

Derin Öğrenme ile 3B İskelet Tabanlı Spor Hareketi Tanıma Sistemi

3D Skeleton-Based Sport Action Recognition System via Deep Learning

Merve DAYAPOGLU¹, Esra ŞAVIRDI¹, İlknur AKCABA¹, Sait ALP^{2,*}, Sefa KUCUK^{3,*}

¹Bilgisayar Mühendisliği Bölümü, Erzurum Teknik Üniversitesi, Erzurum, TÜRKİYE

²Yapay Zeka Mühendisliği Bölümü, Trabzon Üniversitesi, Trabzon, TÜRKİYE

³Elektrik-Elektronik Mühendisliği Bölümü, Erzurum Teknik Üniversitesi, Erzurum, TÜRKİYE

{merve.dayapoglu17, esra.savirdi93, ilknur.akcaba26, sefa.kucuk}@erzurum.edu.tr, saitalp@trabzon.edu.tr

Özetçe —Bu çalışmada, spor hareketlerini tanımak için derin öğrenmeye dayalı bir sistem geliştirilmiştir. İlk aşamada, Microsoft Kinect Xbox One kullanılarak KYM (Kırmızı Yeşil Mavi) görüntülerin ve üç boyutlu - 3B iskelet verilerinin eş zamanlı yakalanmasını sağlayan bir uygulama oluşturulmuştur. Geliştirilen sistem, gerektiğinde veri toplamak ve video kayıtlarını düzenlemek için kullanıcı dostu bir arayüze sahiptir. Beş farklı sporcu hareketini tanımlamak için 10 farklı kişiden çekilen videolar, geliştirilen arayüz ile öğrenme modelleri için kullanılmak üzere hazır veri haline getirilmiştir. Bu sürecin sonucunda, toplamda 500 örnek içeren özel bir veri kümesi oluşturulmuştur. İskelet verilerindeki zamansal bağımlılıkları karakterize etmek için 1D-CNN (Bir Boyutlu Evrişimsel Sinir Ağı) ve LSTM (Uzun Kısa Süreli Bellek) derin öğrenme modelleri kullanılmış, bu modellerin birlikte kullanımının etkinliğini araştırılmıştır. Çapraz doğrulama ile en iyi performansı veren modelin belirlenmesi için farklı aktivasyon fonksiyonları ve en iyileyci (optimizer) kombinasyonlarının detaylı analizleri yapılmıştır. En iyi performans gösteren model ile beş spor eylemi için %87’i aşan bir sınıflandırma doğruluğu elde edilmiştir.

Anahtar Kelimeler—veri toplama aracı, derin öğrenme, CNN, LSTM, iskelet tabanlı insan hareketi tanıma

Abstract—In this paper, we developed a deep learning-based system to recognize customized sports actions. During the first phase, we created an application that enables the simultaneous capture of RGB images and three-dimensional (3D) skeleton data using Microsoft Kinect Xbox One. The developed system has a user-friendly interface for collecting data and editing video recordings as needed. Using the developed interface, we captured data on five sports actions performed by ten people and then processed the data for use in training deep learning models. The created dataset consists of 500 action instances. We employed 1D-CNN (One-Dimensional Convolutional Neural network) and LSTM (long short-term Memory) deep learning models to characterize temporal-spatial dependencies in action’s skeleton data, and investigated their effectiveness. We conducted the cross-validation techniques to determine the best-performing model, involving detailed analyses of different activation functions and optimizer combinations. The top-performing model achieved a classification accuracy exceeding 87% for five sports actions.

Keywords—data collection tool, deep learning, CNN, LSTM, skeleton-based human action recognition

I. GİRİŞ

İnsan hareketlerini tanımlamak, çok sayıda hareket kombinasyonunun bulunması, bireylerin aynı hareketleri farklı şekillerde gerçekleştirmesi ve ortam koşullarının etkisi gibi sebeplerden dolayı oldukça zorlu bir süreçtir [1], [2]. Yalnızca KYM (Kırmızı Yeşil Mavi) görüntülerinin çerçevelerinden çıkarılan özniteliklere dayanan çözümler, zamansal bilgi eksikliği ile ışık, renk ve görüntü kalitesi gibi faktörler nedeniyle insan hareketlerini tespit etmede yeterli değildir [2], [3]. Geleneksel olarak iki boyutlu - 2B görüntüler hareket tanıma için kullanılmış olsa da birçok araştırmacı artık Kinect veya diğer derinlik sensörleriyle elde edilen üç boyutlu veri dizilerini tercih etmektedir [4]–[6].

Yakın zamanda geliştirilen sistemlerden üretilen vücut anahtar eklemlerinin doğruluğu, derinlik algılayan sensörler ile renkli görüntü verilerini içeren KYM-D (Kırmızı Yeşil Mavi-Derinlik) verilerinin daha zengin içerik sunması, araştırmacıları çeşitli uygulamalarda yararlı olabilecek bilgilendirici verileri çıkarmak için insan hareketini anlama ve analiz etme konusunda çalışmaya motive etmiştir [7]. Ancak, üç boyutlu noktaların takibiyle iskelet verilerinden hareket analizi hala zorlu bir araştırma alanıdır [5]. İskelet yapısının eklemli olması nedeniyle tüm vücudun hareketini değerlendirmek zordur [7]. Her eklem bir veya daha fazla ana eklem etkisi altında, sayısız hareket yönü olasılığıyla öngörülemez bir şekilde hareket etmektedir [7]. Bu nedenle, hem KYM görüntülerinin hem de iskelet verilerinin kullanılması, hareketlerin ölçümleri ardışık verilerden oluştuğu için insan hareketini tanımlamayı zorlaştırmaktadır [1].

Hareket yakalama ve işleme teknolojilerinin hızla ilerlemesi, özellikle Microsoft Kinect kameralar ve gerçek zamanlı 3B hareket yakalama (motion capture - MoCap) sistemleri gibi düşük maliyetli derinlik sensörlerinin gelişimi, KYM-D videolardan nesnelerin üç boyutlu iskeletlerinin tespit edilmesini mümkün kılmıştır [4], [5], [8]. Bu gelişmeler, düşme hareketi tespiti, sporda hareket analizi veya insan davranışı analizine yönelik uygulamaların geliştirilmesiyle video dizilerinden insan hareketi tanıma çalışmalarının artmasına olanak sağlamıştır [8]–[10]. Son yıllarda, derin öğrenme yöntemleriyle iskelet

veri kümeleri kullanılarak hareket tanımda önemli ilerlemeler kaydedilmiştir. İskelet veri kümeleri, insanın hareketli eklemlerini içermekte olup, bu tür veri kümeleri videolardaki kişilerin eylemlerini temsil etmek için genellikle daha az miktarda veri kullanılmaktadır. KYM video kümelerinde aydınlatma değişimi ve gürültünün insan hareketini tanımadaki etkileri oldukça büyüktür. İskelet tabanlı hareket tanıma ise bu sorunlara karşı gürbüz bir yapıdadır [5], [8], [11]. Aynı zamanda gerçek dünyada iskelet veri kümeleri kişisel kimliği açığa çıkarmaz, bu nedenle de kişisel verilerin gizliliği korumaktadır. İskelet veri kümelerinin bu avantajları, iskelet veri kümelerinin insan hareketi tanımının temel bileşeni haline gelmesini sağlamıştır [2], [11].

Literatürdeki son yaklaşımlar, insan hareketlerini tanımak için KYM-D videolarından uzay-zaman özelliklerini çıkarmak amacıyla derin öğrenme tekniklerini kullanmaktadır [4]. İnsan aktivitesi tanıma (Human Activity Recognition - HAR), spesifik beden hareketlerini karakteristik algılayıcı işareti örüntülerine dönüştürdükten sonra, bu verileri derin öğrenme teknikleri ile sınıflandırmaya dayanmaktadır [1]. Bu bağlamda, 1D-CNN ve LSTM gibi derin öğrenme modelleri hareket tanıma görevlerinde öne çıkmaktadır [12]. 1D-CNN, zaman serisi verilerinde veri noktalarını ardışık zaman dilimleri boyunca filtreleyerek zamansal bağımlılıkları yakalayabilmektedir. Bununla birlikte, LSTM ağları zaman serisi ve sıralı verilerdeki uzun vadeli bağımlılıkları öğrenmek için tasarlanmış özel bir RNN türü olup, bilgi akışını kontrol eden giriş, unutma ve çıkış kapıları ile donatılmış hafıza hücrelerine sahiptir. Bu sayede zamansal bağımlılıkların daha etkin bir şekilde öğrenilmesine olanak sağlamaktadır [12].

Bu çalışmada öncelikle, sporcu hareketlerinin kolaylıkla alınabilmesi ve alınan kayıtların istenilen şekilde düzenlenebilmesi için kullanıcı dostu özel bir ara yüz geliştirilmiştir. Oluşturulan ara yüz vasıtası ile Microsoft Kinect Xbox One sensörü kullanılarak sporcuların hem KYM hem de iskelet verileri eş zamanlı kaydedilmiştir. Beş farklı sporcu hareketini tanımlamak için 10 farklı kişiden çekimler alınmış, elde edilen video çekimleri geliştirilen arayüz ile öğrenme modelleri için hazır veri haline getirilmiştir. Bu işlemler sonucunda 5 farklı sınıftan toplamda 500 sporcu verisi içeren özel bir veri kümesi oluşturulmuştur. Daha sonra, oluşturulan veri kümesi üzerinde derin öğrenme modelleri ve bu modellerin kombinasyonları uygulanmış, çapraz doğrulama ile en iyi performansı veren modelin belirlenmesi için farklı aktivasyon fonksiyonları ve en iyileşici (optimizer) kombinasyonlarının karşılaştırılması analizleri yapılmıştır.

II. DERİN ÖĞRENME MODELLERİ

A. Tek Boyutlu Evrişimsel Sinir Ağı (1D-CNN)

1D-CNN, zaman serisi işlemlerinde ve HAR gibi alanlarda başarılı olan en popüler derin öğrenme algoritmalarından biridir [13], [14]. Bu model, veri içindeki yakın komşuluk ilişkilerini etkili bir şekilde öğrenme yeteneğine sahiptir ve elle tasarlanmış özellikler yerine yüksek seviyeli özellikleri otomatik olarak çıkarabilmektedir. 1D-CNN'nin düşük hesaplama karmaşıklığı avantajı, onu çeşitli karmaşık görevlerde oldukça yetenekli kılmaktadır [13]. 1D-CNN, ham verilerde gizlenen doğrusal olmayan özellikleri alternatif evrişim katmanı ve havuzlama katmanı aracılığıyla

otomatik olarak çıkarır ve uyarlanabilir özellik öğrenimi sağlar. Bu sayede, geleneksel algoritmalarındaki elle tasarlanmış özellik çıkarma işlemini ortadan kaldırarak uçtan uca bilgi işleme sürecini mümkün kılar [15]. 1D-CNN'nin başarısının temelinde, veri içindeki zamansal bağımlılıkları ve örüntüleri etkili bir şekilde öğrenme kapasitesi yatmaktadır. Ayrıca, verilerin normalizasyon sonrasında doğrudan modele aktarılabilmesi, veri ön işleme sürecini basitleştirir ve modelin daha hızlı eğitilmesine olanak tanır [16]. Evrişim ve havuzlama katmanları, verilerin belirli özelliklerini çıkarırken gereksiz bilgileri filtreleyerek modelin genel performansını artırır. Derin öğrenme teknikleri, geleneksel makine öğrenme yöntemlerine kıyasla daha karmaşık ve yüksek seviyeli özellikleri çıkarma yeteneğine sahip olduğundan, 1D-CNN insan aktivitesi tanıma gibi alanlarda öne çıkmaktadır [17].

B. Uzun Kısa Süreli Bellek (LSTM)

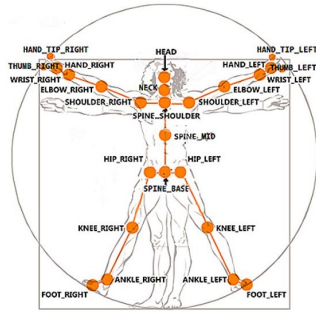
LSTM sinir ağı, yinelenen sinir ağlarının (Recurrent Neural Network - RNN) bir türüdür [18]. RNN'e kıyasla, LSTM hem uzun vadeli bağımlılık sorunlarını ele alabilmektedir hem de gradyan kaybolma olasılığını azaltabilmektedir [18], [19]. LSTM, özellikle uzun vadeli bağımlılıkları korumak için tasarlanmıştır ve bu nedenle zaman serileri gibi ardışık veri türlerinde zamansal bilgileri yakalama yeteneğine sahiptir [20], [21]. LSTM'nin kullanılmasının başlıca avantajları, büyük miktarda veri entegre etme yeteneği ve karmaşık özellikleri otomatik olarak öğrenme kapasitesidir [22]. LSTM'nin temel yapısı, gelen veriyi işlemek için bir döngü içerir. Bu döngü, veri akışının her bir zaman adımında çalışır ve LSTM hücresinin iç durumunu günceller. LSTM hücreleri, birbirine bağlı düğümlerden oluşur ve gelen veri akışını karşılamak için giriş, çıkış ve unutma kapılarına sahiptir [18]. Bu kapılar, hücrenin iç belleğini güncellemek ve ardışık veri akışını kontrol etmek için kullanılır [18]. LSTM'nin donanım ivmecileri, LSTM modellerini gerçek zamanlı uygulamalar için optimize etmek amacıyla kullanılır. Bu ivmeciler, LSTM modellerinin performansını artırabilir ve hesaplama gücünü optimize edebilir [23].

III. KINECT XBOX ONE

Xbox One Kinect Sensörü, 22 Kasım 2013 tarihinde piyasaya sürülmüş olup başlangıçta oyun konsolu için geliştirilmiştir. Cihaz, birçok etkileyici işlev sunmaktadır. Oyuncunun kalp atış hızını ve hareket hızını algılayabilir, yüz ifadelerini tanıyabilir, hareketleri izleyebilir ve 25 farklı eklemin konumunu tespit edebilir (Şekil 1'de temsili olarak gösterilmiştir). Kızılötesi kamera ve yayıcı kullanarak çevresini yüksek doğrulukla algılayabilen Kinect, aynı anda altı kişiyi tanıyabilir ve bu kişilerin iskeletlerini izleyebilir. Bu iskelet bilgileri, insan tanımlama sistemlerinde iskelet uzunluğunu hesaplamak için kullanılabilir [24], [25].

Kinect'in üç farklı sensörü (KYM kamera, derinlik sensörü ve kızılötesi sensör) tek bir cihazda birleştirilmesi, hem görsel hem de derinlik bilgilerini yüksek doğrulukla elde etmeyi sağlar ve cihazın esnekliğini gösterir. Kinect'te yer alan sensörler aşağıdaki özelliklerde veri sağlayabilmektedir.

- KYM kamera: 1920×1080 çözünürlük
- Derinlik sensörü: 512×424 çözünürlük



Şekil 1: Kinect v2'nin 25-eklemlili iskelet konumları [24]

- Kızılötesi sensör: 512×424 çözünürlük

Bu özellikler, Kinect'in eğlence dışında profesyonel ve akademik araştırmalarda da kullanılabileceğini göstermektedir.

IV. YAPILAN ÇALIŞMALAR

Bu bildiride, 1D-CNN ve LSTM modelleri kullanılarak tasarlanan modellerden hangisinin performansının iskelet tabanlı verilerde daha iyi olduğu belirleyebilmek için çalışmalar yapılmıştır. Deneylerde, 5 temel sporcu hareketi belirlenerek istenilen amaca uygun veri elde etmek için özel bir arayüz oluşturulmuş, farklı kişilerden video çekimleri alınmıştır. Bu çekimler sonucunda elde edilen ham veriler geliştirilen arayüz ile düzenlenerek öğrenme modelleri için uygun veri kümesi haline getirilmiştir. Bu bildiri kapsamında yapılan çalışmalar alt bölümlerde detaylandırılmıştır.

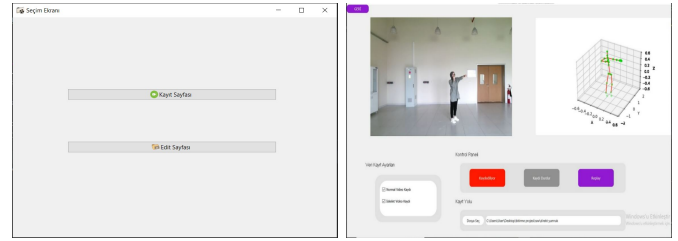
A. Arayüz Tasarımı

Verilerin alınma aşamasında kullanılacak olan Kinect sensöründen KYM görüntüleri ve iskelet verilerini kaydetmek için Python programlama dili ile yazılmış uygulamalarda Qt kütüphanesini kullanmayı sağlayan PySide6 kütüphanesi ile özel bir arayüz tasarlanmıştır. Qt, yüksek performanslı ve platformlar arası uygulamalar geliştirmek için kullanılan popüler bir C++ grafik kullanıcı arayüzü kütüphanesidir. PySide6, bu kütüphanenin Python bağlamalarını sağlamaktadır.

Microsoft'un Kinect desteğini sonlandırması nedeniyle, Kinect sensörünün arayüzde açılması sırasında yaşanan hataların düzeltilmesi için bir sanal ortam oluşturulmuştur. Bu sanal ortamda, gerekli kütüphanelerin belirli sürümleri indirilmiştir. Bu sürümler sayesinde Kinect kameranın arayüzde açılma problemi çözülmüştür.

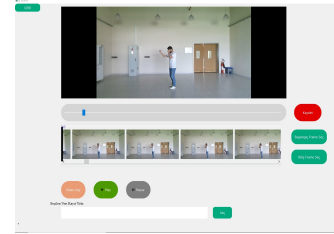
KYM görüntülerini ve iskelet verilerini eş zamanlı kaydederken arayüzün yanıt vermesini engelleyen uzun süreli işlemleri arka planda yürütmek için iş parçacıkları (thread) kullanılmıştır. Bu sayede kullanıcı isteklerine daha hızlı yanıt veren kullanıcı dostu bir arayüz geliştirilmiştir.

Geliştirilen arayüzde yapılacak işlemler için üç farklı ekran tasarlanmıştır. İlk ekran Şekil 2.a'da gösterildiği gibi kayıt ve video parçalama işlemleri için oluşturulan ekranlara yönlendiren butonlardan oluşmaktadır. İkinci ekranda (Şekil 2.b), KYM görüntülerinin ve iskelet verilerinin kayıt işlemleri gerçekleştirilirken, üçüncü ekranda (Şekil 2.c) ise toplanan KYM görüntülerinin başlangıç ve bitiş çerçevelerini seçerek kaydedilen görüntüleri ve eş zamanlı olarak bu görüntülerin



(a) Giriş Ekranı

(b) Kayıt Ekranı



(c) Düzenleme Ekranı

Şekil 2: Geliştirilen kullanıcı arayüzü

metin dosyasında tutulan iskelet verilerini kesme işlemleri yapılmaktadır.

B. Veri Kümesinin Oluşturulması

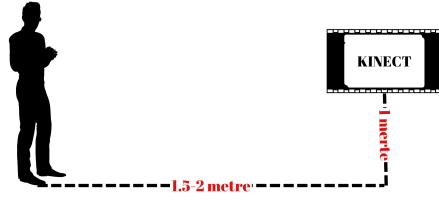
Bu çalışma kapsamında, Şekil 3'te gösterildiği gibi tekme, düz yumruk, servis, kroşe ve aparkat olmak üzere 5 farklı sporcu hareketi ele alınmıştır.

Kinect kameranın derinlik aralığı 8 metre ve iskelet takip aralığı ise 0.5–4.5 metre arasındadır. Veri toplama aşamasında Kinect kameranın konumlandırılması bu aralıklar dikkate alınarak belirlenmiştir. Kinect kamera Şekil 4'te gösterildiği gibi kamera ile kişi arasında yaklaşık 1.5 – 2 metre olacak şekilde konumlandırılmış, yerden yüksekliği ise yaklaşık 1 metre olacak şekilde ayarlanmıştır.

Çekimlerde, farklı boy ve kilolarda on denek ile çalışılmıştır. Yapılan çekimlerde, her bir denek belirlenen beş temel sporcu hareketini sırayla ve iki tekrarla, beş farklı açıdan gerçekleştirmiştir. Her bir hareketin çekimi, beş farklı açıdan yapılarak devam etmiştir. Daha sonra, elde edilen kayıtlar üzerinde video düzenleme ekranında her bir hareketin başlangıç ve bitiş çerçeveleri işaretlenerek video kesme işlemi gerçekleştirilmiştir. Bu işlemler sonunda toplamda 500 veriden oluşan bir veri kümesi elde edilmiştir.

Sporcu Hareketi	Kışt Sayısı	Toplam Veri Sayısı
	9	68
	9	46
	9	67
	8	49
	9	80

Şekil 3: Oluşturulan veri kümesinin detayları

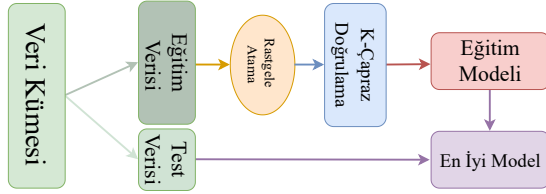


Şekil 4: Çekim alanının temsili gösterimi

C. Veri Kümesinin Ayrılması

Öğrenme modellerine verilmeden önce veri kümesindeki bozuk veriler tespit edilip ayrıldıktan sonra geriye kalan veriler eğitim ve test verileri için %80'i eğitim, %20'si test olacak şekilde ayrılmıştır. Bu işlem sonucunda eğitim ve test için sırasıyla 248 ve 62 veri elde edilmiştir.

Eğitim sırasında aşırı uyum sorununu çözmek için çapraz doğrulama kullanılmıştır. Şekil 5'te K -çapraz doğrulama ile en iyi modelin belirlenme süreci tasvir edilmektedir. Şekilde gösterildiği gibi orijinal veri kümesi eğitim ve test verisi olarak ayrıldıktan sonra K sayısı 5 olarak belirlenmiş yani eğitim için ayrılan veriler eşit olarak 5 alt kümeye bölünmüştür. Eğitim sürecinde dönüşümlü olarak 4 alt küme eğitim verisi olarak kullanılırken geriye kalan tek alt küme doğrulama verisi olarak kullanılmıştır.

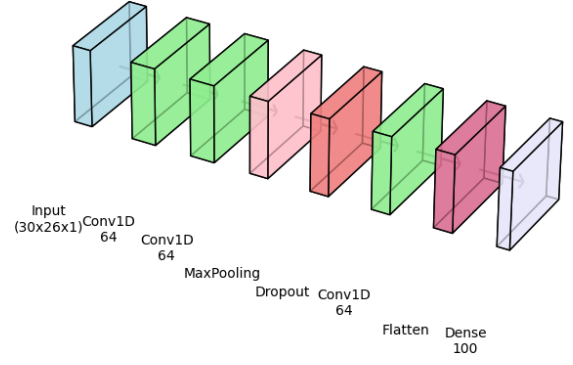


Şekil 5: En iyi modelin belirlenme süreci

D. Ağ Modeli

Veri kümesindeki iskelet verileri farklı çerçeve sayılarına sahip olduğundan, öncelikle her bir iskelet verisini 30 çerçeveye tamamlamak için dolgulama (padding) işlemi uygulanmıştır. Daha sonra iskelet verilerinden, belirli açılar ve mesafeler kullanılarak 26 farklı öznelik çıkarılmıştır. İşlenmiş veriler, Şekil 6'da gösterilen CNN ağı yapısına giriş olarak verilmiştir. Burada, 30 iskelet çerçevesi, her bir çerçeve için 26 öznelik ve tek kanallı veri yapısı kullanılmıştır.

Oluşturulan ağ yapısı beş ana bileşenden oluşmaktadır. Bunlar: evrişim, maksimum havuzlama, sönümlenme, düzleştirme ve yoğun katmanlardır. Evrişim katmanı (convolutional layer), bir özellik matrisi oluşturmak için boyutları önceden belirlenmiş filtreler kullanmaktadır. Veriler bu filtrelerle işlenerek özellikler çıkarılır. Maksimum havuzlama (max pooling) katmanı, matrisin boyutunu azaltmak için kullanılır. Bu katman, belirli bir havuzlama penceresi içinde en yüksek değeri seçerek veri boyutlarını küçültür. Sönümlenme (dropout) katmanı, modelin aşırı öğrenmesini (overfitting) önlemek için kullanılır. Bu katman, eğitim sırasında belirli bir oranda nöronları rastgele seçerek devre dışı bırakır. Düzleştirme (flatten) katmanı, önceki katmanın çıktısını düzleştirerek tek boyutlu bir vektöre dönüştürür. Yoğun (dense) katman, tam bağlantılı



Şekil 6: 3 Katmanlı 1D-CNN ağ modeli

(fully connected) bir katman olup, genellikle ağı son katmanlarında bulunur. Bu katman, özelliklerin son kararını vermek için kullanılır ve girdileri işleyip çıktı nöronlarına ileterek sınıflandırma veya regresyon görevlerini gerçekleştirir.

Modelde aktivasyon fonksiyonu olarak düzeltilmiş lineer ünite (rectified linear unit - ReLU) kullanılmıştır. ReLU aktivasyon fonksiyonu, evrişim ve yoğun katmanlarının çıktısını hesaplamak için kullanılır ve doğrusal olmayan bir dönüşüm uygular. Modelin ağırlıklarını güncellemek için en iyileşici (optimizer) olarak Adam kullanılmıştır. Adam adaptif öğrenme oranı kullanarak hızlı ve etkili bir optimizasyon sağlamaktadır. Modelin hata oranını hesaplamak için kayıp fonksiyonu olarak kategorik çapraz entropi (categorical cross-entropy) seçilmiştir. Kategorik çapraz entropi çok sınıflı sınıflandırma problemleri için yaygın olarak kullanılan bir kayıp fonksiyonudur. Modelin eğitim sürecinde veri kümesinin tamamı üzerinden geçilmesi için dönem/epok (epoch) sayısı 100 olarak belirlenmiştir.

Çapraz doğrulama ile eğitim sonunda elde edilen en iyi model için karışıklık matrisi Şekil 7'de gösterilmektedir. Karışıklık matrisi, beş farklı sınıf için modelin sınıflandırma performansını göstermektedir. Matrisin her satırı gerçek sınıfı, her sütunu ise tahmin edilen sınıfı temsil etmektedir. Model, servis ve tekme sınıflarında oldukça başarılı sonuç elde etmiştir. Bununla birlikte, aparkat sınıfı için oldukça düşük yanlış sınıflandırma oranı üretmiştir. Direkt yumruk ve kroşe sınıflarında bazı örnekler kaçırılmış, özellikle direkt yumrukların bazıları kroşe olarak tahmin edilmiştir. Model genel olarak iyi bir performans sergilerken direkt yumruk ve kroşe sınıflarının benzerlikleri nedeniyle bu iki sınıfı ayırt etmede

aparkat	14		1	1	
direkt yumruk		6	3		
kroşe	1	1	7	1	
servis (voleybol)				13	
tekme					14
	aparkat	direkt yumruk	kroşe	servis (voleybol)	tekme

Şekil 7: En iyi modelin belirlenme sürecinde elde edilen karışıklık matrisi

Tablo I: 3 Katman 1D-CNN modeli üzerinde farklı aktivasyon fonksiyonları ve en iyileycileri için performans değerlendirme sonuçlarının karşılaştırılması

Fonksiyon	En iyileyici	Eğitim DO	Doğrulama DO	Test DO
ReLU	Adam	0.86	0.71	0.87
ReLU	Adamax	0.76	0.62	0.72
Tanh	Adam	0.86	0.68	0.77
Tanh	Adamax	0.91	0.74	0.74
ELU	Adam	0.86	0.69	0.77
ELU	Adamax	0.77	0.64	0.73

zorlanmıştır.

Tablo I'de 3 katman 1D-CNN içeren model üzerinde farklı aktivasyon fonksiyonları ve en iyileycileri için eğitilen modellerin eğitim, doğrulama ve test doğruluk oranları (DO) gösterilmektedir. Tablo I'den ReLU ve Adamax kombinasyonunun düşük doğrulama ve test DO'larına sahip olduğu görülebilmektedir. Tanh ve Adam için eğitim DO'su yüksek, doğrulama ve test DO'ları ise iyi sayılabilecek düzeydedir. Ancak eğitim doğruluğu ile doğrulama ve test doğrulukları arasında belirgin bir fark bulunmaktadır. Benzer bir durum Tanh ve Adamax için de söz konusudur. ELU ile kullanılan en iyileyciler yüksek eğitim DO sunarken yüksek test DO elde edememiştir. Son olarak, ReLU ve Adam kombinasyonu yüksek test DO ile ön plana çıkmaktadır.

Tablo I'de yer alan sonuçlar dikkate alındığında, ReLU ve Adam kombinasyonunun en iyi performansı sergilediği söylenebilmektedir. Bu sonuçlar dikkate alınarak aktivasyon fonksiyonu ReLU ve en iyileyici olarak Adam belirlenmiş, Tablo II'de verilecek farklı mimariler için de bu kombinasyon kullanılmıştır.

V. DENEYSEL SONUÇLAR

Çalışma kapsamında, model eğitimde yapılan birçok değişikliğe rağmen doğruluk oranının artmadığı görülmüş, bu durumun veri kümesindeki bozukluklardan kaynaklanabileceği düşünülmüştür. Bu durum üzerine iskelet verileri yeniden çizdirilerek bozuk veriler tespit edilmeye çalışılmıştır. Bozuk verilerin Kinect Xbox One sensörünün siyah renkli giysilerde iskelet verilerini okurken sebep olduğu ve belirli açılarda görünmeyen eklem noktalarını tahmin ederken hatalı tahminlerde bulunduğu tespit edilmiştir. Bozuk verilerin veri kümesinden çıkarılmasının ardından, 1D-CNN ve LSTM modellerini farklı kombinasyonlarla kullanıldığında doğruluk oranının arttığı gözlemlenmiştir.

Çalışmada performans analizi için doğruluk, kesinlik, duyarlılık ve F1-skoru metrik olarak kullanılmıştır. Doğruluk, modelin doğru tahmin ettiği veri sayısının toplam veri sayısına oranıyla aşağıdaki formül kullanılarak hesaplanmaktadır.

$$\text{Doğruluk} = \frac{\text{Doğru tahmin sayısı}}{\text{Toplam veri sayısı}} \quad (1)$$

Kesinlik, pozitif olarak tahmin edilen verilerin gerçekte kaç tanesinin pozitif olduğunu göstermektedir. Bu ölçüm, modelin performansını her sınıf için ayrı ayrı değerlendirmektedir. Herhangi bir sınıf için kesinlik

$$\text{Kesinlik} = \frac{\text{Gerçek pozitifler}}{\text{Gerçek pozitifler} + \text{Yanlış pozitifler}} \quad (2)$$

Tablo II: Farklı mimarilerin değerlendirme sonuçlarının karşılaştırılması

Model	Sınıf	Kesinlik	Duyarlılık	F1-Score
1 Katman CNN	Aparkat	0.83	0.63	0.71
	Düz Yumruk	0.86	0.67	0.75
	Kroşe	0.55	0.50	0.53
	Servis	0.75	0.92	0.83
	Tekme	0.78	1.00	0.88
2 Katman CNN	Aparkat	0.81	0.81	0.81
	Düz Yumruk	0.78	0.78	0.78
	Kroşe	0.50	0.40	0.44
	Servis	0.87	1.00	0.93
	Tekme	0.93	0.93	0.93
3 Katman CNN	Aparkat	0.93	0.88	0.90
	Düz Yumruk	0.86	0.67	0.75
	Kroşe	0.64	0.70	0.67
	Servis	0.87	1.00	0.93
	Tekme	1.00	1.00	1.00
1 Katman LSTM	Aparkat	0.93	0.81	0.87
	Düz Yumruk	0.80	0.89	0.84
	Kroşe	0.78	0.70	0.74
	Servis	0.72	1.00	0.84
	Tekme	1.00	0.79	0.88
2 Katman LSTM	Aparkat	0.86	0.81	0.83
	Düz Yumruk	0.57	0.88	0.70
	Kroşe	0.60	0.30	0.40
	Servis	0.90	1.00	0.96
	Tekme	1.00	1.00	1.00
2 Katman CNN - 1 Katman LSTM	Aparkat	0.76	0.81	0.79
	Düz Yumruk	0.75	0.67	0.71
	Kroşe	0.55	0.60	0.57
	Servis	0.92	0.92	0.92
	Tekme	0.92	0.86	0.89

ile hesaplanmaktadır. Duyarlılık, pozitif olarak tahmin edilmesi gereken verilerin ne kadarının pozitif olarak tahmin edildiğini gösteren bir metriktir. Modelin bir sınıf için yanlış sınıflandırmaya maruz kalıp kalmadığının belirlenmesine yardımcı olur ve aşağıdaki şekilde hesaplanır.

$$\text{Duyarlılık} = \frac{\text{Gerçek pozitifler}}{\text{Gerçek pozitifler} + \text{Yanlış negatifler}} \quad (3)$$

Son olarak F1-skoru, kesinlik ve duyarlılık değerlerinin harmonik ortalamasıyla hesaplanır. Kesinlik, tahminlerin doğru bulunmasını vurgularken, duyarlılık doğru tahminlerin sayısının artırılmasına öncelik verir. F1-skoru

$$\text{F1-skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (4)$$

ifadesi ile elde edilir.

Farklı 1D-CNN ve LSTM modelleri için elde edilen sonuçlar Tablo II'de sunulmuştur. Bununla birlikte, modellerin doğruluk oranları Tablo III'te gösterilmektedir. Tablo II ve III'te verilen sonuçlar, 1D-CNN'nin doğruluk ve diğer performans metrikleri açısından daha iyi performans sergilediğini göstermektedir. 1D-CNN'nin, yerel örüntüleri yakalama konusunda oldukça etkili olması ve iskelet verilerinde hareketlerin kısa vadeli ilişkilerini yakalayabilmesi belirli hareketlerin veya duruşların tanınması açısından avantaj sağlamaktadır. Özellikle kısa süreli ve ani hareketlerin analizi, 1D-CNN'nin güçlü olduğu alanlardır. İskelet verilerinin doğası gereği, belirli hareketlerin kısa vadeli analizleri daha belirleyici olabilmektedir. Bu nedenle uzun vadeli bağımlılıkların önemli olmadığı veya daha az önemli olduğu durumlarda, 1D-CNN modelleri daha başarılı sonuçlar vermektedir. Ayrıca LSTM tabanlı yaklaşıma göre daha az zaman almakta ve eğitim için

Tablo III: Model sınıflandırma doğruluklarının 100 dönem/epok (epoch) sonrasında karşılaştırılması

Modeller	DO (%)
1 Katman 1D-CNN	75.80
2 Katman 1D-CNN	80.65
3 Katman 1D-CNN	87.09
1 Katman LSTM	83.87
2 Katman LSTM	82.25
2 Katman 1D-CNN - 1 Katman LSTM	79.03

daha az parametre gerektirmektedir, bu da modelin karmaşıklığını azaltmaktadır.

VI. DEĞERLENDİRME

Bu çalışmada, derin öğrenme ile 3B iskelet verileri üzerinden sporcu hareketlerinin tanınması için bir sistem geliştirilmiştir. Özel bir arayüz ile elde edilen verilerin 1D-CNN ve LSTM modelleri üzerindeki performansı incelenmiştir. Katman sayısının artırılmasıyla 1D-CNN modelinin başarısının olumlu yönde etkilendiğini elde edilen sonuçlarla gösterilmiştir. Benzer bir sonucun LSTM modellerinde olmadığı, katman sayısının artırılmasının başarıyı olumsuz yönde etkilediği tespit edilmiştir. Bu durumun, LSTM modellerinin karmaşıklığının artmasıyla birlikte, öğrenme sürecinde aşırı uyum (overfitting), ezberleme veya modelin veriyi yeterince genelleştirememesi sorunlarından kaynaklandığı düşünülmektedir. Yapılan detaylı çalışmalar sonucunda, 3 katman ReLU aktivasyon fonksiyonu ve Adam en iyileyiciye sahip 1D-CNN'nin en iyi performans gösteren model olduğu görülmüştür. Bunlarla birlikte, veri toplama aşamasında kullanılan Kinect Xbox One sensörünün, siyah renkli giysilerden dolayı iskelet noktalarını belirlemede sorun yaşadığı ve gözükmeyen eklemler için tahminlerde anlamsız verilere neden olduğu tespit edilmiştir. Bu durum, sensörün bazı durumlarda performans düşüklüğü yaşayabileceğini ve veri doğruluğunu olumsuz yönde etkileyebileceğini ortaya koymaktadır.

KAYNAKÇA

- [1] N. Tüfek and O. Özkaya, "A comparative research on human activity recognition using deep learning," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, 2019, pp. 1–4.
- [2] R. Yue, Z. Tian, and S. Du, "Action recognition based on rgb and skeleton data sets: A survey," *Neurocomputing*, vol. 512, pp. 287–306, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231222011596>
- [3] T. Singh and D. K. Vishwakarma, "A deeply coupled convnet for human activity recognition using dynamic and rgb images," *Neural Computing and Applications*, vol. 33, no. 1, pp. 469–485, 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s00521-020-05018-y>
- [4] S.-C. Cheng, K.-F. Hsiao, C.-K. Yang, P.-F. Hsiao, and W.-H. Yu, "A novel unsupervised 3d skeleton detection in rgb-d images for video surveillance," *Multimedia Tools and Applications*, vol. 79, pp. 15 829–15 857, 2020. [Online]. Available: <https://link.springer.com/article/10.1007/s11042-018-6292-y>
- [5] H. Rahmani and M. Bennamoun, "Learning action recognition model from depth and skeleton videos," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [6] S. Agahian, F. Negin, and C. Köse, "Improving bag-of-poses with semi-temporal pose descriptors for skeleton-based action recognition," *The Visual Computer*, vol. 35, pp. 591–607, 2019.
- [7] K. Aouaidjia, B. Sheng, P. Li, J. Kim, and D. D. Feng, "Efficient body motion quantification and similarity evaluation using 3-d joints skeleton coordinates," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 2774–2788, 2021.
- [8] C. Li, Q. Zhong, D. Xie, and S. Pu, "Skeleton-based action recognition with convolutional neural networks," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017, pp. 597–600.
- [9] P. Hristov, "Real-time abnormal human activity detection using 1dcnn-lstm for 3d skeleton data," in *2021 12th National Conference with International Participation (ELECTRONICA)*, 2021, pp. 1–4.
- [10] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using kinect camera," in *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, 2012, pp. 28–32.
- [11] P. Zhang, C. Lan, J. Xing, W. Zeng, J. Xue, and N. Zheng, "View adaptive recurrent neural networks for high performance human action recognition from skeleton data," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [12] A. Barua, D. Fuller, S. Musa, and X. Jiang, "Exploring orientation invariant heuristic features with variant window length of 1d-cnn-lstm in human activity recognition," *Biosensors*, vol. 12, no. 7, 2022. [Online]. Available: <https://www.mdpi.com/2079-6374/12/7/549>
- [13] J. Zhu, H. Chen, and W. Ye, "Classification of human activities based on radar signals using 1d-cnn and lstm," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1–5.
- [14] M. G. Ragab, S. J. Abdulkadir, and N. Aziz, "Random search one dimensional cnn for human activity recognition," in *2020 International Conference on Computational Intelligence (ICCI)*, 2020, pp. 86–91.
- [15] W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, and M. Blumenstein, "Rethinking 1d-cnn for time series classification: a stronger baseline," *arXiv preprint arXiv:2002.10061*, 2020.
- [16] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 2017.
- [17] R. Maurya, T. H. Teo, S. H. Chua, H.-C. Chow, and I.-C. Wey, "Complex human activities recognition based on high performance 1d cnn model," in *2022 IEEE 15th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc)*, 2022, pp. 330–336.
- [18] A. K. Özcanlı and M. Baysal, "Islanding detection in microgrid using deep learning based on 1d cnn and cnn-lstm networks," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100839, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352467722001230>
- [19] Z. Wenshu, L. Yuping, W. Yujin, L. Ruilan, L. Daolun, Y. Shan, and X. Yanmei, "Forecasting monthly gas field production based on the cnn-lstm model," *Energy*, vol. 260, p. 124889, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544222017923>
- [20] S. Deep and X. Zheng, "Hybrid model featuring cnn and lstm architecture for human activity recognition on smartphone sensor data," in *2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2019, pp. 259–264.
- [21] K. Chen, Y. Zhou, and F. Dai, "A lstm-based method for stock returns prediction: A case study of china stock market," in *2015 IEEE International Conference on Big Data (Big Data)*, 2015, pp. 2823–2824.
- [22] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A deep learning-based phishing detection system using cnn, lstm, and lstm-cnn," *Electronics*, vol. 12, no. 1, 2023. [Online]. Available: <https://www.mdpi.com/2079-9292/12/1/232>
- [23] R. Yazdani, O. Ruwase, M. Zhang, Y. He, J.-M. Arnau, and A. González, "Lstm-sharp: An adaptable, energy-efficient hardware accelerator for long short-term memory," *arXiv preprint arXiv:1911.01258*, 2019.
- [24] L. Qi, "Using skeleton information for human identification for elderly care and alarm system with the xbox one kinect sensor," Master's thesis, University of Oslo, 2017. [Online]. Available: <http://urn.nb.no/URN:NBN:no-61913>
- [25] A. Corti, S. Giancola, G. Mainetti, and R. Sala, "A metrological characterization of the kinect v2 time-of-flight camera," *Robotics and Autonomous Systems*, vol. 75, pp. 584–594, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889015002195>