

Makine öğrenmesi yöntemi ile dielektron çiftlerinin tanımlanması

Serpil YALÇIN KUZU^{1,*}

¹Fırat Üniversitesi Fen Fakültesi Fizik Bölümü, Elazığ

Geliş Tarihi (Received Date): 31.08.2021

Kabul Tarihi (Accepted Date): 01.12.2021

Öz

Dielektronlar olarak adlandırılan elektron (e^-) pozitron (e^+) çiftleri, evrenin oluşumunu anlamak için yapılan yüksek enerjili parçacık çarpışma deneylerinin çeşitli süreçlerinde üretilen elektromanyetik sinyallerdir. Bu parçacık çiftleri, güçlü kuvvet etkileşimi yapmamaları sebebiyle buldukları ortamın özelliklerinden etkilenmezler ve böylece çeşitli üretim mekanizmaları ile ilgili önemli bilgi sağlarlar. Dielektronları ölçmek için yüksek saflıkta çift sinyalleri gereklidir. Bu sinyalleri, kendisinden çok daha büyük olan arka plan (fon) kaynaklarından (e^+e^+ , e^-e^-) ayırt etmek için karmaşık analiz teknikleri gereklidir. Geleneksel parçacık analiz yöntemleri ile dielektron çiftleri yüksek sistematik belirsizlikler ile üretilir. Son yıllarda çeşitli alanlardaki yapay zeka (AI) uygulamaları, insan çabalarının hızını, doğruluğunu ve verimliliğini artırmak için önem kazanmaktadır. Bu çalışmada dielektron analizinde yapay zeka tabanlı makine öğrenmesi yaklaşımı kullanılmıştır. Çalışmada rastgele orman (RO) sınıflandırma yöntemi Büyük Hadron Çarpıştırıcısı 2010 yılı verisinde bulunan dielektronların elde edilmesine uygulanmıştır. Yapılan çalışmada %90.9 duyarlılık ve %92.0 kesinlik ile RO metodu uygulanmış dielektron analizleri %98.2 başarı göstermiştir.

Anahtar kelimeler: Dielektron, rastgele orman sınıflandırma yöntemi, makine öğrenimi.

Identification of dielectron pairs with machine learning method

Abstract

Dielectrons, electron (e^-) positron (e^+) pairs, are electromagnetic signals produced in various processes of high-energy particle collision experiments to understand the formation of the universe. Since these particle pairs do not interact strongly, they are not affected by the features of their environment. Therefore, they provide significant

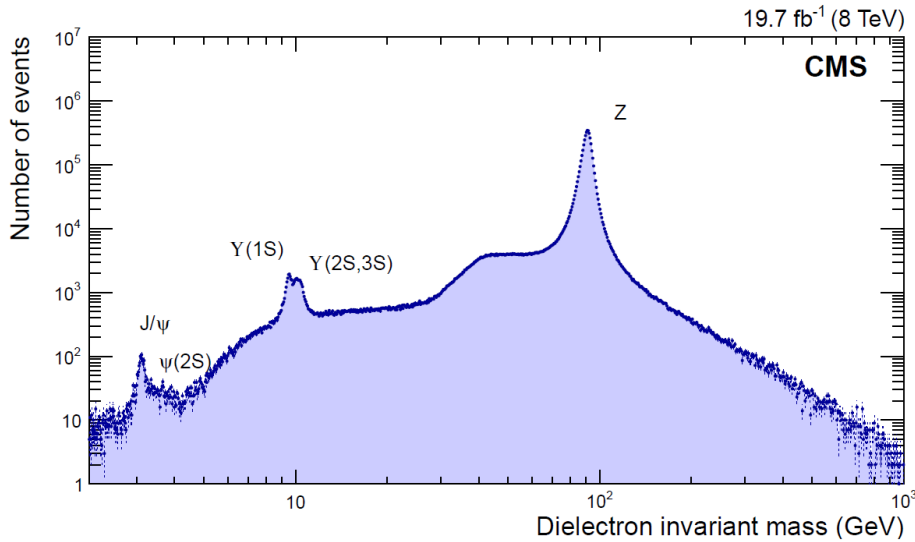
* Serpil YALÇIN KUZU, skuzu@firat.edu.tr, <https://orcid.org/0000-0001-8905-8089>

information about various production mechanisms. High purity pair signals are needed to measure dielectrons. Complex analysis techniques are required to distinguish these signals from much larger background sources (e^+e^+ , e^-e^-). With conventional particle analysis methods, dielectron pairs are produced with high systematic uncertainties. In recent years, artificial intelligence (AI) applications in various fields have gained importance to increase the speed, accuracy and efficiency of human labors. In this study, artificial intelligence-based machine learning approach was used in dielectron analysis. In the study, the random forest (RO) classification method was applied to obtain dielectrons in the Large Hadron Collider 2010 data. In the study, the RO method applied dielectron analysis showed 98.2% success with 90.9% efficiency and 92.0% precision.

Keywords: Dielectron, random forest classifier, machine learning

1. Giriş

Evrenin oluşum dinamiğini laboratuvar koşullarında anlamak ancak küçük bir 'Büyük Patlama' oluşturularak başarılabilir. Bu amaçla yüksek enerjilerde hadronlar [1] ya da iyonlar çarpıştırılır ve oluşan ortamın özellikleri farklı parçacık türlerinin ölçüm aracı olarak kullanılmasıyla incelenir. Dielektonlar, en önemli ölçüm araçlardan biri olan elektron(e^-)-pozitron(e^+) çiftleridir. Bu lepton [2] çiftleri güçlü kuvvet ile etkileşime girmezler. Bu sebeple üretimleri üzerinde buldukları ortamın etkisi gözlenmez. Böylece çarpışmada üretildikleri mekanizma hakkında bozulmamış bilgi verirler. Çarpışmanın herhangi bir aşamasında üretilmeleri dielektronları sistemin tüm evrelerini araştırmak için kullanmaya olanak sağlar.



Şekil 1. Dielektonların değişmez kütle spektrumu [3]. Histogramda x eksenini dielektron değişmez kütle değerlerini, y eksenini olay sayısını göstermektedir.

Dielektonların değişmez kütle spektrumu Şekil 1'de gösterilmiştir [3]. Çarpışma ortamında hadronik maddenin evrimi, faz geçişi ve ortam sıcaklığı gibi temel özellikleri, dielektron çiftlerinin spektrumu incelenerek belirlenebilir. Ayrıca, dağılımın çeşitli kütle aralıkları, ortamın farklı aşamalarına ve fiziksel özelliklerine duyarlıdır.

60-120 GeV/c² arasındaki çok yüksek kütle bölgesindeki çiftler Z bozonunun bozunma ürünüdür [4]. 3-10 GeV/c² arasındaki kütle bölgesi, sistem evriminin ilk aşaması hakkında bilgi sağlar. Bu evredeki çiftler, Drell-Yan süreci olarak adlandırılan kuark-antikuark etkileşimleri tarafından üretilen sanal fotonun bozunma ürünleridir [5]. Bu bölgede, J/ψ ve Y gibi kuarkonyum bozunmaları, sert saçılma hakkında bilgi veren diğer dielektron kaynaklarıdır. Ara kütle bölgesi olarak adlandırılan 1-3 GeV/c² arasındaki çiftler, sürekli dağılım gösteren termalizasyon aşamasında bulunan D^+D^- meson semileptonik bozunmalarının üretimidir. 0-1 GeV/c² arasındaki düşük kütleli dielektron bölgesindeki çiftler ρ ve ω gibi rezonansların ve Dalitz bozunmalarının ürünüdürler[6].

Dielektron çiftleri, uzun işlem süresi gerektiren düşük sinyal verimliliğine sahip yüksek saflıkta örnekler elde etmek için geleneksel kesim tabanlı yöntemlerle tanımlanabilir. Bu yöntemlerde öncelikle detektörler tarafından belirlenen yük (q), pseudo rapidite (η) [7], azimutal açı (ϕ) [7] ve momentum (p) gibi parçacık bilgilerine bazı veto kesintileri uygulanarak e^- ve e^+ parçacıkları elde edilir [7]. Dielektronların belirlenmesi için çiftlerin momentum (P) [7], ham değişmez kütle (M) [7] ve açılma açısı (α) [7] gibi özellikleri hesaplanır. Dielektron ham değişmez kütle dağılımları foton gibi çeşitli kaynaklar nedeniyle çok geniş bir arka plana sahiptirler [8-9]. Bu sebeple ham değişmez kütle spektrumunda bulunan fon dağılımları karışık olay (mixed event) veya benzeri işaret (like sign) tekniklerinin uygulanmasıyla tanımlanır [8-12]. Yöntemlerden herhangi biri ile belirlenen arka plan dağılımları ham kütle spektrumlarından oldukça büyük olmaları sebebiyle normalize edilirler [9]. Normalizasyon sonucu elde edilen fonların ham dağılımlardan çıkartılmasıyla dielektron değişmez kütle spektrumu elde edilir [9]. Veto kriterlerinin uygulanması, karışık olay veya benzer olay ile arka plan tanımlanması ve normalizasyon gibi işlemlerin yapılması sistematik belirsizlik kaynaklarıdır [8-9] ve çift bilgilerinin kaybolmasına sebep olabilir. Dolayısıyla kesim tabanlı yöntemlerin kullanılması uzun zaman almasının yanında yüksek sistematik belirsizliklere ve çarpışma sonrası oluşan ortamın dinamikleri hakkında bilgi kaybına neden olur [8-9, 13]. Yapay zeka (AI) tabanlı makine öğrenme tekniklerinde klasik kesimler yerine parçacıkların genel özellikleri kullanılarak veri analizi yapılmaktadır. Kesme değerleri bulunmamaktadır. Bu sebeple makine öğrenmesi yöntemiyle sistematik belirsizlik problemleri aşılabilir.

Parçacık fiziği analizlerine makine öğrenme tekniklerinin uygulanması, parçacık sinyali ve fon arasında girişime neden olan kuantum mekaniğinin doğası nedeniyle diğer araştırma alanlarındaki kullanımına kıyasla zordur [14]. Higgs bozonu oluşumu [15-16] ve CERN Büyük Hadron Çarpıştırıcısında (LHC) [17] ışın dinamiği analizi gibi fizik sonuçlarını geliştirmek için farklı makine öğrenme yaklaşımları çalışılmıştır. Denetimli makine öğrenmesi yöntemlerinden biri olan rastgele orman (RO) sınıflandırması [18] ile parçacık tanımlamasının [19] ve veri kalitesi belirlemenin [20] mümkün olduğu ortaya çıkmıştır. Bu nedenle, bu çalışmada klasik kesim tabanlı analiz yöntemlerine alternatif bir metot geliştirmek için RO sınıflandırması dielektron değişmez kütle spektrumunun belirlenmesinde kullanılmıştır.

1.1. Denetimli makine öğrenmesi

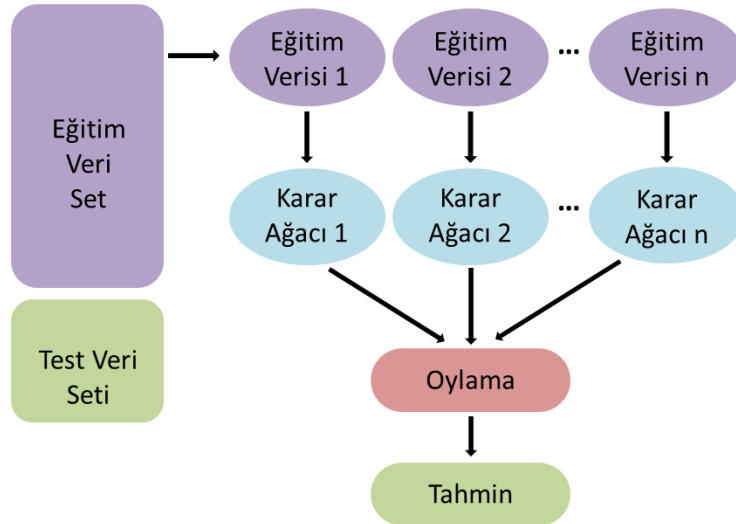
Denetimli makine öğrenmesi [21], sınıflandırma ve regresyon için kullanılan başarılı makine öğrenmesi tekniklerinden biridir. Bu teknik iki ya da daha fazla sınıf ayrımı yapmak için kullanılabilir. Yaklaşımın amacı, doğru ve genellenebilir sonuçlar veren bir model geliştirmektir. İlk olarak eğitim veri seti üzerine bir model uygulanır.

Genelleme yapmak için model test veri seti olarak tanımlanan eğitim seti ile aynı özelliklere sahip daha önceden görülmemiş verilere uygulanır. Veri setlerinin yeterince ortak noktası varsa, modelin doğru olduğu sonucuna varılabilir ve sonuçlar eğitimden test verilerine genellenebilir [21]. Bu yaklaşımda, bir modelin seçimi ve uygulanması için iki anahtar kavram vardır. Bunlar 'aşırı' ve 'eksik' öğrenmedir. Veri bilgisi miktarı ile karşılaştırıldığında, seçilen model çok karmaşık ve eğitim setinde iyi çalışıp ancak testi genelleştirmiyorsa bu durum aşırı öğrenmedir. Tercih edilen modelin basit olması, eksik öğrenme olarak adlandırılan eğitim setinde daha az öğrenme ile sonuçlanır ve model verilerin tüm yönlerini yakalayabilir [21].

1.1.1. Rastgele orman (RO) sınıflandırma yöntemi

RO sınıflandırma yöntemi [18], karar ağaçlarından oluşan denetimli öğrenme modellerinden biridir. Modelde kararlar, bir dizi alt karara göre üretilir. Her karar ağacı, sırasıyla özellikleri ve kararları temsil eden düğümlere ve yapraklara sahiptir. Düğümlere, geçerli ağacı eğitmek için uygulanan özelliklerden en iyilerinin seçilmesiyle karar verilir. Her bir özellik için düğüm ayırma kalitesi, entropi kazancı veya Gini indeksi ile bulunabilir [19]. RO sınıflandırmasının oluşturma ve tahmin olmak üzere iki bölümü vardır. İlk olarak model, eğitim veri setinin rastgele seçilmiş bir alt kümesinden ağaçlar üretir. İkinci aşamada test veri setinin nihai kararını vermek için farklı karar ağaçlarından oylar alır [18-24]. Sınıflandırma algoritması Şekil 2'de şematik gösterilmiştir.

RO sınıflandırıcısının diğer modellere göre birçok avantajı vardır. Makine öğrenimi modellerinde temel sorunlardan biri aşırı öğrenme sonucu modelin eğitim veri kümesi dışındaki veriler için doğru sınıflandırma yapamamasıdır. Karar ağaçları ormanına sahip olması nedeniyle RO yöntemi aşırı öğrenmeye karşı dirençlidir. Ayrıca modelin yorumlanabilir olmasını sağlayan ve özelliklerle kararlar arasındaki ilişkiyi ortaya çıkaran her bir özelliğin önemi hesaplanabilir.

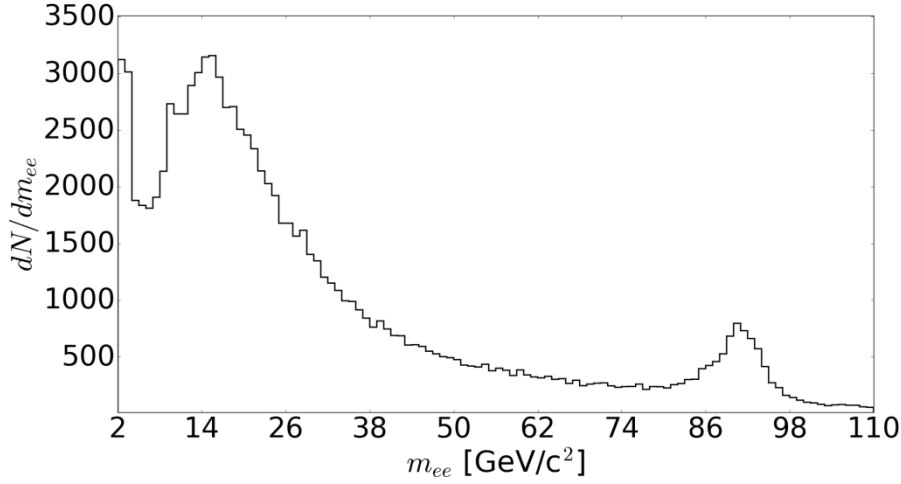


Şekil 2. RO sınıflandırma modelinin şematik gösterimi.

2. RO Modeli ile Dielektron Analizi

2.1. Veri seti

Kompakt Müon Solenoit Deneyi (CMS) [3] 2-110 GeV/c² enerji aralığı için e^-e^- , e^+e^+ ve dielektronlardan oluşan veri seti bu çalışmada kullanılmıştır [25]. Toplam 99912 olay, scikit-learn paketi [26] tarafından sağlanan RO sınıflandırma yönteminin Python uygulaması ile analiz edilmiştir. Tüm çiftlerin ham değişmez kütle spektrumu Şekil 3'de sunulmuştur.



Şekil 3. Analizde kullanılan e^+e^+ , e^-e^- ve dielektron çiftlerinin ham değişmez kütle dağılımı. Histogramda x eksenini çiftler için ham kütle değerlerini, y eksenini 1 GeV/c² kütleye karşılık gelen çift sayısını göstermektedir.

Veri setinde her bir çifte ait parçacığın yükü (q), momentum bileşenleri (p_x , p_y , p_z), enine momentumu (p_t) [7], enerjisi (E), η ve ϕ bilgileri bulunmaktadır. Ayrıca her bir çifte ait kütle değeri (M)'de mevcuttur. Sınıflandırma yönteminde özellik olarak kullanılmak üzere çiftlerin α , P ve her bir parçacığın p değerleri hesaplanmıştır. Farklı korelasyonları çözmek için ek özellik olarak $\Delta\eta$ ve $\Delta\phi$ farkları hesaplanmıştır [27].

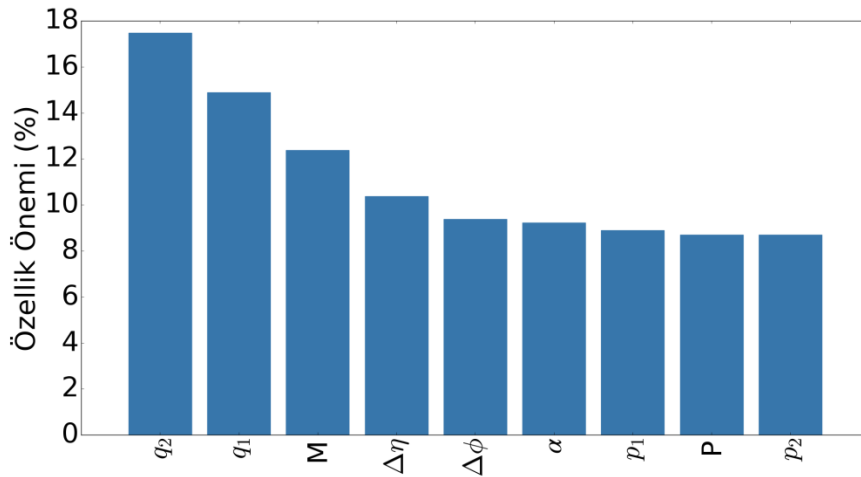
2.2. Modelin uygulanması

RO sınıflandırma modeli ile dielektron çiftlerinin tanımlanması incelenmiştir. Çiftlerin sinyal sınıfı klasik analizlerde olduğu gibi her bir çiftin zıt yüklere sahip olmasına bakılarak tanımlanmıştır. Bu sınıflandırma ile veri setinin %57.02 dielektronlardan ve %42.98 arka plan çiftlerinden (e^+e^+ , e^-e^-) oluşmaktadır. Sinyal ve arka plan olaylarının sayısını içeren analiz yöntemi ve sinyal seçim kriteri Tablo 1'de özetlenmiştir. RO ile çiftleri belirlemek için klasik dielektron analizlerinde kullanılan özellikler kullanılmıştır [8-10, 13]. Kesim tabanlı analizlerde veto kriterlerinin uygulandığı p , η ve ϕ gibi parametreler herhangi bir kesinti uygulamadan modelde çiftleri belirleyici özellik olarak kullanılmıştır. Geleneksel metotlarda dielektron tanımlanmasında önemli role sahip olan q , M , α ve P bilgileri de modelde karar özellikleri olarak kullanılmıştır [7]. RO modelinde her bir özelliğin önemi, eğitim veri setinde kesin olarak bilinen dielektron çiftlerinin özellik analizi sonucunda belirlenmiş olup Şekil 4'te sunulmuştur [18, 21, 26].

Tablo 1. RO analizinde kullanılan sinyal ve arka plan olay sayıları.

Analiz	Sinyal Tanımı	Sinyal	Fon
Dielektron Çifti	Birbirine zıt yüklere sahip çiftler	56968	42944

Bu çalışmada verilerin %75'i eğitim, kalanı ise test veri seti olarak seçilmiştir. Tahmin performansını artırmak için modelin iki temel parametresi olan ormandaki ağaç sayısı ve her bir ağacın maksimum derinliği optimize edilmiştir. Ayrıca her bir özellik için düğümde bölünme kalitesi *Gini indeksi* ile değerlendirilmiş ve maksimum özellik parametresi otomatik olarak seçilmiştir. Uygulanan RO metodunda her bir özelliğin dielektron tahminlerindeki önemi Şekil 4'te gösterilmiştir. Klasik parçacık analizinde olduğu gibi RO yöntemi veri seti için karar verirken öncelikle parçacık çiftlerinin yüklerini (q_1 , q_2) ve birlikte oluşturdukları kütle (M) göz önüne almıştır. Geleneksel yöntemlerden farklı olarak parçacıkların çarpışma geometrisinden dolayı sahip oldukları momentum ve açılma açısı gibi kinematik özellikleri çifti tanımlamada benzer oranda etki etmiştir.



Şekil 4. RO modelinde dielektron tanımlaması için kullanılan her bir özelliğin önemi.

3. Model Değerlendirilmesi

3.1. RO model performans değerlendirilmesi

RO model performansı bir parçacık tanımlama metriği olan sinyal duyarlılığı ve makine öğrenimi metriklerinden kesinlik, F1 skor ve ROC eğrisi [28] altındaki alan (ROC-AUC) [28] ile değerlendirilmiştir. Sinyal duyarlılığı, kesinlik ve F1 skor değerleri modelin hata matrisinin [21, 26] hesaplanması sonucu elde edilen parametrelerin kullanılmasıyla belirlenir. Doğru tahmin edilmiş dielektron çiftlerinin yüzdesini ölçmek için kullanılan sinyal duyarlılığı:

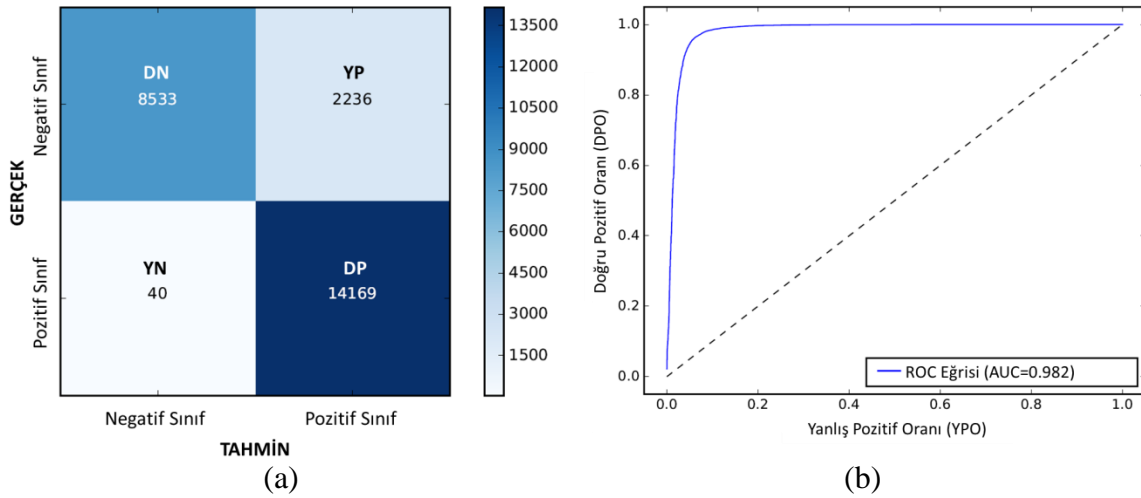
$$\text{Sinyal Duyarlılığı} = DP/(DP + YN) \quad (1)$$

denklemi ile tanımlanır. Denklem (1)' de DP ve YN sırasıyla doğru sınıflandırılmış sinyal sayısını ve yanlış sınıflandırılmış arka planı temsil eder. Kesinlik

$$\text{Kesinlik} = DP/(DP + YP) \quad (2)$$

denklemini kullanarak yapılan tahminler için modelin doğruluğunu tanımlar. Denklem (2) ' de YP yanlış tanımlanmış dielektron sinyallerini temsil eder. Sinyal duyarlılığı ve kesinliğin harmonik ortalaması olan F1 skor ile modelin iki yönlü değerlendirilmesi yapılabilir. RO modeli hata matrisi DP, YN, YP ve doğru sınıflandırılmış arka planı gösteren DN değerleri ile Şekil 5 (a)' da gösterilmiştir. Hata matrisinde negatif sınıf arka planı ve pozitif sınıf dielektron sinyalini temsil etmektedir.

ROC eğrisi makine öğreniminin en önemli metriklerinden biri olup, modelin yanlış sınıflandırdığı sinyal oranı ile gerçek sinyal oranı ilişkisini gösterir. ROC eğrisinin altındaki alan (ROC-AUC), bir parametrenin iki tanı sınıfının ne kadar iyi ayırt edebileceğinin bir ölçüsüdür. Yüksek ROC-AUC değeri, modelin sınıfları ayırt etme yeteneğinde ki artışı belirler. RO modeli ROC eğrisi Şekil 5 (b)' de gösterilmiştir. ROC eğrisinde y eksenini doğru pozitif oranı (DPO) olarak da tanımlanan sinyal duyarlılığını, x eksenini YP/(YP+DN) olarak hesaplanan yanlış tanımlanmış dielektron oranını (yanlış pozitif oranı-YPO) göstermektedir. DPO ile doğru tahmin edilen dielektron sinyallerinin yüzdesi ve YPO ile yanlış tahmin edilen arka plan yüzdesi elde edilir. Modelin sinyal ve fon ağırlıkları göz önüne alınarak ortalama sinyal duyarlılığı, kesinliği ve F1 skoru hesaplanmıştır [26]. Bu veriler ROC-AUC değeri ile birlikte Tablo 2'de listelenmiştir. RO sınıflandırma yöntemi dielektron çiftlerini %92.0 kesinlik ve %90.9 sinyal duyarlılığı ile belirlemiştir. Modelin başarısı %98.2'dir.



Şekil 5. RO modeli hata matrisi (a) ve ROC eğrisi (b).

Tablo 2. RO sınıflandırma yöntemi kesinlik, sinyal duyarlılığı, F1 skor ve ROC-AUC değerleri.

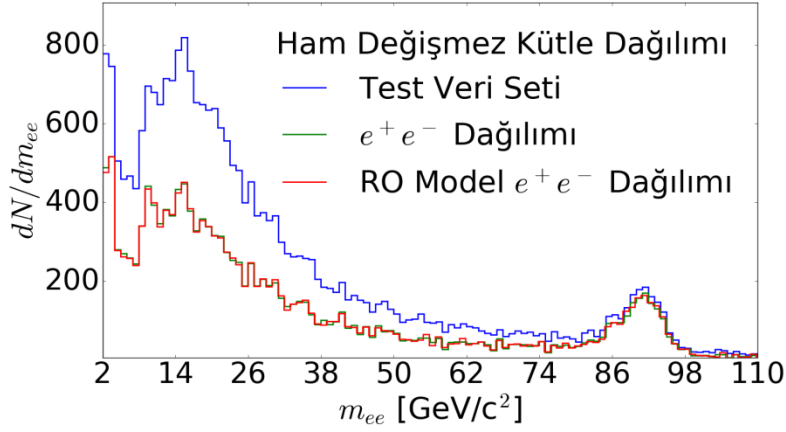
Analiz	Kesinlik	Sinyal Duyarlılığı	F1 Skor	ROC-AUC
Dielektron Çifti	0.920	0.909	0.907	0.982

3.1. Öngörülen dielektron kütle spektrumu

Parçacık fiziği analizinin zorlu kısmı, birbiriyle örtüşen parçacık sinyalini çok daha hacimli arka plandan çıkarmaktır. Bu çalışmada RO sınıflandırma modeli uygulanarak net dielektron sinyalleri belirlenmiştir. Test veri seti için ham değişmez kütle dağılımı ölçülen ve model ile belirlenmiş dielektron çiftlerinin değişmez kütle dağılımı ile Şekil 6'da gösterilmiştir.

3. Sonuçlar

Bu çalışmada, dielektron çiftlerinin belirlenmesinde makine öğrenimi modellerinden biri olan RO sınıflandırma yöntemi kullanılmıştır. RO modeli rastgele karar ağaçları ormanından oluşumu ile makine uygulamalarında sıkça karşılaşılan aşırı ve eksik öğrenme problemlerinin üstesinden geldiği için tercih edilmiştir. Yürütülmüş analizler, klasik yüklü parçacık çifti tanımlama teknikleri ile uyumludur. Dielektron çiftlerinin belirlenmesinde model öncelikli olarak her bir parçacığın zıt yük ilişkisini incelemiştir. Sınıflandırıcının performansını anlamak için hem parçacık hem de makine öğrenimi değerlendirme metrikleri kullanılmıştır. RO modelinin başarısı %98.2'dir. Model e^+e^- çiftlerini %92.0 kesinlik ve %90.9 sinyal duyarlılığı ile tespit etmiştir. Elde edilen sonuçlar RO sınıflandırma metodu ile nükleer, parçacık ve yüksek enerji fiziği çalışmalarının zorlayıcı ve yüksek hesaplamalı analiz teknikleri kullanılmadan yapılabileceğini göstermiştir.



Şekil 6. Test veri seti için ham değişmez kütle dağılımı ile ölçülen ve RO modeli ile öngörölmüş dielektron çiftlerinin değişmez kütle dağılımı. Histogramlarda x eksenini çiftler için ham kütle değerlerini, y eksenini 1 GeV/c^2 kütleyle karşılık gelen çift sayısını göstermektedir. Histogramlarda ham dağılım mavi, ölçölmüş dielektron dağılımı yeşil ve model tarafından belirlenmiş dielektron dağılımı kırmızı ile gösterilmiştir.

Teşekkür

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) 119F302 numaralı proje ile desteklenmiştir. Ayben Karasu Uysal'a faydalı önerileri için özel teşekkür ederim.

Kaynaklar

- [1] Wong, S. S. M., **Introductory nuclear physics**, 29, Wiley-VCH Press, Weinheim, (2004).
- [2] Griffiths, D. J., **Introduction to elementary particles**, 74, Wiley-VCH Press, Weinheim, (2008).

- [3] CMS Collab., Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}=8$ TeV, **JINST**, 10, P06005, (2015).
- [4] CMS Collab, Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV, **JHEP**, 10, 132, (2011).
- [5] Drell, S. D. ve Yan, T. M., Massive lepton-pair production in hadron-hadron collisions at high energies, **Physical Review Letters**, 25, 316, (1970).
- [6] Biernat, J., Measuring di-electron Dalitz decays of baryon resonances with HADES and PANDA, Doktora Tezi, Jagiellonian Üniversitesi, Nükleer Fizik Enstitüsü, Krakow, (2017).
- [7] Nourbakhsh, S., Studio degli eventi J/ψ in due elettroni con i primi dati di CMS, Doktora Tezi, Roma La Sapienza Üniversitesi, Matematik, Fizik ve Doğa Bilimleri Fakültesi, Roma, (2010).
- [8] ALICE Collab., Dielectron and heavy-quark production in inelastic and high-multiplicity proton-proton collisions at $\sqrt{s}=13$ TeV, **Physics Letters B**, 788, 505, (2019).
- [9] STAR Collab., Measurements of Dielectron Production in Au+Au Collisions at $\sqrt{s_{NN}}=200$ GeV from the STAR Experiment, **Phys. Rev. C**, 92, 024912, (2015).
- [10] ALICE Collab., Measurement of dielectron production in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, **Phys. Rev. C**, 99, 024002, (2019).
- [11] STAR Collab., J/ψ production at high transverse momenta in p+p and Cu+Cu collisions at $\sqrt{s_{NN}}=200$ GeV, **Phys. Rev. C**, 80, 041902, (2009).
- [12] ALICE Collab., J/ψ production as a function of charged-particle pseudorapidity density in p-Pb collisions at $\sqrt{s_{NN}}=5.02$ TeV, **Physics Letters B**, 776, 91, (2018).
- [13] ALICE Collab., Dielectron production in proton-proton collisions at $\sqrt{s}=7$ TeV, **JHEP**, 64, 1809, (2018).
- [14] Schwartz M. D., Modern Machine Learning and Particle Physics. arXiv:2103.12226, 2021.
- [15] Chen T., He T., Higgs Boson Discovery with Boosted Trees, **Proceedings of the 2014 International Conference on High-Energy Physics and Machine Learning**, 69-80, Montreal, (2014).
- [16] CMS Collab., Machine Learning Techniques in the CMS Search for Higgs Decays to Dimuons, **Proceedings of 23rd International Conference on Computing in High Energy and Nuclear Physics**, 06002, Sofia, (2019).
- [17] Arpaia P., Azzopardi G., Blanc F., Bregliozzi G., Buffat X., Coyle L., et al., Machine learning for beam dynamics studies at the CERN Large Hadron Collider, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 985, 164652, (2021).
- [18] Breiman, L., Random Forests, **Machine Learning**, 45, 5–32, (2001).
- [19] Trzcinski, T., Graczykowski, L. K. ve Glinka, M., Using Random Forest Classifier for particle identification in the ALICE Experiment, **Proceedings of Information Technology, Systems Research and Computational Physics**, 3-17, Krakow, (2019).
- [20] Trzcinski T. and Deja K., Assigning Quality Labels in the High-energy Physics Experiment ALICE Using Machine Learning Algorithms, **Proceedings of NICA days**, 647-655, Warsaw, (2017).
- [21] Müller A. C. ve Guido, S., **Introduction to Machine Learning with Python**, 28-30, O'Reilly Media Inc., Sebastopol CA, (2016).

- [22] Azhari, M., Alaoui, A., Achraoui, Z., Ettaki, B. ve Zerouaoui, J., Adaptation of the Random Forest Method, **Proceedings of the 4th International Conference on Smart City Applications - SCA '19**, 1141–1146, Warsaw, (2019).
- [23] Azhari, M., Alaoui, A., Abarda, A., Ettaki, B. ve Zerouaoui, J., **Big Data and Networks Technologies**, 183-189, Springer 81, (2020).
- [24] Azhari, M., Alaoui, A., Abarda A., Ettaki, B. ve Zerouaoui, J., A Comparison of Random Forest Methods for Solving the Problem of Pulsar Search, **Proceedings of the Fourth International Conference on Smart City Applications**, 1-6, Cham, (2020).
- [25] McCauley, T., Events with two electrons from 2010, CERN Open Data Portal, (2014). <https://opendata.cern.ch/record/304>, (30.08.2021).
- [26] Pedregosa, F., Varoquaux G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. ve Duchesnay, E., Scikit-learn: Machine learning in Python, **Journal of Machine Learning Research**, 12, 2825-2830, (2011).
- [27] NA61/SHINE Collab., Two-particle correlations in azimuthal angle and pseudorapidity in inelastic p+p interactions at the CERN Super Proton Synchrotron, **Eur. Phys. J. C.**, **77**, 59, (2017).
- [28] Bradley, A. P., The use of the area under the ROC curve in the evaluation of machine learning algorithms, **Pattern Recognition**, 30, 1145-1159, (1997).