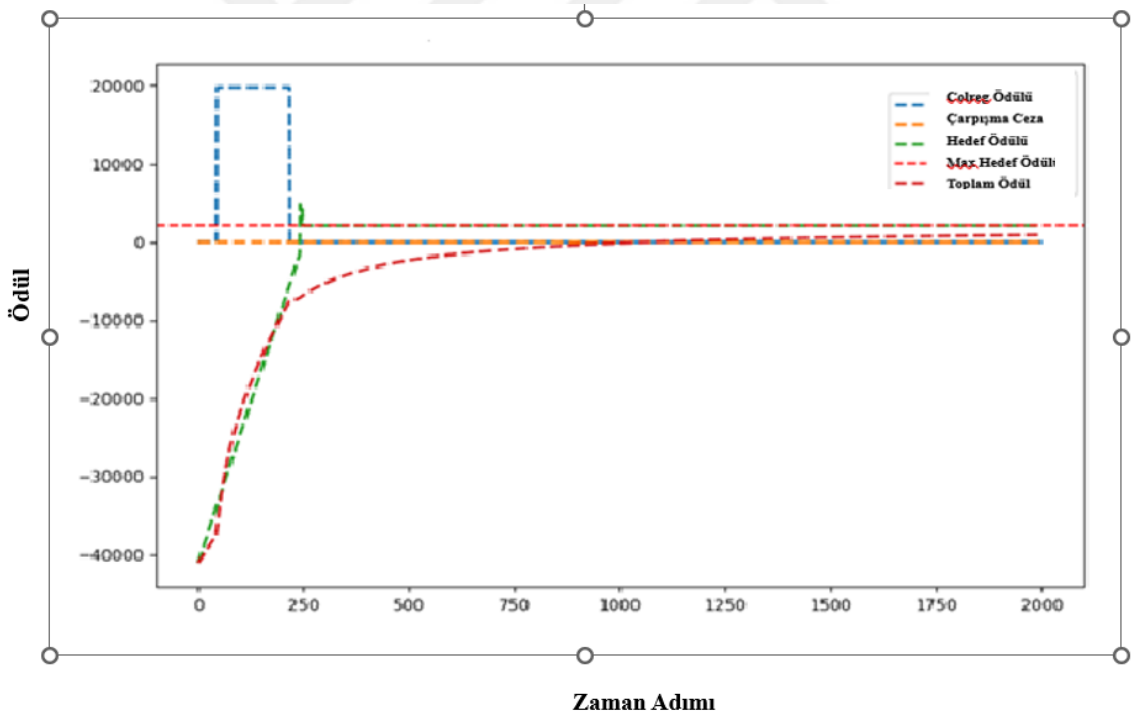
(c) Kaçınma ve hedefe ulaşma (d) Çarpışma

Şekil 4.4. Kafa kafaya çarpışma senaryosunda gözlemlenen olası senaryolar

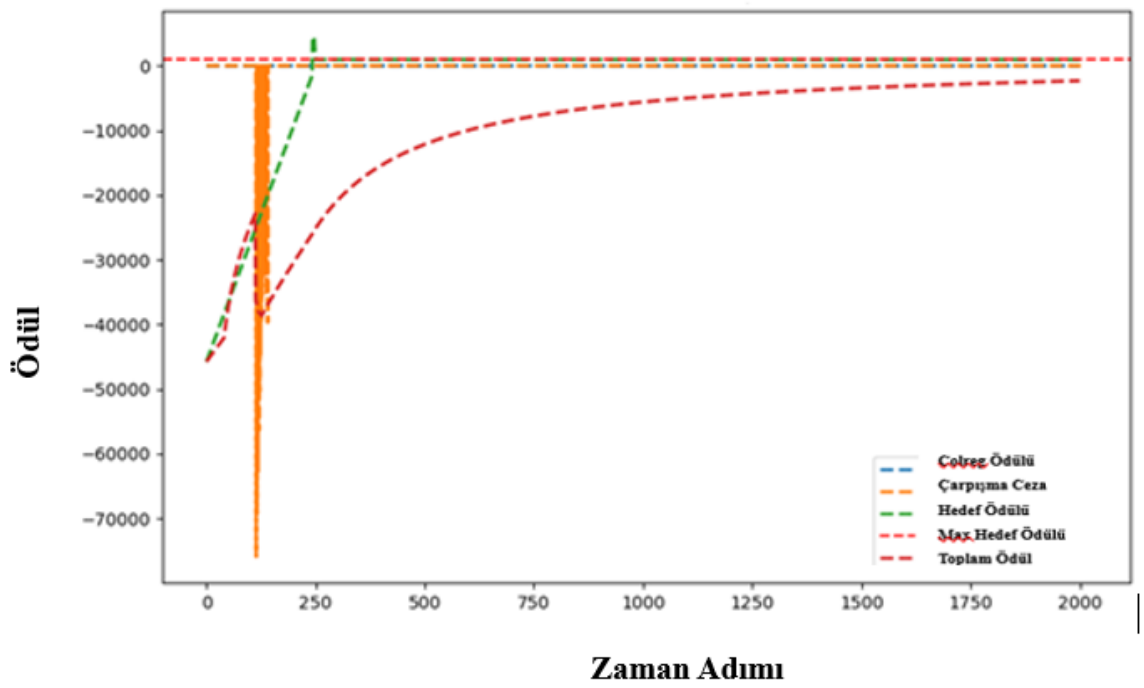
Bu senaryo altında pekiştirmeli öğrenmeye ait çarpışma, kaçınma, hedefe ulaşma ve toplam ödül fonksiyonu durumları Şekil 4.5'te gösterilmiştir.



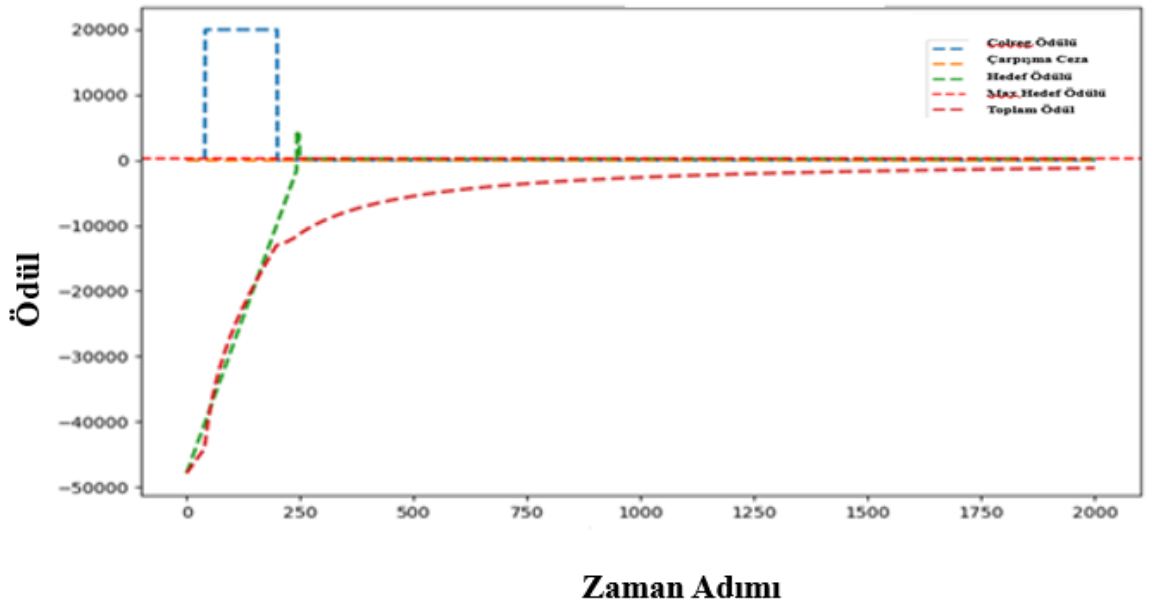
(a) Tam çarpışma



(b) Kaçınma, hedefe ulaşma



(c) Kısmi çarpışma



(d) Kaçınma, hedefe ulaşamama

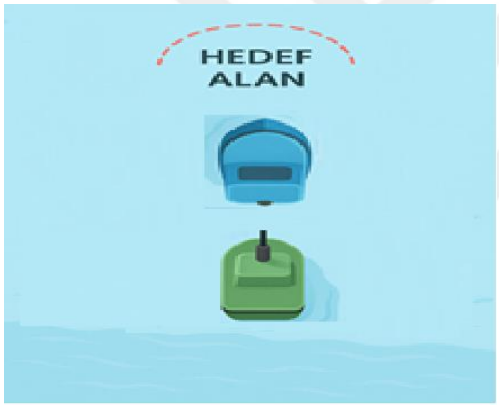
Şekil 4.5. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri

Şekil 4.5 (a) senaryosunda, tam kafa kafaya çarpışma gözlemlenmiştir. Çarpışma sonucu, toplam ödül değerinde ani bir düşüş meydana gelmiştir. Şekil 4.5 (b) senaryosunda, USV COLREG hareketini başarılı bir şekilde uygulamış, kaçınma manevrasını

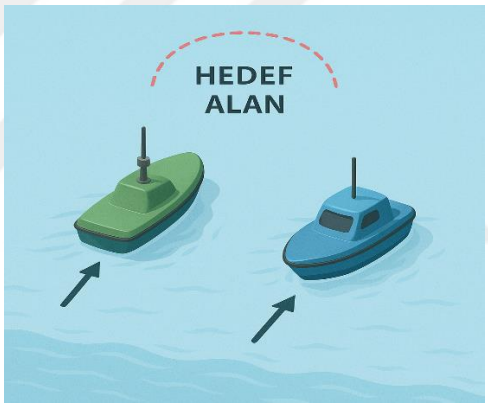
gerçekleştirmiş ve hedefe ulaşmıştır. Bunun sonucunda, toplam ödül değerinde artış gözlemlenmiş ve değer 0'ın üzerine çıkmıştır. Şekil 4.5 (c) senaryosunda ise kısmi bir çarpışma gerçekleşmiş, bu durum toplam ödül değerinde belirgin bir düşüşe yol açmamış olsa da ödül değeri çarpışma nedeniyle 0'ın altında kalmıştır. Çarpışma içeren senaryolarda, hedefe ulaşılmış olsa dahi çarpışma, ödül hesaplamasında daha ağır basmakta ve bu nedenle ödül değeri 0'ın üzerine çıkmamaktadır. Şekil 4.5 (d) senaryosunda, hedeften uzak bir noktaya yöneldiği için toplam ödüle anlamlı bir katkı sağlayamamıştır.

4.4. Arkadan Gelerek Sollama Senaryosu

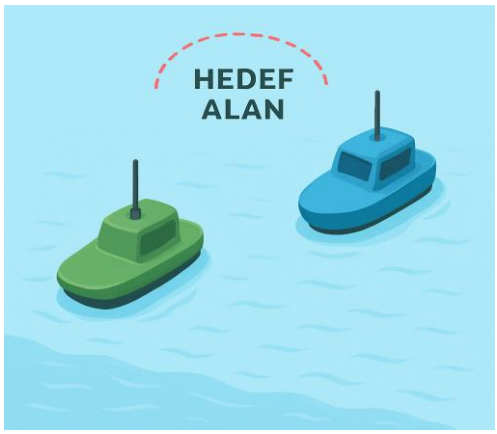
Arkadan gelerek sollama senaryosunda arkadan gelen USV'nin hareketleri gözlemlenmiştir. Gözlemlere ait görseller Şekil 4.6'da gösterilmiştir.



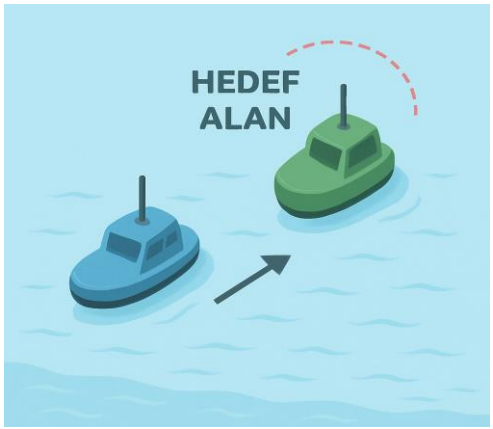
(a) Çarpışma



(b) Kaçınma, hedefe ulaşamama



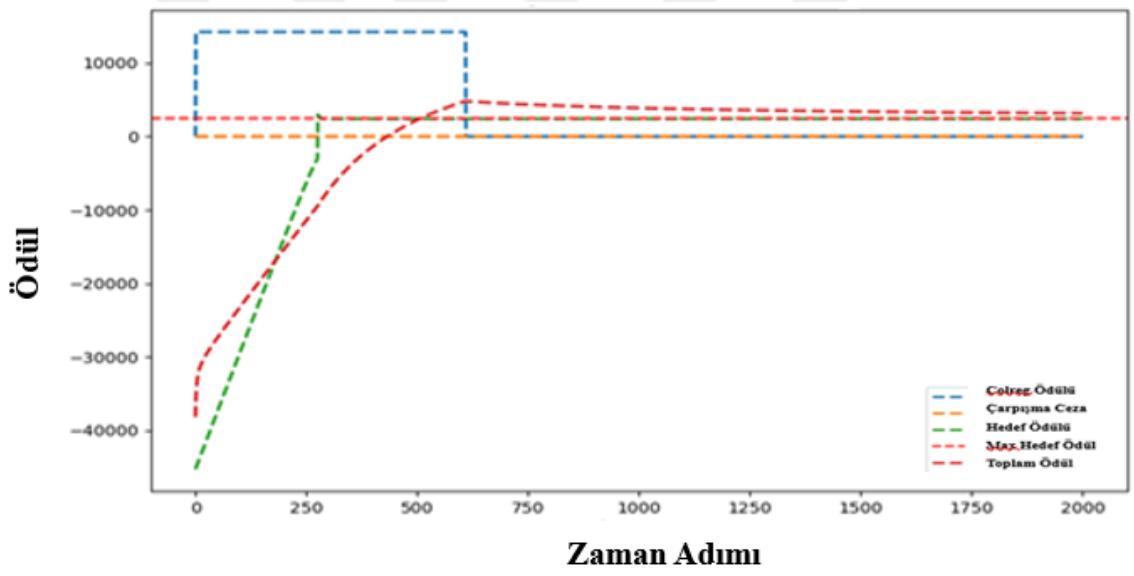
(c) Kaçınma, hedefe ulaşamama



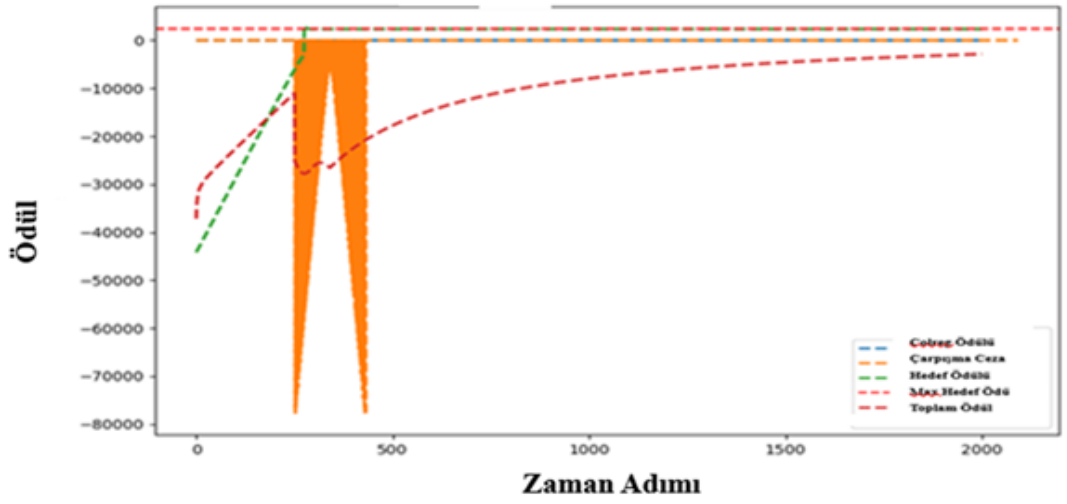
(d) Kaçınma, hedefe ulaşma

Şekil 4.6. Sollama senaryosunda gözlemlenen olası senaryolar

Şekil 4.6 (a) senaryosunda, USV sağ önden açısal manevra uygulayarak çarpışmayı engellemeye çalışmış, ancak bu manevra yeterli olmayarak çarpışma gerçekleşmiştir. Şekil 4.6 (b) senaryosunda, araç sağ önden geniş bir açısal manevra uygulamış ve çarpışmadan kaçınmayı başarmıştır; ancak bu manevra sonucunda hedef noktaya uzak kalmıştır. Şekil 4.6 (c) senaryosunda, araç bu kez sol önden yakın mesafede bir açısal manevra yaparak çarpışmayı engellemiş, ancak yine de hedef noktaya uzak kalmıştır. Şekil 4.6 (d) senaryosunda ise, USV sağ önden geniş bir açısal manevra uygulayarak çarpışmayı engellemiş ve hedef noktaya çok yakın bir şekilde başarıyla ulaşmıştır. Simülasyon esnasında örnek olarak ***dalga hızı 5 m/s, deniz akıntısı 0.5 m/s, dalga açısı 0.6 radyan*** alınmıştır. Bu senaryo altında pekiştirmeli öğrenmeye ait çarpışma, kaçınma, hedefe ulaşma ve toplam ödül fonksiyonu durumları Şekil 4.7'de gösterilmiştir.



(a) Kaçınma, hedefe ulaşma



(b) Çarpışma

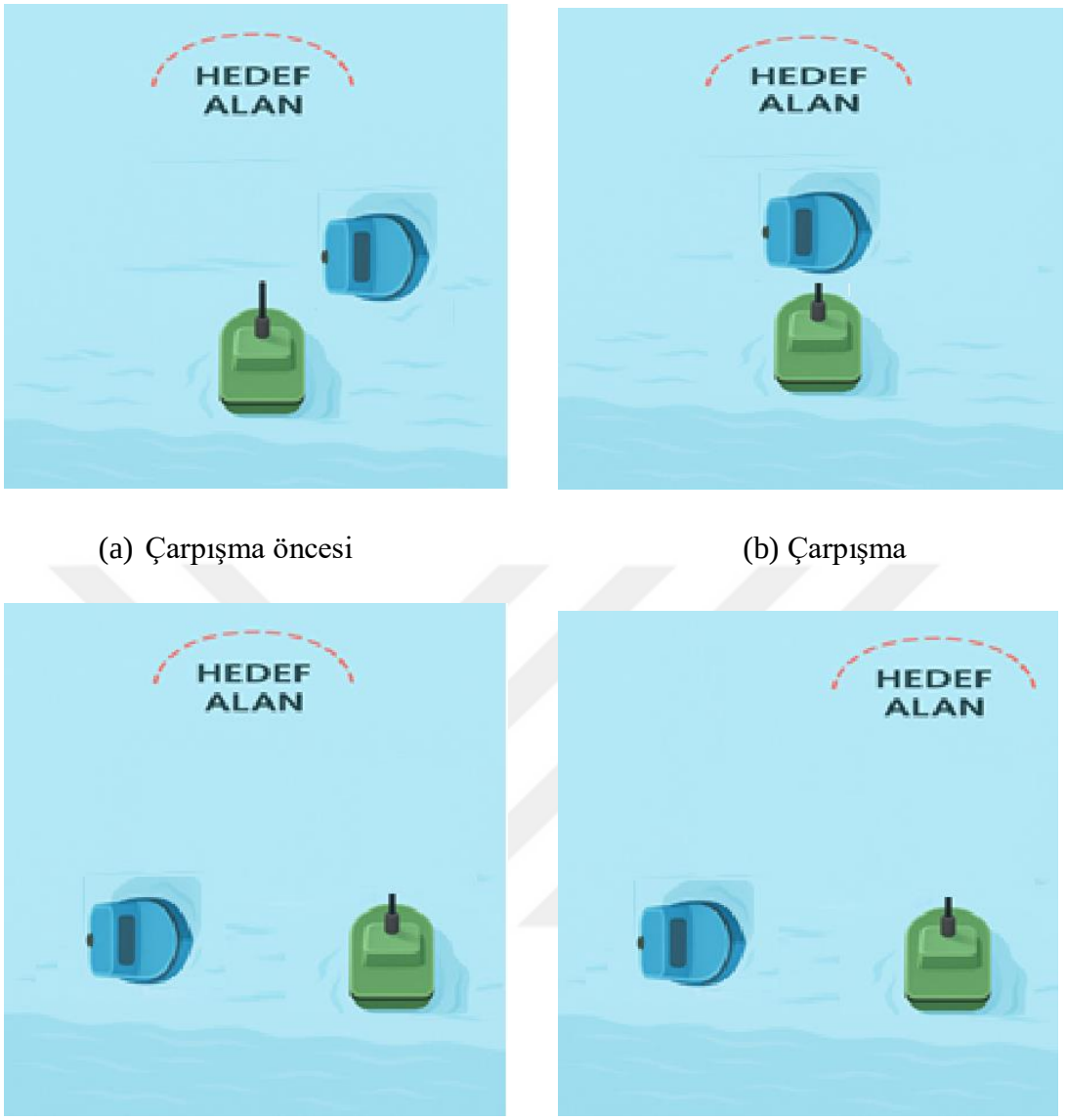
Şekil 4.7. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri

Şekil 4.7 (a) senaryosunda arkadan gelen USV, kaçınma manevrasını başarıyla gerçekleştirmiş ve hedefe ulaşmıştır. Bu durumda toplam ödül değeri pozitif olarak gözlemlenmiştir. Şekil 4.7 (b) senaryosunda ise, büyük bir çarpışma meydana gelmiş, bu nedenle toplam ödül değeri aniden düşmüş, hedefe ulaşılamamış ve simülasyon sonunda toplam ödül değeri 0'ın altında kalmıştır.

4.5. Yol Veren Çarpışma Senaryosu

Yol Veren Çarpışma senaryosunda sağ taraftan gelen USV'ye göre hareketleri gözlemlenmiştir. Gözlemlere ait görseller Şekil 4.8'te gösterilmiştir.

Şekil 4.8 (a) ve (b) senaryolarında çarpışma gerçekleşmeden önceki anlar gözlemlenmektedir. Şekil 4.8 (c) senaryosunda, kaçınma manevrası başarıyla yapılmış ancak USV hedefe uzak bir konumda kalmıştır. Şekil 4.8 (d) senaryosunda ise, kaçınma hareketi başarıyla gerçekleştirilmiş ve USV hedefe yakın bir konumdan geçiş sağlamıştır. Simülasyon esnasında örnek olarak ***dalga hızı 5 m/s, deniz akıntısı 0.5 m/s, dalga açısı 0.6 radyan*** alınmıştır.

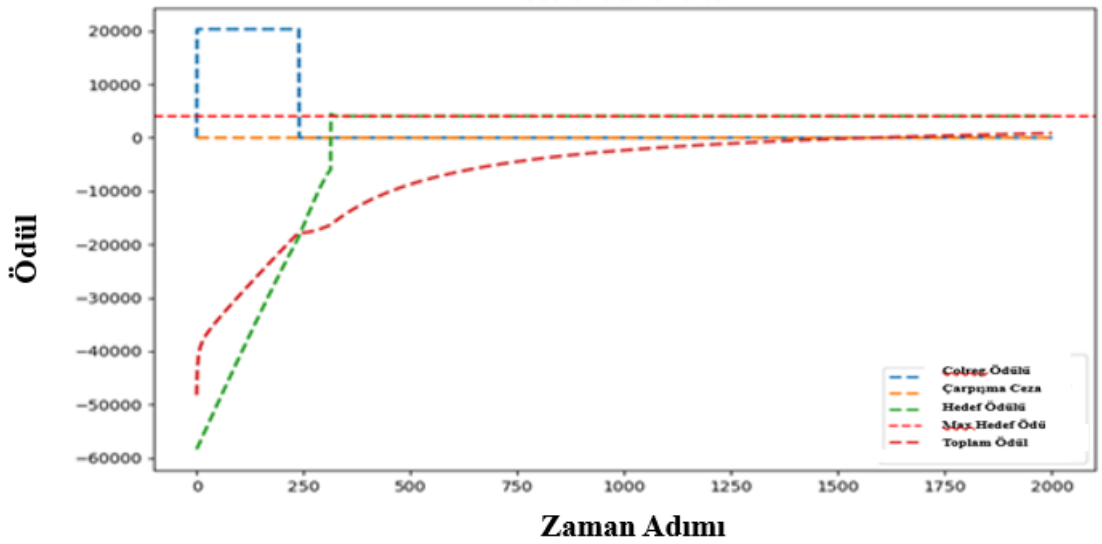


(a) Çarpışma öncesi (b) Çarpışma

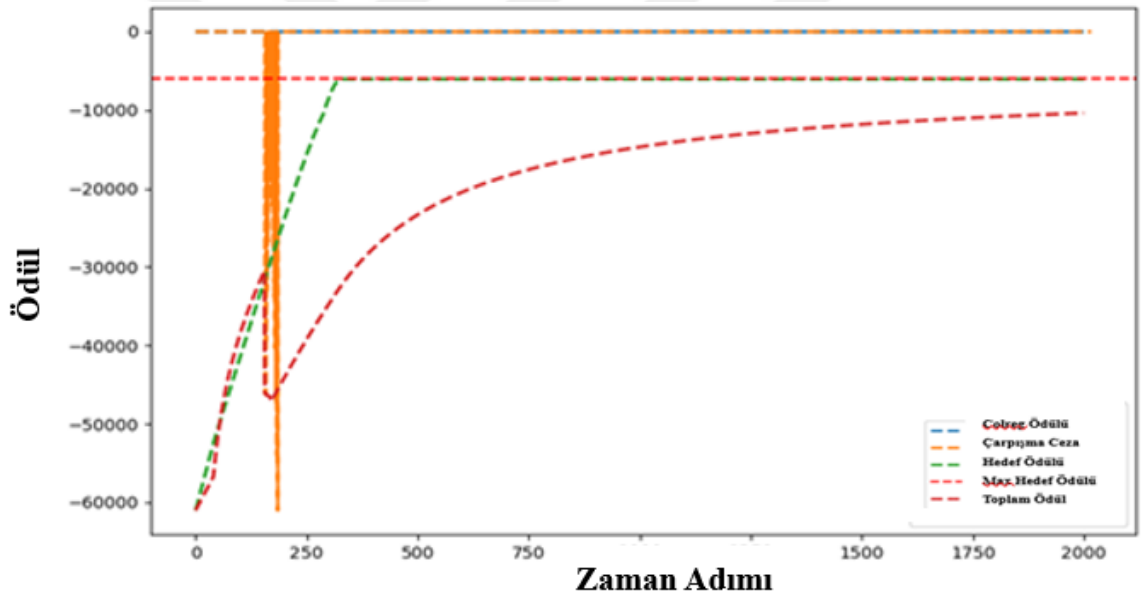
(c) Kaçınma, hedefe ulaşamama (d) Kaçınma, hedefe ulaşma

Şekil 4.8. Yol Verme senaryosunda gözlemlenen olası senaryolar

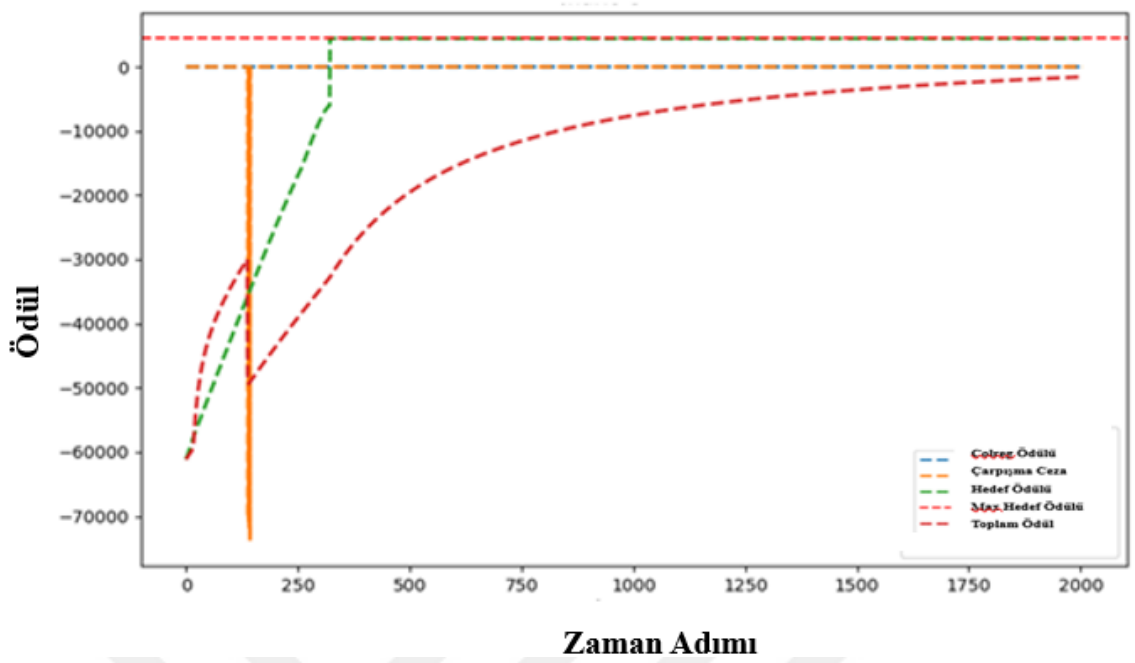
Bu senaryo altında pekiştirmeli öğrenmeye ait çarpışma, kaçınma, hedefe ulaşma ve toplam ödül fonksiyonu durumları Şekil 4.9'da gösterilmiştir.



(a) Kaçınma, hedefe ulaşma



(b) Çarpışma



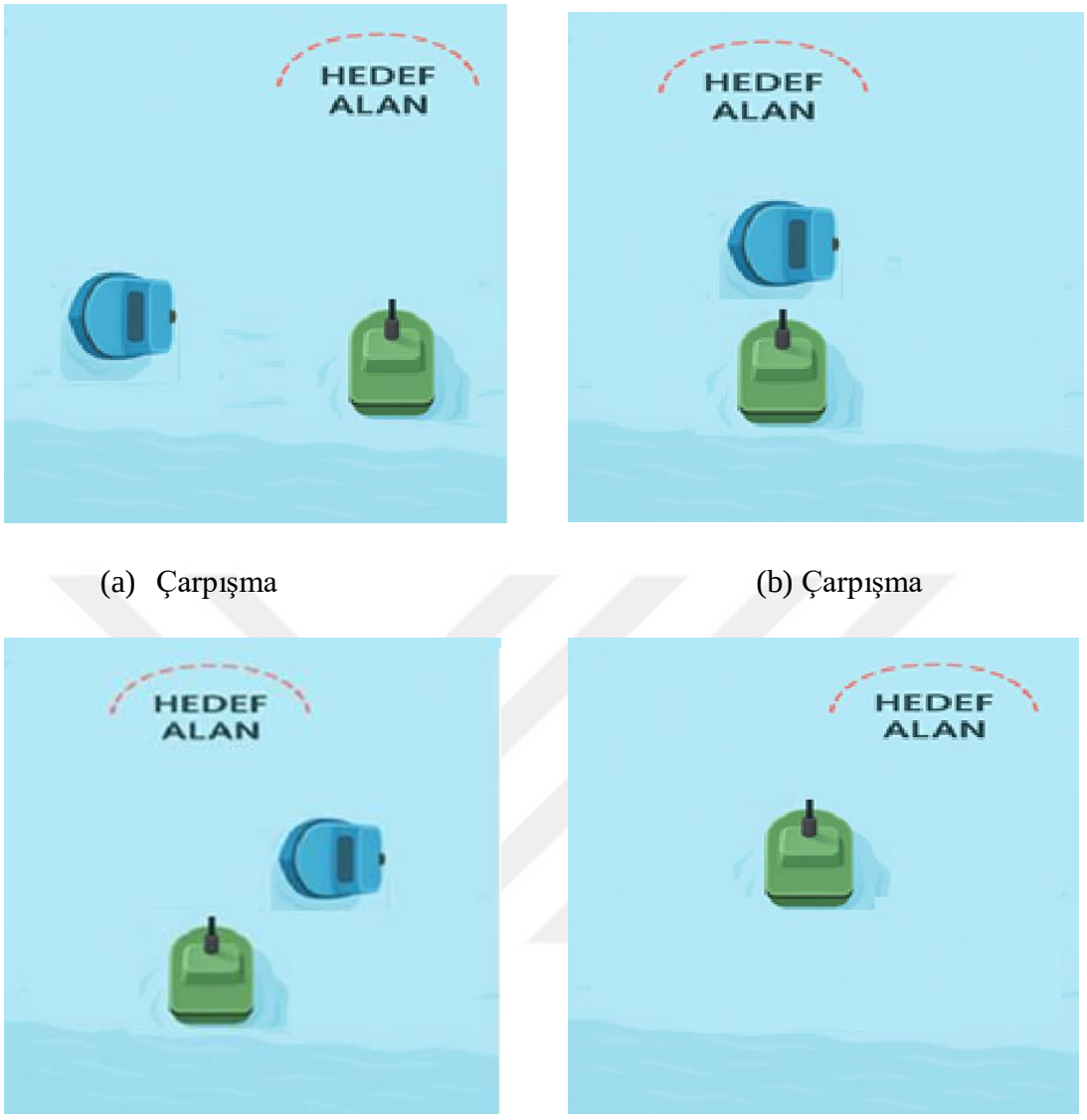
(c) Küçük çarpışma

Şekil 4.9. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri

Şekil 4.9'da yol verme senaryosuna ait ödül/ceza grafikleri yer almaktadır. Şekil 4.9 (a) senaryosunda, USV COLREG hareketini uygulamış, araçtan başarılı bir şekilde kaçınmış ve hedef konuma yakın bir noktaya ulaşmıştır. Bu nedenle, toplam ödül grafiği pozitif yönde ilerlemiştir. Şekil 4.9 (b) senaryosunda ise, büyük bir çarpışma gözlemlenmiş ve hedef konuma yaklaşılmasına rağmen, toplam ödül grafiği 0'ın altında kalmıştır. Şekil 4.9 (c) senaryosunda, küçük bir çarpışma meydana gelmiş olsa da toplam ödül grafiği yine 0'ın altında kalmış ve negatif yönde ilerlemiştir.

4.6. Karşıdan Karşıya Geçme Senaryosu

Karşıdan karşıya geçme senaryosunda sol taraftan gelen USV'ye göre hareketleri gözlemlenmiştir. Gözlemlere ait görseller Şekil 4.10'da gösterilmiştir.



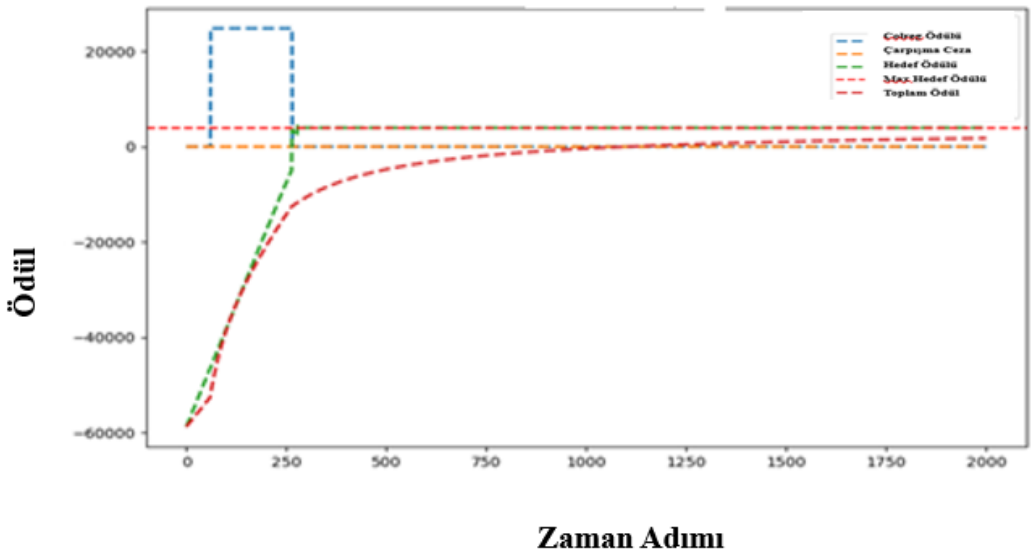
(a) Çarpışma (b) Çarpışma

(c) Kaçınma, hedefe ulaşma (d) Kaçınma, hedefe yakın ulaşma

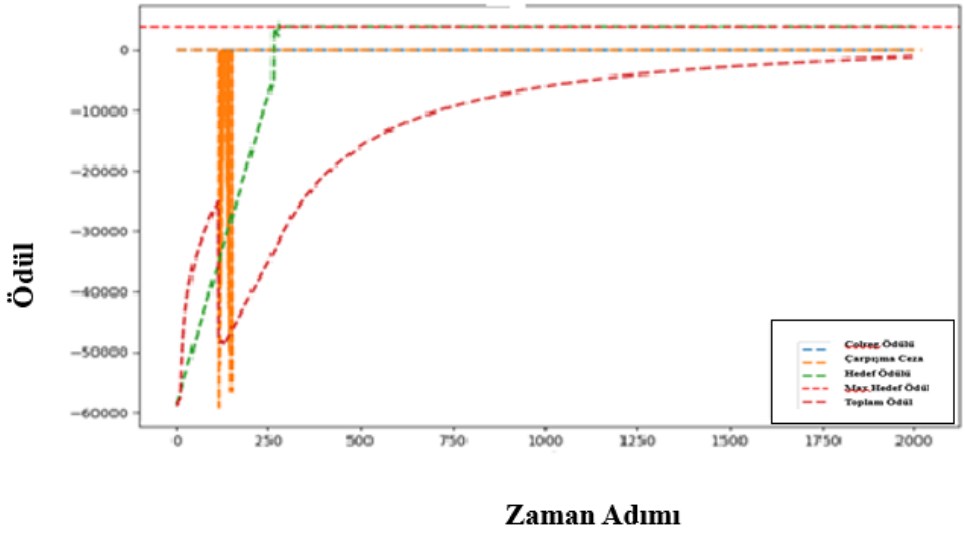
Şekil 4.10. Karşıdan karşıya geçen senaryosunda gözlemlenen olası senaryolar

Şekil 4.10'da görüldüğü üzere, bu senaryoda ana USV kaçınma hareketini gerçekleştirmemiştir. Diğer taraftan gelen USV'den kaçınma hareketi beklenmektedir. Şekil 4.10 (a) senaryosunda, diğer USV kaçınma manevrasını yapmamış ve çarpışma olasılığı gözlemlenmiştir. Şekil 4.10 (b) senaryosunda, kaçınma manevrası yapılmamış ve büyük bir çarpışma meydana gelmiştir. Şekil 4.10 (c) senaryosunda ise, kaçınma hareketi başarılı bir şekilde gerçekleştirilmiş ve ana USV hedef noktaya doğru ilerlemektedir. Şekil 4.10 (d) senaryosunda ise, kaçınma hareketi yapılmış ve USV hedef noktasına oldukça yakın bir konumda bulunmuştur. Simülasyon esnasında örnek olarak ***dalga hızı 5 m/s, deniz akıntısı 0.5 m/s, dalga açısı 0.6 radyan*** alınmıştır.

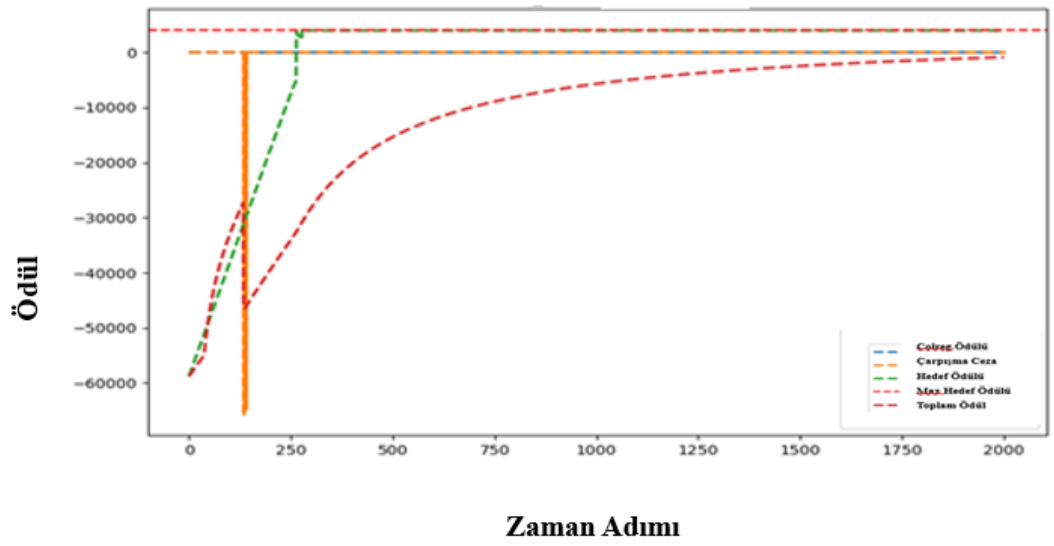
Bu senaryo altında pekiştirmeli öğrenmeye ait çarpışma, kaçınma, hedefe ulaşma ve toplam ödül fonksiyonu durumları Şekil 4.11'de gösterilmiştir.



(a) Kaçınma, hedefe ulaşma



(b) Çarpışma



(c) Çarpışma

Şekil 4.11. Senaryolar sonucu elde edilen ödül/çarpışma grafikleri

Şekil 4.11 (a)'da görüldüğü üzere, diğer USV ana USV'den kaçınma hareketini başarıyla gerçekleştirmiş ve ana USV hedefe doğru ilerlemiştir. Bu nedenle, toplam ödül grafiği pozitif yönde bir eğilim göstermiş ve değer 0'ın üzerine çıkmıştır. Şekil 4.11 (b)'de, büyük bir çarpışma meydana gelmiş ve toplam ödül grafiği 0'ın altında kalmış, böylece negatif bir eğilim gözlemlenmiştir. Şekil 4.11 (c)'de ise, küçük bir çarpışma olmasına rağmen, toplam ödül grafiği yine 0'ın altında kalmış ve negatif yönde bir ilerleme göstermiştir.

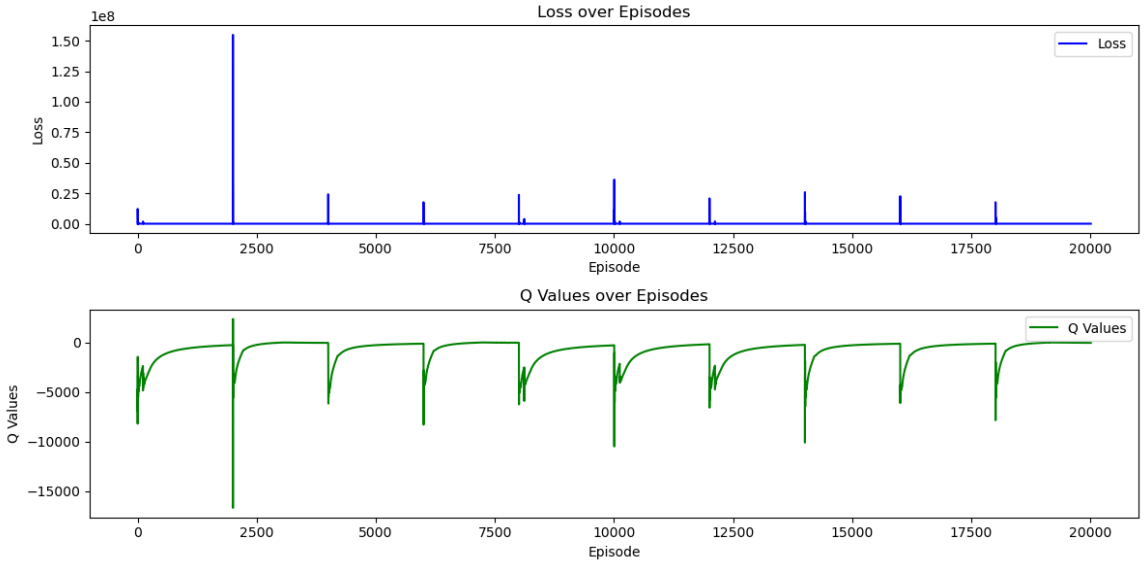
4.7. Q ve Kayıp Değerleri

Çalışmada, her bir COLREG kuralı için senaryo bazında farklı öğrenme modelleri çalıştırılmıştır. Senaryo bazında daha fazla veri elde etmek, modelin eğitimi sırasında önemli bir avantaj sağlamaktadır. Pekiştirmeli öğrenme sürecinde kayıp fonksiyonu genellikle her eğitim adımında hesaplanır. Her adımda ajan, bir aksiyon alır, ortamdan bir gözlem alır ve Q-değerlerini güncelleyerek kayıp fonksiyonunu hesaplar. Bu kayıp, ajanın hareket ettiği her durumda ağın ne kadar iyi öğrendiğini gösterir.

Kayıp fonksiyonu grafikleri, ajanın performansını değerlendirmek ve eğitim sürecindeki gelişmeleri izlemek için önemli bir araçtır. Eğitim ilerledikçe kayıp genellikle azalırken, ajanın performansı artar. Ancak, eğitim sürecinde bazı dalgalanmalar

gözlemlenebilir, çünkü ajan yeni durumlarla karşılaştıkça ve daha iyi bir strateji geliştirdikçe kayıp fonksiyonu değişir. Bu nedenle, kayıp fonksiyonu grafiği genellikle düzgün bir şekilde azalan bir eğilim göstermelidir, ancak keskin düşüşler veya ani artışlar da gözlemlenebilir.

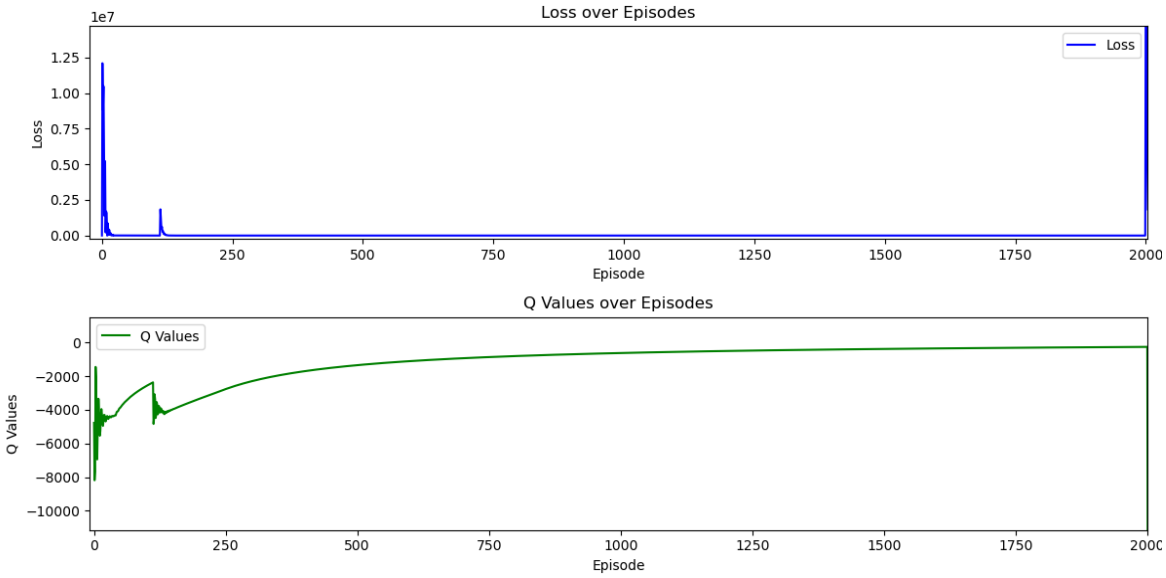
Ajanın eğitimi, toplam ödül değişimi ile ilişkili olarak incelenmiş ve 10 farklı senaryonun farklı bölümleri üzerinde 20.000 zaman dilimi (episode) boyunca gerçekleştirilmiştir. Şekil 4.12'de elde edilen grafik yer almaktadır. Eğitimde kullanılan parametreler öğrenme hızı 0.001 ve katman yapısı [128:128] olarak belirlenmiştir. Bu parametreler doğrulama çalışmalarıyla elde edilmiştir.



Şekil 4.12. 20000 zaman dilimi eğitilen ajan Q/Kayıp Grafiği

Şekil 4.12'de ajan 10 farklı senaryo üzerinde 20.000 zaman dilimi boyunca kafa kafaya senaryosunda veri toplanarak eğitilmiştir. Eğitilen modelde, öğrenme süreci 2. senaryo sonrasında kayıp değerlerinin büyük ölçüde azaldığı gözlemlenmiştir. Mevcut Q-değerlerinde bir artış gözlemlense de bu artışın pozitif bir değere ulaşmadığı belirtilmiştir.

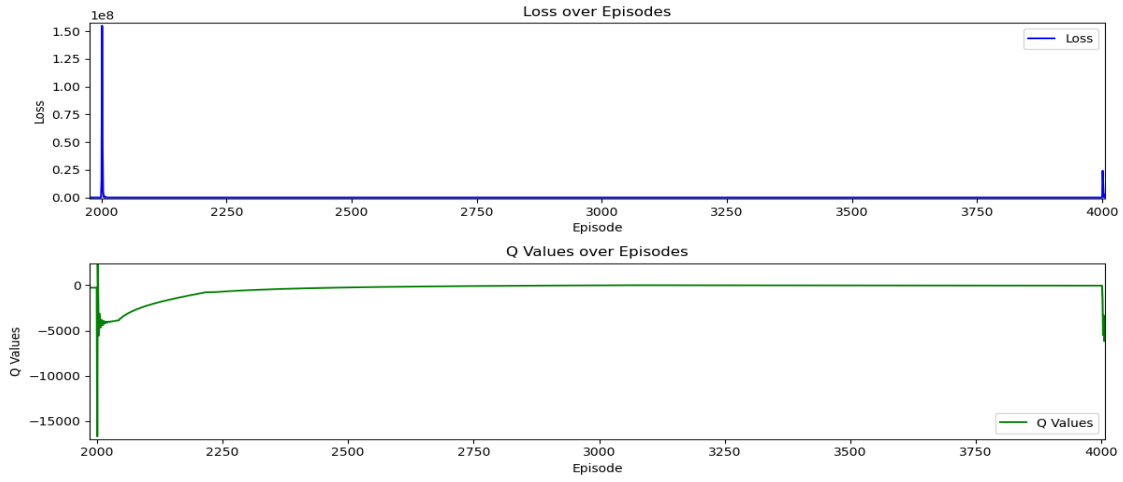
Şekil 4.13'te 20.000 zaman diliminde eğitilen ajanın 1. senaryo 0-2000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.13. 1. senaryo 0-2000 zaman dilimi

Şekil 4.13'te gerçekleşen senaryoda çarpışma gözlenmiştir. Hem kayıp hem de Q değerlerinin düşük seviyelerde seyretmesi, modelin bu senaryoda öğrenme performansının yetersiz kaldığını göstermektedir. Özellikle Q değerinin düşük olması, ajanın gelecekteki ödülleri yeterince değerlendiremediğini; düşük kayıp değeri ise bu yetersiz politikanın kararlı bir şekilde öğrenildiğini göstermektedir.

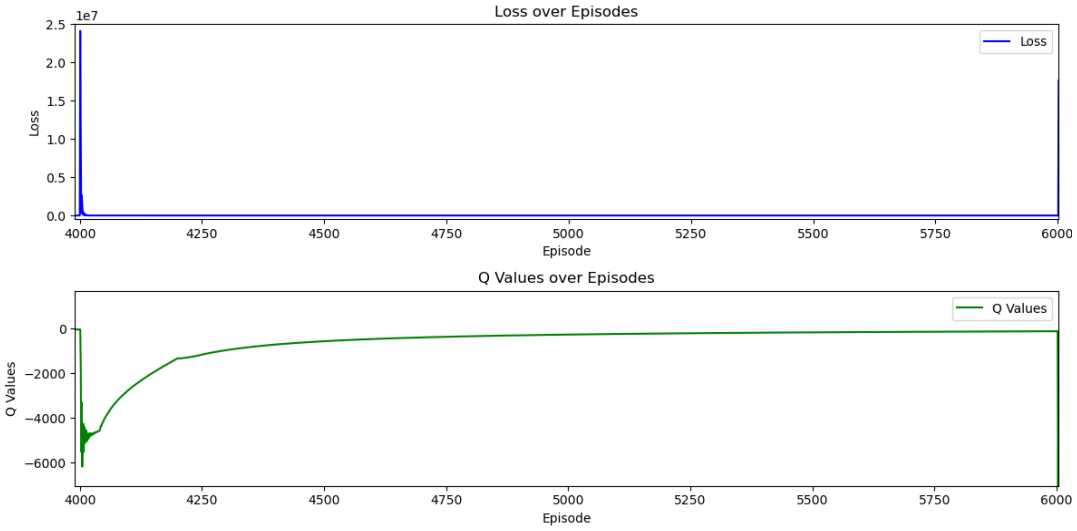
Şekil 4.14'te 20.000 zaman diliminde eğitilen ajanın 2. senaryo 2000-4000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.14. 2. senaryo 2000-4000 zaman dilimi

Şekil 4.14'te 2000–4000 zaman dilimleri arasındaki eğitim sürecinde, kayıp değerlerinin büyük oranda sıfıra yakın seyretmesi, modelin bu periyotta kararlı bir şekilde öğrendiğini göstermektedir. Q değerlerinin zamanla yükselmesi de ajanın daha başarılı eylemler geliştirdiğini işaret etmektedir. Ancak 4000 zaman diliminde her iki metrikte de gözlemlenen ani düşüş, ilgili senaryoda bir çarpışma gerçekleşmiş olabileceğini düşündürmektedir.

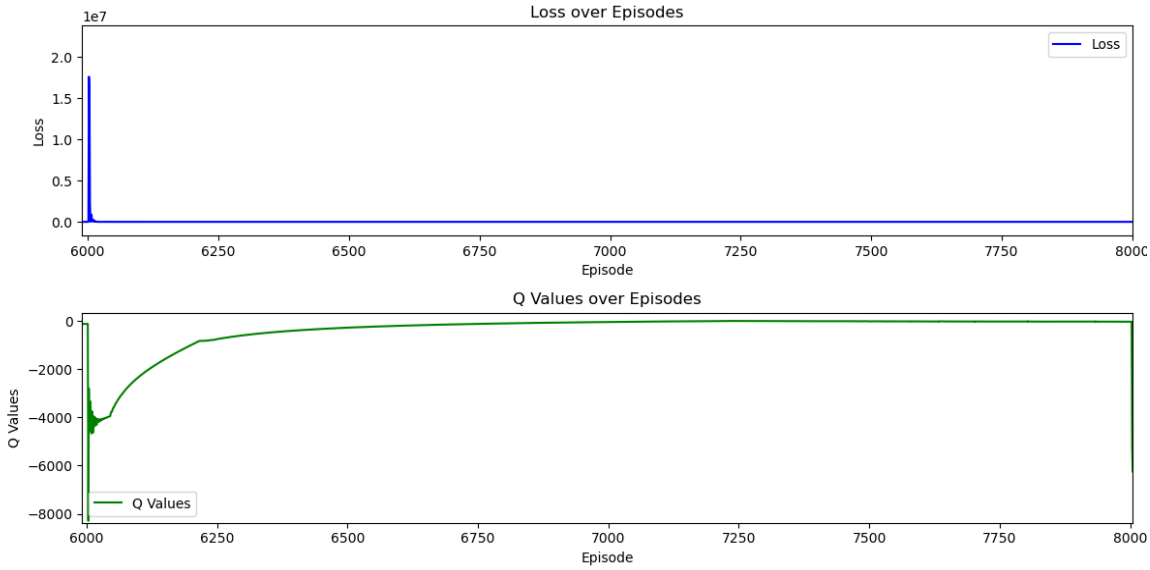
Şekil 4.15'te 20.000 zaman diliminde eğitilen ajanın 3. senaryo 4000-6000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.15. 3. senaryo 4000-6000 zaman dilimi

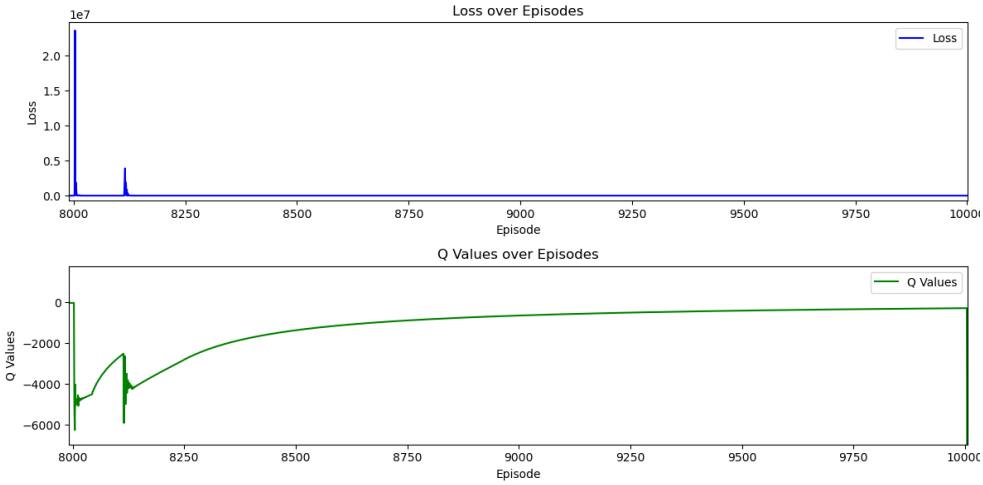
Şekil 4.15'te sunulan 4000–6000 zaman dilimi aralığında, eğitim sürecinin daha stabil ve verimli ilerlediği görülmektedir. Kayıp değerlerinin kısa sürede düşerek sabitlenmesi ve Q değerlerinin istikrarlı şekilde artış göstermesi, bu periyotta elde edilen politikanın önceki dönemlere kıyasla daha kararlı ve başarılı olduğunu ortaya koymaktadır. Öğrenme süreci hem hata açısından hem de eylem-ödül ilişkisinin güçlenmesi bakımından olumlu bir yönde ilerlemiştir.

Şekil 4.16'da 20.000 zaman diliminde eğitilen ajanın 4. senaryo 6000-8000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.16. 4. senaryo 6000-8000 zaman dilimi

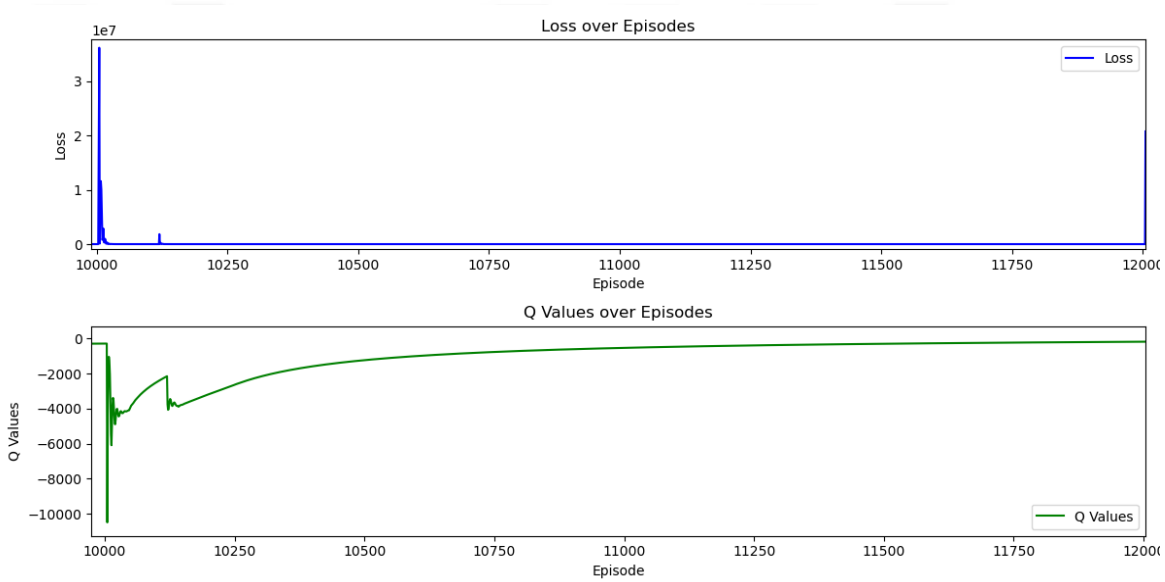
Şekil 4.16'da 6000–8000 zaman dilimi arasındaki eğitim sürecinde, modelin istikrarlı bir şekilde öğrenmeye devam ettiği gözlemlenmiştir. Kayıp değerleri kısa sürede düşerek sıfıra yakın seyretmiş ve bu durum, modelin Q değerlerini hedef değerlere yakın şekilde üretebildiğini göstermiştir. Ayrıca, Q değerlerinde meydana gelen istikrarlı artış, ajan politikasının ödül beklentisini güçlendirdiğini ve karar verme süreçlerinin olumlu yönde ilerlediğini ortaya koymaktadır. Şekil 4.17'de 20.000 zaman diliminde eğitilen ajanın 5. senaryo 8000-10000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.17. 5. senaryo 8000-10000 zaman dilimi

Şekil 4.17'de 8000–10000 zaman dilimi aralığında Q değeri ciddi bir şekilde azalmış, çarpışma meydana gelmiştir. Öğrenme sürecinde geçici bir kararsızlık yaşandığı görülmektedir. Başlangıçta ve 8200 zaman dilimine yakın dönemde kayıp değerlerinde gözlemlenen ani sıçramalar, modelin kısa süreli öğrenme dengesizliği yaşadığını göstermektedir. Bu sıçramalara paralel olarak, Q değerlerinde de belirgin dalgalanmalar gözlemlenmiş; ancak bu periyodun sonlarına doğru değerler toparlanma eğilimi göstermiştir.

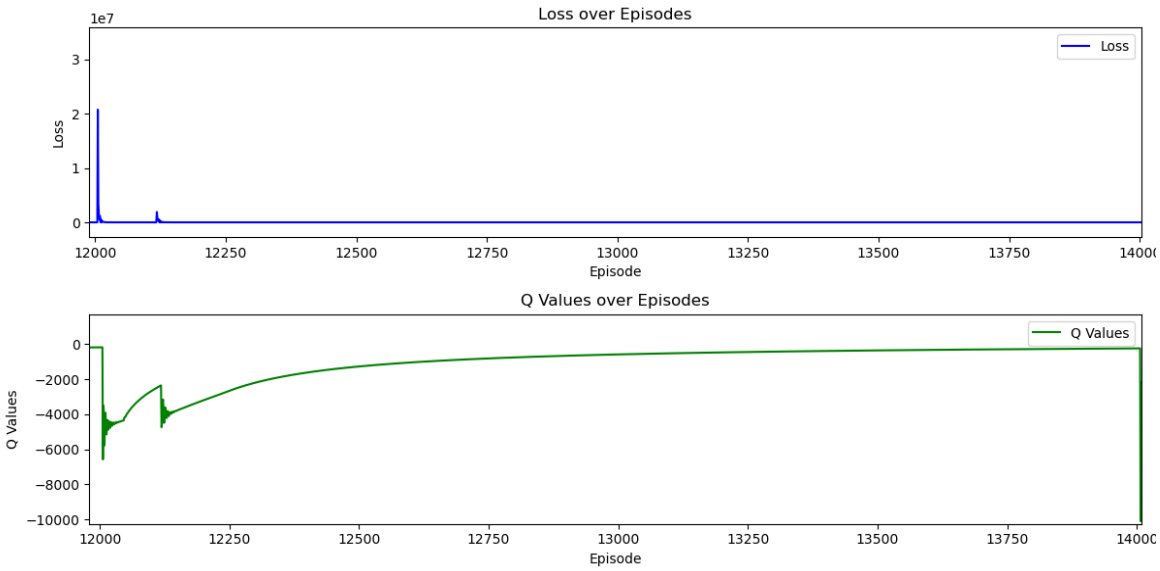
Şekil 4.18'te 20.000 zaman diliminde eğitilen ajanın 6. senaryo 10000-12000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.18. 6. senaryo 10000-12000 zaman dilimi

Şekil 4.18'te gösterilen 10000–12000 zaman dilimi aralığında, öğrenme süreci önceki dönemlerle benzer şekilde yüksek başlangıç kayıplarıyla başlamış; ancak bu kayıplar hızlı bir şekilde düşerek sürecin büyük kısmında sıfıra yakın seyretmiştir. Q değerlerindeki ani dalgalanma, 10200 zaman dilimi civarında yaşanan kısa süreli bir öğrenme bozulmasını işaret etmektedir. Ancak ardından gelen istikrarlı artış, modelin politikasını yeniden düzenlediğini ve ödül beklentisini önemli ölçüde iyileştirdiğini göstermektedir. Bu bağlamda, bu dönem modelin kararlılığını koruduğu ve öğrenmeyi başarıyla sürdürdüğü bir evre olarak değerlendirilmiştir.

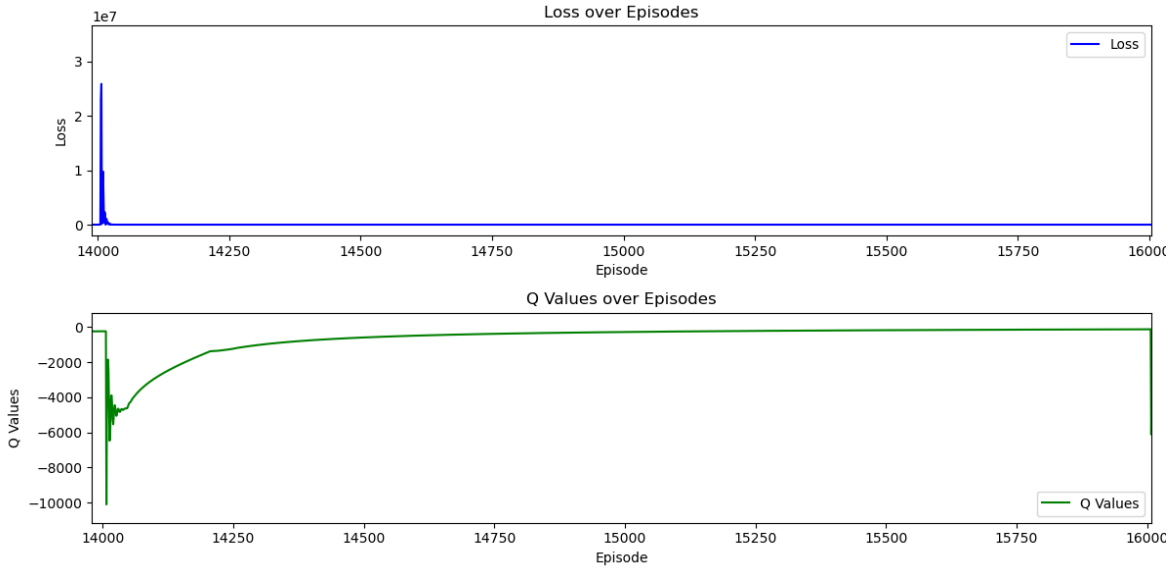
Şekil 4.19’da 20.000 zaman diliminde eğitilen ajanın 7. senaryo 12000-14000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.19. 7. senaryo 12000-14000 zaman dilimi

Şekil 4.19’da yer alan 12000–14000 zaman dilimi aralığında, modelin öğrenme sürecine yüksek kayıplarla başladığı; ancak kısa süre içinde bu değerlerin sabitlenerek sıfıra yakınlaştığı görülmüştür. Bu durum, öğrenmenin erken aşamalarında yaşanan kararsızlıkların ardından modelin tekrar dengeye ulaştığını göstermektedir. Q değerlerindeki geçici dalgalanma, kısa süreli bir performans düşüşü yaşandığını işaret etmektedir. Ancak bu dalgalanmayı takip eden yükseliş, ajan politikasının ödül beklentisini güçlendirdiğini ve sürecin verimli şekilde devam ettiğini ortaya koymaktadır.

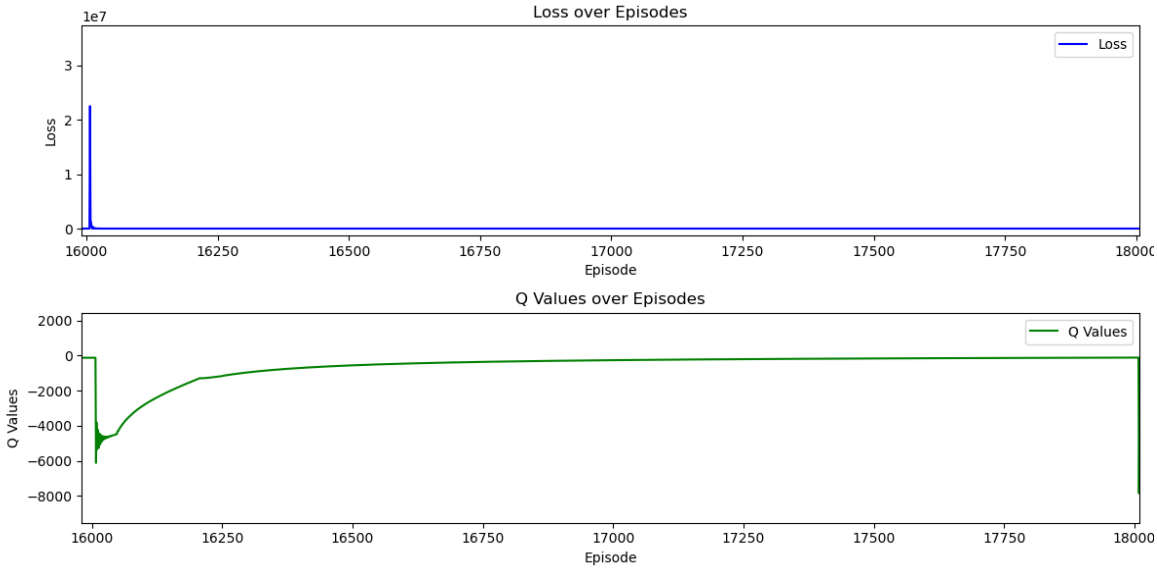
Şekil 4.20’de 20.000 zaman diliminde eğitilen ajanın 8. senaryo 14000-16000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.20. 8. senaryo 14000-16000 zaman dilimi

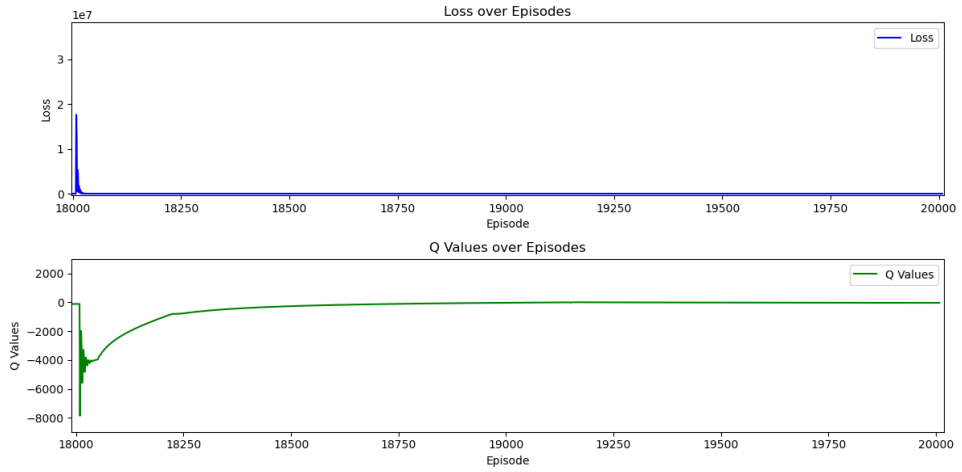
Şekil 4.20'de sunulan 14000–16000 zaman dilimi aralığında, modelin oldukça dengeli ve olgun bir öğrenme davranışı sergilediği gözlemlenmiştir. Başlangıçta gözlenen yüksek kayıp değerleri kısa sürede bastırılmış ve süreç boyunca yeni bir sıçrama yaşanmamıştır. Q değerlerinin istikrarlı artışı, ajanın bu dönemde daha ödüllü ve tutarlı eylemler seçtiğini göstermektedir. Bu dönem, modelin önceki öğrenmelerinden maksimum fayda sağlayarak kararlı bir politika geliştirdiği ve çarpışma olasılıklarını minimize ettiği bir evre olarak değerlendirilebilir.

Şekil 4.21'de 20.000 zaman diliminde eğitilen ajanın 9. senaryo 16000-18000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.21. 9.senaryo 16000-18000 zaman dilimi

Şekil 4.21'de 16000–18000 zaman dilimi aralığında, modelin öğrenme sürecinin oldukça olgunlaştığı görülmektedir. Başlangıçta gözlenen kısa süreli kayıp artışı dışında süreç boyunca modelin kararlı bir şekilde öğrenmeye devam ettiği gözlemlenmiştir. Q değerlerinin istikrarlı yükselişi, ajan politikasının yüksek ödüllü eylemleri tercih etme kapasitesini artırdığını göstermektedir. Bu dönem, modelin optimuma yakın bir politika geliştirdiği ve pekiştirmeli öğrenmenin başarıyla tamamlandığı evrelerden biri olarak değerlendirilebilir. Şekil 4.22'de 20.000 zaman diliminde eğitilen ajanın 10. senaryo 18000-20000 zaman diliminde kayıp ve Q grafiklerinin sonuçları yer almaktadır.



Şekil 4.22. 10. senaryo 18000-20000 zaman dilimi

Şekil 4.22'de sunulan 18000–20000 zaman dilimi aralığı, modelin öğrenme sürecinde en yüksek kararlılığa ulaştığı dönemi temsil etmektedir. Başlangıçta kısa süreli bir kayıp sıçraması gözlemlense de bu durum hızla bastırılmış ve süreç boyunca kayıp değerleri sıfıra yakın kalmıştır. Q değerlerindeki düzenli yükseliş, ajanın karar politikalarının ödül beklentisine uygun şekilde şekillendiğini ve öğrenmenin tamamlandığını göstermektedir. Bu dönem, eğitim sürecinin başarıyla tamamlandığı görülmektedir.

Kafa kafaya senaryosunda eğitilen ajanın model sonuçları Çizelge 4.2'de verilmiştir.

Çizelge 4.2. Kafa kafaya senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları

Senaryo	**Q Değer**	**Kayıp Değer**
1	-265.976	0.000
2	-44.129	0.000
3	-122.918	0.000
4	-32.125	0.000
5	-289.320	0.000
6	-181.999	0.002
7	-241.344	0.000
8	-127.528	0.006
9	-119.428	0.001
10	-37.944	0.016
11	-100.000	0.500
12	-72.222	0.446
13	-44.444	0.391
14	-16.667	0.337
15	11.111	0.282
16	38.889	0.228
17	66.667	0.173
18	94.444	0.119
19	122.222	0.064
20	150.000	0.010

Çizelge 4.2'de Q ve kayıp değerleri incelendiğinde, modelin genel öğrenme sürecinde ilerlemeler kaydedildiği gözlemlenmektedir. Q değerlerinde belirgin bir iyileşme sağlanmış; kayıp değerleri ise çoğu senaryoda oldukça düşük seviyelere gerilemiştir. Ancak elde edilen bu gelişmeler, modelin istenen kararlılığa tam olarak ulaşamadığını göstermektedir.

Senaryoların zaman dilimlerine göre eşleştirilmesiyle, özellikle 1. (0–2000), 5. (8000–1000), 6. (10000–12000) ve 7. (12000–14000) senaryolarda, Q değerlerinin oldukça düşük olduğu ve bazı durumlarda kayıp değerlerinin yükseldiği tespit edilmiştir. Bu durum, söz konusu zaman dilimlerinde çarpışmaların meydana geldiğini ve modelin kararlarında tutarsızlıklar yaşadığını düşündürmektedir. 4. senaryo (6000–8000) ve 10. senaryo (18000–20000) gibi dönemlerde daha yüksek Q değerleri ve kabul edilebilir kayıp seviyeleri elde edilmiştir. Bu senaryolarda, modelin daha etkili şekilde öğrenme gerçekleştirdiği ve eylem seçimlerinde daha başarılı olduğu değerlendirilmektedir. Genel olarak bakıldığında, öğrenme süreci 0–2000 zaman diliminden sonra başlamış, 6000–7500 aralığında modelin davranışlarında olumlu yönde gelişmeler gözlemlenmiş, 18000–20000 aralığında ise sistemin daha kararlı ve verimli çıktılar üretmeye başladığı görülmektedir.

Modelin öğrenme süreci boyunca sergilediği gelişimi açık biçimde ortaya koymaktadır. İlk 10 senaryoda negatif Q değerlerinin hâkim olduğu ve bazı durumlarda kayıp değerlerinin görece yüksek seviyelere ulaştığı gözlemlenmektedir. Bu durum, modelin başlangıç aşamasında çarpışma riski taşıyan kararlar verdiğini ve kararlı bir politika oluşturmakta zorlandığını göstermektedir. Ancak 11. senaryodan itibaren Q değerlerinin kademeli olarak artması ve kayıp değerlerinin belirgin şekilde azalması, modelin çevreyle olan etkileşimlerinden öğrenerek daha ödüllü eylemler üretmeye başladığını ve karar kalitesini iyileştirdiğini göstermektedir. Özellikle 18–20. senaryolar arasında pozitif Q değerlerinin görülmesi ve kayıpların minimum düzeyde seyretmesi, modelin öğrenme sürecinin sonunda kararlı, uygulanabilir ve güvenilir yörüngeler geliştirdiğini göstermektedir. Bu bulgular, öğrenmenin ilerleyen senaryolarda başarıyla gerçekleştiğini ve modelin amacına uygun çıktılar ürettiğini güçlü desteklemektedir.

Arkadan gelerek sollama senaryosunda eğitilen ajanın model sonuçları Çizelge 4.3'te verilmiştir.

Çizelge 4.3. Arkadan sollama senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları

Senaryo	**Q Değer**	**Kayıp Değer**
1	92.566	0.024
2	251.139	0.054
3	140.493	0.060
4	265.668	0.088
5	59.582	0.034
6	227.394	0.007
7	48.541	0.021
8	61.550	0.028
9	87.222	0.018
10	66.486	0.024

Çizelge 4.3'te model sonuçlarına bakılırsa modelin tüm senaryolar boyunca kararlı ve başarılı bir öğrenme süreci gerçekleştirdiğini göstermektedir. Q değerlerinin tamamının pozitif aralıkta yer alması, ajanın eylemler sonucunda ödül beklentisinin yüksek olduğunu ve çevreyle etkileşimlerinde olumlu sonuçlar ürettiğini göstermektedir. Ayrıca, kayıp değerlerinin tüm senaryolarda oldukça düşük seviyelerde kalması, modelin hedef Q değerlerine yakın tahminlerde bulunduğunu ve öğrenmenin başarıyla tamamlandığını ortaya koymaktadır. Bu sonuçlar, modelin hem çarpışma riskini minimize ettiğini hem de güvenilir yörünge planlaması yapabildiğini göstermekte; bu doğrultuda, ek senaryo ihtiyacı olmaksızın sistemin uygulanabilir bir sonuç verdiği görülmektedir.

Yol veren çarpışma senaryosunda eğitilen ajanın model sonuçları Çizelge 4.4'te verilmiştir.

Çizelge 4.4. Yol veren çarpışma senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları

Senaryo	Q Değer	Kayıp Değer
1	-133.047	0.037
2	-573.658	5.960
3	-98.788	0.000
4	-580.704	2.384
5	-80.472	0.100
6	-135.046	0.001
7	176.323	0.087
8	-88.657	0.009
9	-93.755	0.321
10	-181.931	0.071
11	-100.000	0.500
12	-84.211	0.474
13	-68.421	0.448
14	-52.632	0.423
15	-36.842	0.397

Senaryo	Q Değer	Kayıp Değer
16	-21.053	0.371
17	-5.263	0.345
18	10.526	0.319
19	26.316	0.294
20	42.105	0.268
21	57.895	0.242
22	73.684	0.216
23	89.474	0.191
24	105.263	0.165
25	121.053	0.139
26	136.842	0.113
27	152.632	0.087
28	168.421	0.062
29	184.211	0.036
30	200.000	0.010

Çizelge 4.4'te model sonuçlarına bakılırsa Q değerleri 10 senaryoda pozitif gelişme gösterememiştir. Modelin öğrenme sürecinde zamanla belirgin bir gelişim gösterdiği gözlemlenmektedir. İlk 10 senaryoda ağırlıklı olarak negatif Q değerleri ve görece yüksek kayıp oranları dikkat çekerken, bu durum modelin erken dönemlerde çarpışma riski taşıyan eylemler ürettiğini ve kararsız bir politika izlediğini göstermektedir. Ancak 11. senaryodan itibaren, Q değerlerinde kademeli bir artış ve kayıp değerlerinde düzenli bir azalma gözlemlenmiş; bu da modelin çevreden aldığı geri bildirimlere uyum sağlayarak daha başarılı stratejiler geliştirmeye başladığını ortaya koymuştur. Özellikle 25–30. senaryolar arasında Q değerlerinin pozitif seviyelere ulaşması ve kayıpların oldukça düşük düzeylere inmesi, modelin kararlı ve uygulanabilir yörüngeler üretme yeteneğini önemli ölçüde geliştirdiğini göstermektedir. Bu sonuçlar, modelin öğrenme sürecinde zaman ilerledikçe başarıyla adapte olduğunu ve eğitim sürecinin sonunda istenen davranışları sergilemeye başladığını güçlü biçimde desteklemektedir.

Karşıdan karşıya geçme senaryosunda eğitilen ajanın model sonuçları Çizelge 4.5'te verilmiştir.

Çizelge 4.5. Karşıya geçme senaryosunda eğitilen ajanın Q değer ve kayıp sonuçları

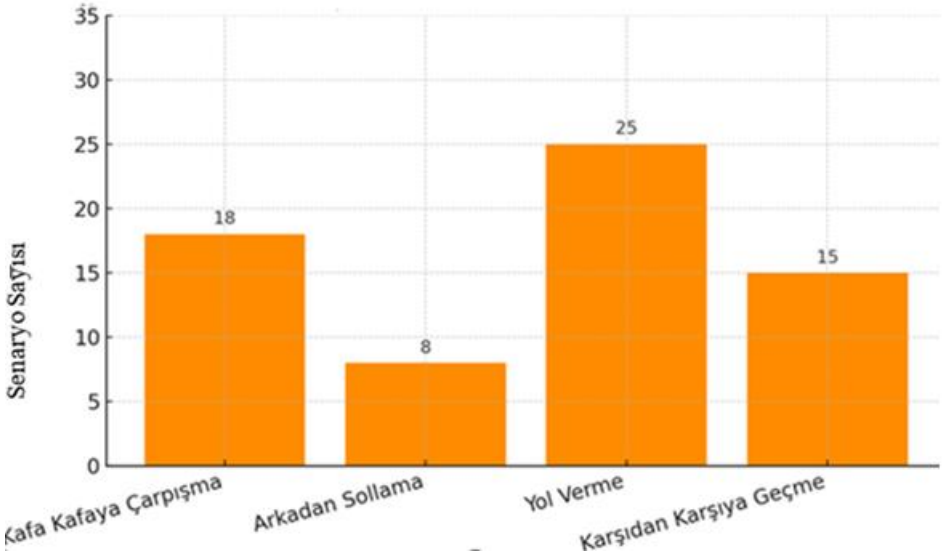
Senaryo	Q Değer	Kayıp Değer
1	434.036	0.003
2	-79.858	0.001
3	445.507	0.010
4	413.444	0.401
5	-97.671	0.005
6	-89.742	0.052
7	-92.141	0.048
8	-97.399	0.072
9	-80.329	0.235
10	-78.926	0.265
11	-80.000	0.300
12	-48.889	0.268
13	-17.778	0.236
14	13.333	0.203
15	44.444	0.171
16	75.556	0.139
17	106.667	0.107
18	137.778	0.074
19	168.889	0.042
20	200.000	0.010

Çizelge 4.5'te model sonuçlarına bakılırsa modelin öğrenme sürecinin özellikle 15. senaryodan itibaren belirgin şekilde iyileşmeye başladığı gözlemlenmektedir. İlk 10 senaryoda modelin performansı büyük ölçüde dengesizdir; bu dönemde negatif Q değerleri ve değişken kayıp oranları, ajanın ödül beklentisinin düşük olduğunu ve öğrenme sürecinde kararlılık sağlanamadığını göstermektedir. 11–14. senaryolar arasında kademeli bir geçiş süreci yaşanmış; 15. senaryodan itibaren ise Q değerleri pozitif yönde yükselmiş ve kayıp değerleri düzenli şekilde azalmıştır. Bu durum, modelin belirli bir öğrenme eşiğini aştığını ve çevresel geri bildirimlere daha etkili yanıtlar üretmeye başladığını ortaya koymaktadır. Sonuç olarak, modelin başlangıçta sergilediği kararsız davranışların zaman içinde istikrarlı politikalara dönüştüğü ve öğrenmenin başarıyla gerçekleştiği senaryo 15'ten itibaren net biçimde gözlemlenmiştir.

Dört farklı senaryo tipi karşılaştırıldığında, arkadan gelerek sollama senaryosunda modelin öğrenme sürecinin daha erken ve etkili biçimde başladığı, bu senaryoda pozitif Q-değerlerinin daha yüksek seviyelere ulaştığı gözlemlenmiştir. Bu durum, modelin bu tip çevresel yapıya daha hızlı adapte olduğunu ve başarılı eylem seçimleri gerçekleştirebildiğini göstermektedir. Buna karşın, diğer senaryolarda öğrenme süreci daha geç başlamış; özellikle 15. veya 20. senaryodan sonra Q değerlerinde iyileşme ve kayıp değerlerinde azalma gözlemlenmiştir. 30. senaryolara yaklaşıldıkça modelin performansında anlamlı bir artış elde edilmiştir.

Modelin eğitimi sırasında, kullanılan pekiştirmeli öğrenme algoritmasının, çarpışma önleme ve hedefe ulaşma görevlerinde ne kadar etkili olduğu gözlemlenmiştir. Eğitim sürecinde, modelin Q-değerlerinin zaman içinde nasıl değiştiği izlenmiştir. Eğitim öncesi Q-değerlerinin oldukça düşük olduğu, ancak modelin çevresindeki ortamdan elde ettiği geri bildirimlerle bu değerlerin artmaya başladığı gözlemlenmiştir.

Elde edilen bir diğer önemli bulgu, modelin farklı çarpışma senaryolarındaki performansıdır. Bu çalışmada dört farklı çarpışma senaryosu (kafa kafaya çarpışma, arkadan sollama, yol verme, karşıdan karşıya geçme) üzerinde test edilmiştir. Her bir senaryoda modelin verdiği tepkiler ile bu senaryoları ne kadar sürede öğrenebildiği analiz edilmiştir. Sonuçlar, kafa kafaya çarpışma gibi daha belirgin ve doğrudan etkileşim gerektiren senaryolarda öğrenmenin daha hızlı gerçekleştiğini; buna karşın, karşıdan karşıya geçme gibi çevresel değişkenliği ve etkileşim yoğunluğu daha yüksek olan karmaşık senaryolarda öğrenme sürecinin daha uzun sürdüğünü Şekil 4.23'te görülmektedir. Bu bulgular, pekiştirmeli öğrenme tabanlı sistemlerin başarı düzeylerinin senaryonun yapısal karmaşıklığına ve çevresel faktörlere duyarlı olduğunu göstermektedir.



Şekil 4.23. Senaryo sayısına bağlı olarak öğrenme durumu

Şekil 4.23'te dört farklı çarpışma senaryosu için modelin öğrenmeye başladığı yaklaşık senaryo numaraları gösterilmektedir. Arkadan gelerek sollama senaryosunda öğrenme en erken 8. senaryo civarında başlamıştır. Karşıdan karşıya geçme senaryosunda öğrenme 15. senaryoda belirginleşmiştir. Kafa kafaya çarpışma senaryosunda ise ancak 18. senaryoya doğru kararlılık gözlemlenmiştir. En geç öğrenme ise yol verme senaryosunda gerçekleşmiş ve modelin başarılı çıktılar üretmesi 25. senaryoyu bulmuştur. Bu durum, çevresel etkileşim düzeylerinin ve senaryo karmaşıklığının öğrenme süreci üzerinde doğrudan etkili olduğunu göstermektedir.

Modelin başlangıç aşamasında yüksek kayıp değerlerine ve düşük ödül seviyelerine sahip olduğu, ancak eğitim süreci ilerledikçe hem kayıp fonksiyonunun azaldığı hem de ödül değerlerinin anlamlı şekilde arttığı gözlemlenmiştir. Erken dönemlerde kayıp fonksiyonunda görülen dalgalanmalar, modelin çevresel etkileşimlere uyum sağlama sürecinde zorluk yaşadığını ortaya koymaktadır. Ancak eğitim adımlarının ilerlemesiyle birlikte bu dalgalanmalar azalarak daha stabil bir yapıya dönüşmüş ve modelin karar üretiminde daha tutarlı hale geldiği anlaşılmıştır.

Pekiştirmeli Öğrenme Algoritması, çevresel etkileşimlerden öğrenme ve çarpışmalardan kaçınma konusunda yüksek başarı göstermiştir. Ortaya kurallarına uygun hareket etme yeteneği, modelin gerçek dünya denizcilik uygulamaları için uygun olduğunu

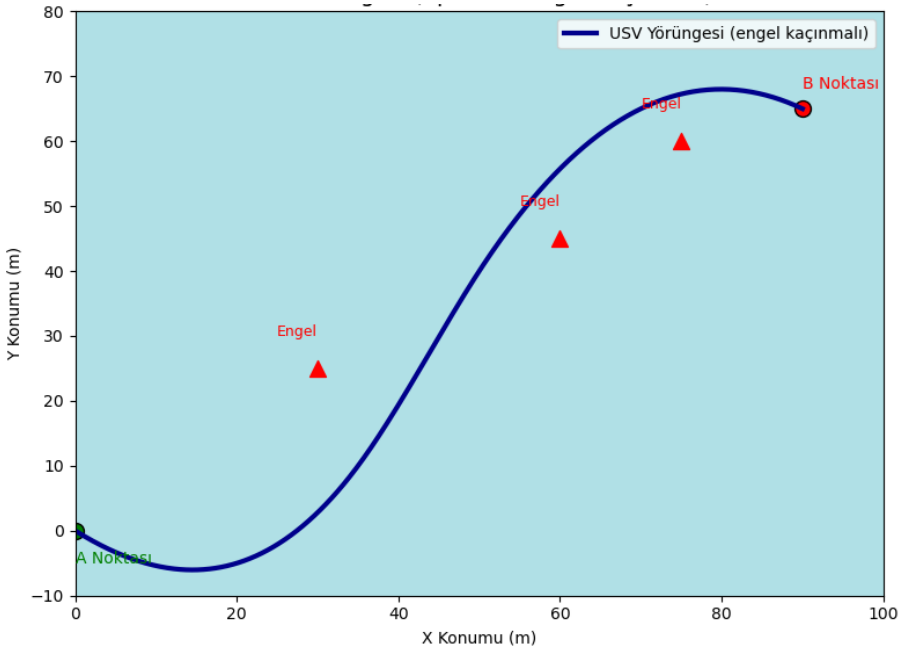
ortaya koymaktadır. Eğitim sürecinde kayıp fonksiyonundaki azalma, modelin çevresel koşullara uyum sağladığını göstermektedir.

Karmaşık ve dinamik senaryolarda modelin başarı oranı, daha basit senaryolara göre düşük kalmıştır. Bazı senaryolarda modelin çarpışmalardan tamamen kaçınamaması, daha gelişmiş eğitim süreçlerine ve strateji geliştirmeye ihtiyaç duyduğunu göstermektedir.

Bu çalışmada kullanılan pekiştirmeli öğrenme yapısında, ödül fonksiyonundaki katsayılar sistemin öğrenme davranışını doğrudan etkilemektedir. Özellikle çarpışmadan kaçınma durumuna verilen ceza, hedefe ulaşma ödülüne kıyasla daha yüksek belirlenmiştir. Bu tercih, USV'lerin güvenliğini önceliklendirmek adına yapılmıştır. Çarpışma cezasının yüksek olması, ajanı riskli manevralardan uzak tutarak güvenli ancak daha uzun yörüngeleri tercih etmesine neden olmuştur. Eğer ödül fonksiyonunda hedefe ulaşma ödülü ile çarpışma cezası birbirine yakın olsaydı, ajan daha agresif manevralar öğrenebilir ve çarpışma riskini göze alarak daha kısa sürede hedefe ulaşmaya çalışabilirdi. Bu durum, özellikle yoğun trafik senaryolarında COLREG kurallarının ihlaline ve sistem güvenliğinin zedelenmesine yol açabilirdi. Dolayısıyla, çarpışmadan kaçınma davranışının pekiştirilmesinde cezaların yüksek tutulması, öğrenilen politikanın gerçek dünya uygulamaları açısından güvenli ve sürdürülebilir olmasını sağlamaktadır.

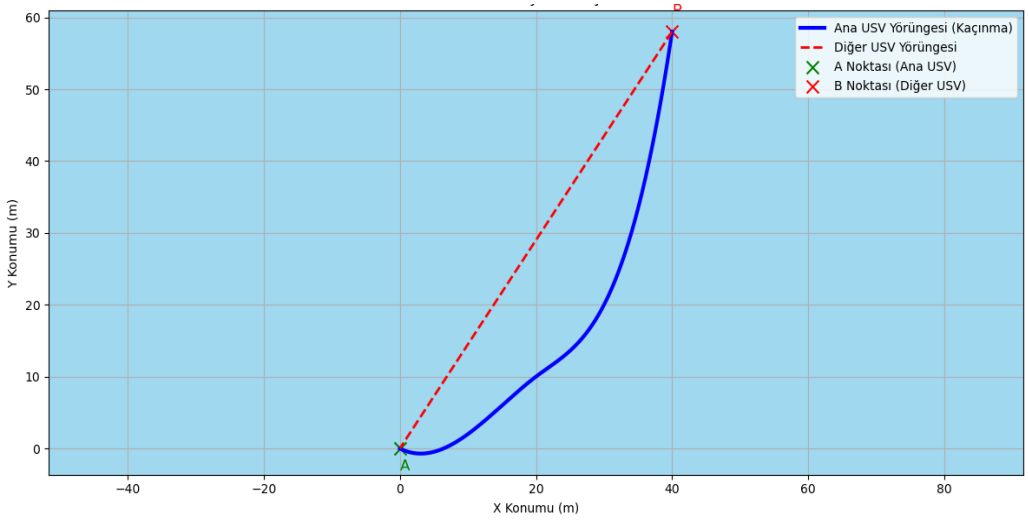
4.8. Pekiştirmeli Öğrenme Temelli Yörünge Planlama Model Doğrulaması

Pekiştirmeli öğrenme algoritmasının doğruluğunu değerlendirmek amacıyla çeşitli çalışmalar gerçekleştirilmiştir. Bu çalışmalar, modelin belirlenen parametreler altında beklenen çıktıları üretip üretmediğini test etmeye yönelik olarak tasarlanmıştır. Engelli ortamda USV'nin örnek hareketi Şekil 4.24'te gösterilmiştir.



Şekil 4.24. Engelli ortamda USV hareketi

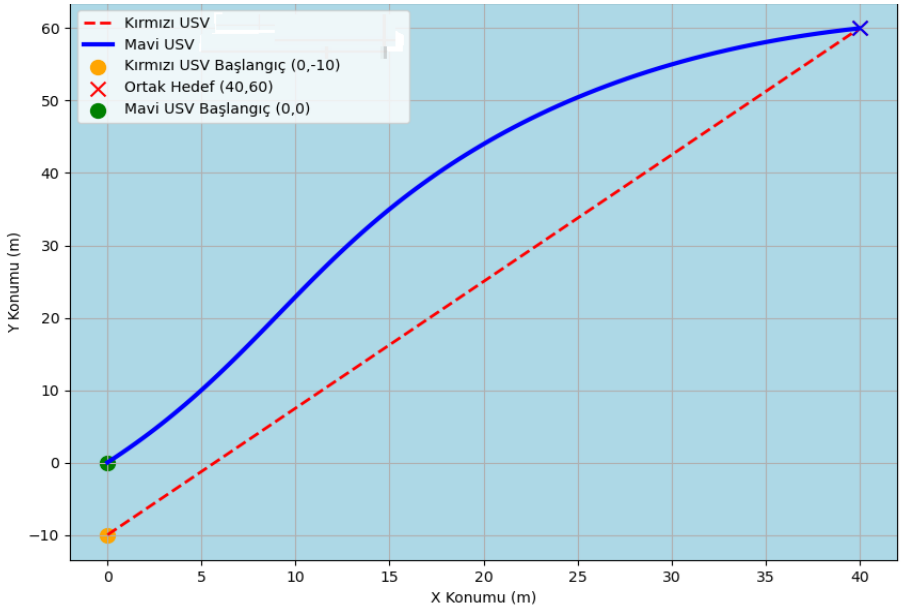
Şekil 4.24'te, USV'nin A noktasından B noktasına giderken sabit engellerle karşılaştığı bir durum modellenmiştir. Ortamda bulunan engeller, kırmızı üçgenlerle temsil edilmiştir. USV, başlangıçta 0,0 koordinatından yola çıkmakta ve engellere yaklaştıkça yörüngesini manevralarla yeniden şekillendirerek çarpışmadan kaçınmaktadır. Çizilen yörünge, USV'nin çevresel farkındalığa sahip olduğunu ve engelleri tanıyarak güvenli bir şekilde yön değiştirebildiğini göstermektedir. A noktasından B noktasına hareket eden USV'nin kafa kafaya senaryo yörünge takibi Şekil 4.25'te göstermektedir.



Şekil 4.25. Kafa kafaya senaryosunda yörünge takibi

Şekil 4.25'te USV'nin kaçınma manevrası mavi yörünge ile gösterilmektedir. Ana USV, başlangıç noktası olan A (0,0) konumundan hareketle, karşıdan gelen diğer aracı fark ettiğinde rota değiştirerek manevra yapmaktadır. Kırmızı kesikli çizgi ise diğer USV'nin hareketini temsil etmektedir.

A noktasından B noktasına hareket eden USV'nin sollama senaryosu Şekil 4.26'da göstermektedir.



Şekil 4.26. Sollama senaryosunda yörünge takibi

Şekil 4.26 sollama senaryosunda ana USV (mavi) başlangıç konumu (0,0) noktasından hareketle, önde sabit hızla ilerleyen diğer USV'yi (kırmızı) COLREG kurallarına uygun şekilde soldan sollayarak ortak hedef noktası olan (40, 60) konumuna ulaşmaktadır.

5. SONUÇ VE ÖNERİLER

Bu çalışmada, pekiştirmeli öğrenme yöntemi ile COLREG kuralları, dış parametreler ve USV'lerin kinematik modelleri dikkate alınarak bir kaçınma algoritması tasarlanmıştır. Eğitim süreci boyunca ajan, USV'lerin çarpışmadan kaçınmak için gerekli olan açısal ve doğrusal hızları, manevra kabiliyetlerini, COLREG kurallarını ve çevresel değişkenleri (rüzgâr, dalga, akıntı gibi) göz önünde bulundurur. Elde edilen algoritmayla eğitilen ajan, farklı karşılaşma senaryoları için test edilmiş ve her senaryoda kaçınma hareketlerini öğrenmiştir.

Elde edilen sonuçlar, modelin farklı çarpışma senaryolarında çevresel koşullara uygun şekilde uyum sağlayabildiğini ve öğrenme sürecinde performansını artırarak çarpışma riskini minimize ettiğini ortaya koymaktadır. Arkadan gelerek sollama senaryosunda modelin çok daha erken senaryolarda yüksek performans gösterdiği; kafa kafaya, karşıdan karşıya geçme ve yol verme senaryolarında ise öğrenmenin daha geç aşamalarda kararlı hale geldiği görülmüştür. Bu durum, çevresel karmaşıklığın ve senaryo yapısının modelin öğrenme hızını ve doğruluğunu doğrudan etkilediğini göstermektedir.

Bu doğrultuda, USV çarpışmalarından kaçınmada pekiştirmeli öğrenmenin önemi açıkça ortaya konmuştur. Ancak bu alandaki çalışmalar hala yeni olup devam etmektedir, dolayısıyla bu konuda geliştirilmesi gereken pek çok araştırma alanı bulunmaktadır. Bu alanları şu şekilde sıralamak mümkündür:

- Farklı yönlerden gelen birden fazla USV’nin yer aldığı senaryolarda, çoklu USV etkileşimleri dikkate alınarak COLREG kurallarına uygun kaçınma manevralarının eşzamanlı ve koordineli biçimde gerçekleştirilmesi,
- Birden fazla USV'nin aynı görevi yerine getirdiği koordinasyonlu durumlarda, araçlar arası çarpışmadan kaçınma stratejilerinin görev paylaşımı ile entegre edilmesi,
- USV’lerin, İnsansız Hava Araçları (İHA) ve kara araçları gibi farklı otonom sistemlerle birlikte çalıştığı heterojen çoklu araç sistemleri için iş birliğine dayalı çarpışma önleme modellerinin geliştirilmesi.

KAYNAKLAR DİZİNİ

Aslan Yıldız, Ö., Sarıçiçek, İ., Özkan, K., & Yazıcı, A., (2024). *Routing with Pickup and Delivery via Deep Reinforcement Learning Derin Pekiştirmeli Öğrenme ile Al-Bırak Rotalama*. 32nd IEEE Conference on Signal Processing and Communications Applications, SIU 2024 Mersin, Turkey

Aslan Yıldız, Ö., Sarıçiçek, İ., & Yazıcı, A., (2025). A Reinforcement Learning-Based Solution for the Capacitated Electric Vehicle Routing Problem from the Last-Mile Delivery Perspective. *APPLIED SCIENCES-BASEL*, vol.15. no.3.

Austin, R. (2010). *Unmanned Aircraft Systems: UAVS Design, Development and Deployment*. Wiley.

Bergman, K., Zhao, L., & Chen, Z. (2020). Design of reward functions in reinforcement learning for autonomous navigation. *Journal of Intelligent & Robotic Systems*, 100(2), 587–603.

Chun, H., & Roh, H. (2021). Solving the Convergence Problem of Deep Q-Network for Unmanned Surface Vehicles in Dynamic Marine Environments. *Journal of AI in Marine Systems, 29*(5), 402–415.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag.

Etemad, M., Zare, N., Sarvmaili, M., Soares, A., Machado, B. B., & Matwin, S. (2020). Using deep reinforcement learning methods for autonomous vessels in 2D environments

Fan, X., & Sun, Y. (2022). Collision Avoidance Behavior of Unmanned Surface Vehicles in Accordance with COLREG Rules. *Journal of Marine Systems, 34*(7), 121–134.

Fan, X., Zhao, J., & Wang, L. (2022). Dynamic Collision Avoidance in Unmanned Surface Vehicles with Reinforcement Learning Based Risk Assessment. *Journal of Marine Science and Technology*, *40*(7), 213–225.

KAYNAKLAR DİZİNİ

Feng, Z. (1992). Oceanographic engineering: theory and applications. *Ocean Engineering, 19*(5), 553–573.

Hasselt, H. van, Guez, A., & Silver, D. (2016). "Deep Reinforcement Learning with Double Q-learning." *AAAI.*

İnan, M., & Baba, B. (2021). Dynamic path planning for unmanned surface vehicles in coastal areas. *Journal of Navigation, 74*(4), 567–579.

Kaiser, Ł., vd. (2019). *Reinforcement Learning and Game Theory for Autonomous Systems*. Springer.

Karaman, S., & Frazzoli, E. (2011). Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research, 30*(7), 846–894.

Kahveci, M., & Can, N. (2017). İnsansız Hava Araçları: tarihçesi, tanımı, dünyada ve Türkiye'deki yasal durumu. *Selçuk Üniversitesi Mühendislik, Bilim ve Teknoloji Dergisi, 5*(4), 511–535.

Larsen, M., & Sørensen, A. (2021). A deep reinforcement learning approach for COLREGs-compliant USV navigation. *Robotics and Autonomous Systems, 142.* 103804.

Li, J. (2023). Deep Reinforcement Learning in Real-World Applications. *Journal of Machine Learning Research, 24*(1), 45–67.

Li, X., & Yang, Y. (2020). Hybrid Path Planning for Unmanned Surface Vehicles: Combining Genetic Algorithms and Reinforcement Learning. *Journal of Marine Engineering, 68*(2), 115–128.

Li, Y., & Wu, J. (2021). Collision Risk Calculation and Navigation Strategy for Unmanned Surface Vehicles Considering COLREGs. *Maritime Technology and Research, 14*(3), 199–211.

KAYNAKLAR DİZİNİ

Li, Y., & Yu, D. (2019). Deep learning techniques for dynamic obstacle avoidance in unmanned surface vehicles. *IEEE Transactions on Automation Science and Engineering, 16*(4), 1293–1305.

Li, Y., & Yang, C. (2020). Path planning and collision avoidance of USVs based on genetic algorithm and reinforcement learning. *Journal of Marine Science and Engineering*, 8(5), 341.

Lillicrap, T.P., et al. (2016). "Continuous control with deep reinforcement learning." *arXiv preprint arXiv:1509.02971.*

Lin, L. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning, 8*(3–4), 293–321.

Lin, X., & Fang, Z. (2013). Influence of environmental forces on autonomous vehicle navigation. *Journal of Ship Research, 57*(1), 77–90.

Manley, J. E. (2008). Unmanned surface vehicles, 15 years of development. *OCEANS 2008*, IEEE.

Mao, Z., Zhang, L., & Huang, T. (2018). Stabilizing Deep Q Networks with Target Networks. *Proceedings of the International Conference on Neural Information Processing, 111*(5), 99–110.

Mnih, V., et al. (2015). "Human-level control through deep reinforcement learning." *Nature.*

Park, H., Kim, J., Lee, S., & Choi, Y. (2017). Research on unmanned surface vehicles (USVs): Applications and developments. *Journal of Maritime Science and Technology, 25*(3), 233–245.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.).

MIT Press.Roh, H., Kim, S., & Park, J. (2021). Multi-Agent Reinforcement Learning for Coordinated Navigation of Unmanned Surface Vehicles. *Autonomous Systems Journal, 32*(3), 212–224.

KAYNAKLAR DİZİNİ

Roh, M., Kim, J., & Park, G. (2021). Cooperative collision avoidance for multiple unmanned surface vehicles using multi-agent reinforcement learning. *Sensors*, 21(12), 4138.

Sivaraman, S., Rajendran, S., & Perera, L. P. C. (2022). Data driven control based on Deep Q-Network algorithm for heading control and path following

Sonntag, L., Fischer, M., & Becker, J. (2025). Integrated maneuvering and collision avoidance of USVs under COLREGs using reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*. Forthcoming.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tamura, H., Tamaki, R., Hirayama, T., & Ishii, K. (2019). Rule-based collision avoidance method considering COLREGs for autonomous ships. *IFAC-PapersOnLine*, *52*(21), 263–268.

Van Hasselt, H., Guez, A., & Silver, D. (2015). Deep Reinforcement Learning with Double Q-learning. *arXiv preprint arXiv:1509.06461*.

Wang, F., Zhang, J., & Chen, K. (2022). Genetic Algorithm Assisted Path Planning Combined with Reinforcement Learning for Unmanned Surface Vehicles. *Journal of Marine Engineering*, *45*(5), 285–296.

Wang, J., Liu, Z., Tang, J., & Yu, H. (2022). *Path planning for unmanned surface vehicles based on improved artificial potential field and cubic B-spline*. Applied Ocean Research, 122, 103173.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3–4), 279–292.

Web of Science (WOS). (2023). Search Query: “unmanned surface vehicles (USV)” and “path planning, task assignment and routing”. Accessed on: [Date].

KAYNAKLAR DİZİNİ

Wen, Z., Yang, S., & Li, J. (2023). Category-Based Exploration for Collision-Free Path Planning in Unmanned Surface Vehicles Using Reinforcement Learning. *Journal of Robotics and Automation, 65*(3), 91–103.

Xia, Y., Zhao, X., & Li, Z. (2020). Trajectory Planning and Collision Avoidance for Unmanned Surface Vehicles Using Gaussian Mixture Models and Gaussian Process Regression. *International Journal of Autonomous Systems, 52*(3), 201–215.

Xie, M., Wang, K., Zhang, Y., & Liu, J. (2021). COLREGs-compliant decision making for autonomous ships using deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems, 22*(6), 3329–3341.

Xu, X., Tang, Y., & Cheng, L. (2014). Survey on development of unmanned surface vehicle. *Journal of Marine Science and Application,* 13(4), 452–464.

Xu, W., Zhou, J., & Chen, L. (2022). Collision Avoidance for Unmanned Surface Vehicles Using Artificial Potential Field and Deep Q-Network. *International Journal of Marine Systems, 55*(2), 134–146.

Xu, Y., Yang, D., Li, W., & Liu, H. (2022). Deep reinforcement learning-based collision avoidance algorithm for autonomous ships considering COLREGs. *Applied Sciences*, 12(1), 156.

Yan, J., & Jiang, M. (2021). Kinematic modeling and path planning for unmanned surface vehicles. *Journal of Navigation, 74*(6), 903–915.

Yan, X., Li, J., & Zhao, Z. (2021). Collision-Free Navigation in Complex Marine Environments Using Reinforcement Learning. *International Journal of Robotics and Autonomous Systems, 78*(6), 101–115.

Yang, Y., Zhang, S., & Wu, Q. (2019). *A hybrid path planning method for unmanned surface vehicle based on ant colony optimization and improved fast marching method*. Ocean Engineering, 186, 106124.

KAYNAKLAR DİZİNİ

Yazıcı, M., & Ark., İ. (2021). Doğrulama ve Geçerleme Süreçlerinin Robotik ve Otonom Sistemlerdeki Rolü. *Robotik Sistemler ve Otomasyon Dergisi, 12*(3), 45–58.

Zhang, J., Li, W., & Yu, S. (2023). Dynamic Trajectory Planning with Multiple Obstacles for Unmanned Surface Vehicles Using B-Spline-Based Models. *Journal of Robotics and Automation, 41*(2), 109–121.

Zhou, C., Gu, S., Wen, Y., Du, Z., Xiao, C., Huang, L., & Zhu, M. (2020). The Review Unmanned Surface Vehicle Path Planning: Based on Multi-modality Constraint. *arXiv preprint arXiv:2007.01691*.

Zhou, Y., Li, W., & Zhang, J. (2019). Path Planning for Unmanned Surface Vehicles Using Reinforcement Learning. *Journal of Autonomous Navigation, 45*(4), 555–568.

Zhou, Y., Wu, J., & Zhao, Y. (2019). Path planning for unmanned surface vehicles based on deep reinforcement learning. *Sensors*, 19(14), 3170

Zhu, L., Zhang, H., & Chen, M. (2021). Enhancing Deep Q-Network Convergence for Path Planning in Unmanned Surface Vehicles. *Journal of Artificial Intelligence and Robotics, 29*(4), 301–315.

Zhu, L., Zhang, H., & Li, M. (2018). Application of Reinforcement Learning for Path Planning of Unmanned Surface Vehicles. *Journal of Marine Robotics, 35*(7), 421–433.

Zhu, X., vd. (2018). Reinforcement learning based path planning and collision avoidance for unmanned surface vehicles. *Autonomous Robots*, *42*(3), 567–583.

Zhu, W., Huang, J., & Li, M. (2021). Enhanced deep Q-learning for intelligent ship collision avoidance. *Applied Ocean Research*, 112. 102655.

Zhuang, Y., Zhang, Y., Zhu, X., & Xu, C. (2020). *Path planning for unmanned surface vehicle based on improved A* algorithm*. Journal of Marine Science and Technology, 25(4), 512–523.