TÜBİTAK

# Büyük Hadron Çarpıştırıcısındaki İleri Veri Çözümleme Teknikleri İçin Derin Öğrenim Metotları

**Program Kodu:** 3001

**Proje No:** 116F026

Proje Yürütücüsü:
**Doç. Dr. M. Altan ÇAKIR**

Bursiyerler:
İlknur Köseoğlu
Sevda Yeliz Nar

MAYIS 2018

İSTANBUL

# ÖNSÖZ

Proje süresince proje ekibi aşağıda yer alan yayın, sözlü sunum, bildiri, çalışta ve araştırmalara katılım gibi çalışmaların yanı sıra, araştırmalarda kullanılmak üzere TUBİTAK veya herhangi bir devlet kurumunun desteği dışında yapılandırılmış bir araştırma laboratuvarı ve proje kapsamında alınmış ödül ve burslar bulunmaktadır. Aynı zamanda proje yöneticisi bu projeye paralel olarak Türkiye'nin önde gelen şirketlerinde akademi-sanayi iş-birliği kapsamında tanımlanan TEYDEB projelerine danışmanlık yapmakta ve bu proje kapmasında yapılan araştırma alt-yapısını bu çalışmalarda başarı ile entegre ederek çalışmaktadır. Bu çerçevede Türkiye'de araştırmalarına yeni başlayan araştırmacılara destek olmak üzere yapılandırılmış tarafıma verilen 116F026 numaralı 3001 başlangıç ar-ge desteği, bilimsel çalışmalar ile birlikte yürütülen akademi-sanayi iş-birliğini de içine alacak şekilde başarı ile tamamlanmıştır. Projede yer alan iş paketleri özellikle 3. Bölümde yer alan gereç ve yöntemler kısmında anlatılmış, sonuç ve bulgular kısmında proje iş paketleri ve proje kapsamında elde edilen başlangıç ar-ge gelişimi için yapılan yazılım araçları paralelinde yapılan diğer akademi dışı çalışmalar detaylı olarak özetlenmiştir.

Bilimsel çalışmalarımız:

- 1 tane yayınlanmış makaleye katkı
- 2 tane konferans özeti (CERN Makalesi – Physics Analysis Summary)
- 1 tane uluslararası indekslerde kayıtlı bildiri
- 2 tane tamamlanmak üzere olan makale
- 1 tane yüksek lisans ve 2 tane lisans bitirme tezi yayınlanmıştır

Proje kapsamında alınan makine teçhizat desteği (Intel-Tarentum Yapay Zekâ):

- Çalışma sunucusu ve ekipmanları: Intel Power Edge R620 Rock Server, Raid SSD disk hafıza sunucusu, Rock server kabini, HP ağ anahtarları. Veri seti eğitimleri iş koşturmaları, paralel programlama, GPU tabanlı resim yapısal inceleme ve ileri sayısal hesaplayıcılar
- 8 adet dizüstü Lenovo G50-i7 çalışma bilgisayarları
- D-link gözetleme (surveillance) kamerası: 1 adet ek profesyonel kamera cihazı, 3 adet Intel akıllı kızılötesi kamera
- 4 adet Intel Galileo geliştirme kartı ve IoT çalışma kitleri: Özellikle sensörlerin bir araya getirilmesi ve uygulama çalışmalarında kullanılmaktadır
- 1 Adet NVidia Jetson TX2 GPU Gömülü Geliştirme kiti: Yapay zeka uygulamaları kapsamında görüntü ve ses işleme, gözetleme kamerasında gelen görüntülerin işlenmesi ve öğrenilmesi, sensör çözümlemeleri ve büyük veri akışını sağlama modülleri çalışmaları yürütülmektedir.
- 1 Adet LG marka 42" televizyon
- 6 kişilik profesyonel çalışma masaları ve sandalyeleri

Proje kapsamında yaptığımız çalışmalar uyarınca alınan ödüller:

- Proje yürütücüsü: "Fermi National Laborotary (Fermilab) Guest Visitor Faculty Fellowship." Fermi araştırma laboratuvarı 9 aylık araştırma ödülü

- Proje öğrencisi: “Fermi National Laborotary (Fermilab) Student Scholarship.” Fermi araştırma laboratuvarı 3 aylık araştırma bursu
- Proje öğrencisi: CERN School of Computing araştırma ve katılım bursu
- Proje öğrencisi: CERN-TAEK araştırma bursu ile CERN-CMS deneyinde araştırmalar yapmak üzere, 3 aylık burs.
- İlknur Köseoğlu yurt dışına Almanya’da yer alan bir üniversite tarafından sağlanan doktora araştırma bursu alması

Proje kapsamında akademi-sanayi iş-birliği ve proje konusu ile alakalı alınmış projeler:

- Proje yürütücüsü proje çerçevesinde uzmanlığını kullanarak yazdığı 1501 TEYDEB programı kapsamında Hürriyet Gazetecilik ve Medya A.Ş kapsamında proje almaya hak kazanmış ve proje benzeri konular çerçevesinde danışmanlık yapmaktadır.
- Proje yürütücüsü benzer konular çerçevesinde iki ayrı TEYDEB projesi yazmış ve hakem süreçleri tamamlanmıştır. Bu projelerde ayrıca yakın zamanda işleme konularak benzer konu başlıklarının sanayi de olan uygulamalarına devam edilecektir.
- Journal of Cognitive Science dergisinde editörlük

Proje kapsamında üniversite içi ev dışı etkinliklere katılım:

- Proje yürütücüsü bu proje kapsamında İstanbul Teknik Üniversitesi ve Kalkınma Bakanlığı çerçevesinde kurulacak iki ayrı merkez, Büyük Veri ve Yapay Zekâ, için yapılan kurulda yer almakta ve projelere katkılarda bulunmaktadır.

Proje süresince proje yürütücüsü birçok akademi-sanayi kuruluşlarında proje içeriğini ve teknikleri içeren önemli konuşmaları ‘KeyNote’ konuşmacı olarak vermiştir.

**Önsöz** ii-iii

**İçindekiler** iv

**Özet** v

**Abstract** vi

**1. Giriş** 7

**2. Literatür Özeti** 10

3. **Gereç ve Yöntem** 12

3.1. Monte-Carlo (MC) simülasyonlarının (sinyal ve ardalan) ortak çalışma sunucusuna kurulumu ve kullanıma hazır hale getirilmesi ve kıyaslanması
3.2. Detektör simülasyonu ve MC'dan üretilen kuramsal modellerin tekrar oluşturulması
3.3. İleri veri çözümleme tekniklerinin uygulanması için gerekli veri çözümleme programları
3.4. Sinyal/Ardalan veri çözümleme uygulamaları makine öğrenimi ve yapar sinir ağları kullanımı

**4. Bulgular** 18

4.1. CMS Kolaborasyonun da Direk Stop Araştırmaları
4.2. Peltier Tabanlı Bulut Odası ve Parçacık İzleri Sınıflandırma Çalışmaları
4.3. Derin Öğrenme Tekniklerinin Web Analitiğine Uygulanması
4.4. LHC'de Tek Lepton Kanalında Direk Stop Araştırmaları ve Derin Öğrenme Tekniklerinin Uygulanması
4.5. TTbar Çiftlerinin Spin Korelasyonu ve Polarizasyon bilgilerinin İki Lepton Sonlu Durumunda CMS Deneyinde Hassas Ölçümlenmesi ve Yeni Fizik Arayışları

**5. Sonuç** 33

**6. Kaynakça** 36

# ÖZET

Günümüz parçacık fiziği bilimsel araştırmalarında yaygın olarak kullanılan ileri veri çözümleme teknikleri mikroskobik boyutta oldukça karmaşık fiziksel etkileşimlerin anlaşılması ve yeni parçacıkların keşfedilmesini için çeşitli öneriler sunmaktadır. Bu veri çözümleme teknikleri günümüzde 'Big Data – Büyük Veri' başlığı gibi oldukça önem arz eden ve gündelik hayatımızda yaygın olarak kayıt altına alınmaya başlanan veri akışının işlenmesi, anlaşılması ve uygulamaları konuları çerçevesinde benzerlikler göstermektedir. Bu projede, oluşturulması planlanan çoklu değişken metotları ve yapar sinir ağların teknikleri aracılığı ile Büyük Hadron Çarpıştırıcısında 13 TeV ve zamanla 14 TeV merkez çarpışma enerjisindeki etkileşimler içerisinde bulunması beklenen Standart Model ötesi yeni fizik modellerinin bulunması ve bu çerçevedeki iler seviye parçacık fiziği veri uygulamaları düşünülmektedir. Proje çerçevesinde lisansüstü öğrencilerin parçacık fiziği üzerine ileri veri çözümleme metotları üzerine uzmanlaşmalarının sağlanması, projenin devamında da burada öğrenilen tekniklerin veri işleme, bilişim ve ekonomi alanında uygulamaları için gerekli alt-yapının oluşturulması planlanmaktadır. Projenin önemi, araştırmaya katılması planlanan öğrencilerin bursları ve eğitimlerine destek olmasının yanı sıra, bundan sonraki daha büyük ölçekteki proje ve araştırmalar için ilgili teknik alt-yapı ve eğitim desteğinin yaratılması beklenmektedir.

# ABSTRACT

Advanced data analysis techniques used in particle physics provide the broad understanding of complex processes and interactions at the microscopic level. The processes involved ultimately resulted in the creation of atoms and the complex molecules that led to our existence. The intellectual curiosity embodied in particle physics is also at the foundation of engineering, philosophy, art and other scientic disciplines which, together, have shaped the modern world. The study of particle physics challenges our preconceptions, inspires and seeks to move human knowledge forward at a basic level – wherever that may lead. The Large Hadron Collider (LHC), which collides protons at an energy of 13 TeV, produces hundreds of exabytes of data per year, making it one of the largest sources of data for particle physics meaasurements and searches in the world today. In this project we will concentrate on the advanced multivariate analysis and neural networks techniques, which will be used for background-to-signal decision making, in the context of the standard model and beyond. Furthermore we plan to provide scholarships for graduate students and also to teach the advanced statistical methods which can be ultimately used later on science and technology applications. With this start-up project, the project author targets the infrastructure of the machine learning and neural network techniques for further developments.

# 1. GİRİŞ

Günümüz itibari ile ulaştığımız veri ağının büyüklüğü ve akışı ele alındığında, büyük ölçekteki verinin işlenmesi ve anlaşılması oldukça önemli duruma gelmektedir. Bu çerçevede bilimsel araştırmalarda kullanılan veri çözümleme teknikleri ve bu alanlarda kullanılan çeşitli yazılımlar, başta karmaşık fizik ve matematik problemlerinin anlaşılmasının yanı sıra bu metotların gündelik hayatta derin olarak uygulanmasını beraberinde getirmektedir.

Bu araştırmalar arasında en yoğun uygulama alanı olan ve başta fizik ve matematik alanları olmak üzere bilişim ve mühendislik alanlarını derinden etkileyen araştırma alanı parçacık fiziğidir. Parçacık fiziği yaklaşık yüzyıldan fazla kuramsal modelleme ve deneyler sonucu oluşturulmuş, şu anda yaşadığımız evreni anlamaya ve çözümlemeye çalışan bir alandır. Şu ana kadar elde edilen tüm verilerin ve hesaplamaların temel oluşturduğu, günümüzde bilinen tüm temel parçacıkları ve aralarında ki etkileşimleri tasvir eden kuramsal modelleme Standart Model olarak adlandırılır. Parçacık fiziğinin Standart Modeli (SM) olarak adlandırılan bu yapı (Figür 1-sol), birçok kuramsal hesaplamalar ve deneysel gözlemler sonucu oluşturulmuş kuvvetli bir kuantum alan teorisi kuramıdır (Abers ve Lee 1973). Günümüz parçacık fiziğinin birçok araştırma konusu bu model çerçevesinde incelenmekte olup bu modelin getirdiği sonuçlar, bunlar arasından yapılan tahminler ve ölçümler, doğada gözlemlenmekte olan fakat bu model çerçevesinde anlaşılamamış yapılar üzerinden ilerlemektedir. SM'in açıklayamadığı durumlar, SM ötesi kuramsal modellemeler çerçevesinde SM'i içerecek ve genişletecek şekilde detaylıca incelenmekte, elde edilen veriler ileri matematiksel hesaplamalar, istatiksel metot ve algoritmalar aracılığı ile çözümlenmektedir. Bulunduğumuz evreni anlamak adına SM'in cevaplamakta yetersiz kaldığı durumlar aşağıdaki gibi birkaç temel başlıkta ele alınmaktadır:

- Evrendeki toplam maddenin yaklaşık %26,8'ini oluşturan Kara Madde (KM) için herhangi bir açıklama getirmemektedir ve KM için herhangi bir aday parçacık model içerisinde bulunmamaktadır
- SM nötrino olarak adlandırılan parçacıklar kütlesiz olarak tanımlanmaktadır fakat en son gözlemler bu parçacıkların bir kütlelerinin olduğunu göstermektedir
- SM bünyesinden bulunan güçlü etkileşim, zayıf etkileşim ve elektromanyetik etkileşimleri tasvir eden ayar kuplaj sabitlerinin yüksek enerji değerlerinde bir araya getirerek tek bir kuvvet terimleri olarak yazmak mümkün olmamaktadır (Birleştirilmiş alan kuramı)
- KM adayı için herhangi bir önerisi bulunmayan SM'in, bir diğer önemli problemi ise kuantum mekaniksel etkiler karşısında kararlı davranamayışıdır (ayar hiyerarşisi problemi)
- Çekim kuvveti SM'de yer almamaktadır
- Parçacıklar arası kütle farkı büyüklüğü (Kütleler arası hiyerarşi) ve kökenleri konusunda bize herhangi bir bilgi aktarmamaktadır

SM ötesi modellerin aranması için BHÇ'de uygulanan sonlu durum veri çözümleme metotları uygulanmaktadır. Bu metotlarda detektörlerde gözlemlenen ve çeşitli objeleri temel alan alt-detektörlerin verdiği analog sinyaller, detektör elektroniği ve özel yazılımlar aracılığı ile dijital sinyale dönüştürülerek parçacıkların sahip oldukları momentum (px, py, pz) ve enerji değerleri 'olay' olarak kaydedilerek elimizdeki fiziği anlamak üzere bize gerekli olan 4-vektör bilgisi sağlanmaktadır. BHÇ`de yer alan deneyler, ATLAS, CMS, LHCb ve ALICE, parçacık fiziği alanı içeresinde kendilerine verilmiş belirgin görevleri yerine getirmek üzere tasarlanmışlardır. Bu deneylerin temel hedefleri, özellik çok amaçlı ATLAS ve CMS deneyleri olmak üzere, yeni keşfedilmiş olan Higgs bozonun anlaşılması esnasında yapılan SM hassaslık ölçümler ile birlikte yeni fizik olarak adlandırılan SM ötesi modellerin aranmasıdır. Bu deneylerden bizim çalışmamızda kullanılacak olan detektör Kompak Muon Selonoid detektörüdür (Bayatian 2006). CMS detektörünün temel özellikleri Şekil 1-sol`da gösterilmiştir. Genel olarak 21.5m uzunluğunda, 15m yüksekliğinde ve yaklaşık olarak 12500 Ton ağırlığındadır. CMS algıcının en temel özelliklerinden biri 6m`lik alan içerisine alınmış kalorimetre sisteminin geniş bir süper-iletken selonoid içerisinde yer almasıdır. Bu tasarım aynı zamanda algıca kompakt bir tasarım sağlamaktadır. Selonoid proton ışın eksenine paralel yaklaşık 4T`lık homojen bir manyetik alan uygulamaktadır, böylece iç takipçide (inner tracker) yer alan yüklü parçacıkların yörüngeleri değiştirilerek ölçümlerinin yapılması sağlanmaktadır. Parçacıkların momentum bilgileri takipleyici (tracker), enerji bilgileri kalorimetriler tarafından sağlanmaktadır.

SM ve ötesi kurumlar incelenmek üzere detektör elektroniğinden gelen büyük verilerin işlenmesi gerekmektedir. Bu çerçevede bilim insanları aşağıda yer alan sonlu durumları temel alarak çeşitli SM ve ötesi analizler gerçekleştirmektedir. Bu incelemelerde oluşan parçacıkların kaskad bozunumları sonucu oluşturdukları yapının detektör içerisinde tanımlanmaları ve elde edilen tüm veri arasında bir bağlantı kurulması sağlanarak veri çözümlemesi yapılmaktadır. Çarpışmalarda ortaya çıkan kuarklar durağan olmadıkları için detektör içerisinde bir çoklu dağılım olarak gözükmekte ve parçacık veri çözümleme dilinde jet olarak adlandırılmaktadır. Jetler, içlerinden etkileşim kuvveti gluon ve hafif ve ağır kuarkları bulundurabilirler (Berends, H. Kuijf, 1991). CMS detektöründen gelen bir nitelik ve niceliğini taşıyan olayın içerisindeki parçacıklar, 4-vektör değerleri ve bunlar arasındaki koordinat bilgileri net olarak ölçümlemesiyle birlikte analizler yapılabilmektedir. Bozunumların topolojik bilgileri dahilinde incelenebilecek bu durumlar birkaç alt başlıkta tanımlanabilir:

Bu çerçevede proje kapsamında geliştirilmesi planlanan makine öğrenimi ve yapay sinir ağları metotlarının yüksek enerji ve parçacık fiziğindeki ileri uygulamaları aşağıdaki şekilde ana ve alt başlıklar altında özetlenmektedir:

- Kuram ve deney verilerinin üretildiği Monte-Carlo (MC) simülasyonlarının anlaşılması
- Veri/MC uyuşmazlığı ve sistematik belirsizlikler
- Enerji regresyonu ve olay sınıflandırması
- Piksel kümelemesi, Takip kalite sınıflandırması, veri sıralama

Makine öğrenimi ve uygulamalarının yüksek enerji ve parçacık fiziğindeki uygulamaları:

- Yapı/şekil tanınması: detektör etkileşimlerinin kümelenmesi
- Takip sınıflandırılması: kalite seçimi ve yinelenen durumların silinmesi
- Enerji/Momentum regresyonu: photon, elektron, jetler, b-jetler
- Obje tanımlanması: elektron seçimi, b/c kuarkların belirlenmesi, ardalan bilgisi
- Tüm olayın (event) sınıflandırılması: sinyal olaylarının ardalan olaylarından ayrılması: Likelihood metodu, sinir ağları, Ötelenmiş karar ağacı (Boosted Decision Tree) 1 ve/veya 2 boyutta çoklu değişken çıktılarının uyum kontrolü (fit)
- İleri kullanım durumları- karıştırılmış kullanım: Daha basit seçimler için çıktıların sınıflandırılmasına bakılması

Tüm bu uygulamaların temelinde yatan ana neden faz uzayını kesmeye dayalı (cut-based) sınıflandırmadan öğrenilmiş durumların denetlenmesi ve makine öğretimi (supervised learning techniques and machine learning) durumuna doğru olan geçişlerdir ve bu projede bunlar uygulanmıştır. İlgili ardalan ve sinyal örnekleri çalışılmış ve bu örnekler üzerinden yapay zekâ özelinde parçacık fizikçileri tarafından kullanılmakta olan TMVA (A Toolkit for Multivariate Analysis) ve TensorFlow gibi araçlar analize dahil edilmiştir. Snowmass örnekleri olarak adlandırılan ardalan örnekleri üzerinden çeşitli pileup senaryoları dahilinden çalışmalarımız yürütülmüş ve bu kapsamda önemli çıktılar elde edilmiştir.

# 2. LİTERATÜR ÖZETİ

CMS Deneyi, 2017 yılı sonu itibari ile yaklaşık toplam ~100/fb entegre parlaklığa ulaşmıştır (LHC Lumi). Bu parlaklık seviyesi, Standart Model (SM) ötesi yeni fizik kuramlarının araştırılabilmesi için uzun yıllardır ihtiyaç duyulan kuram/deney karşılaştırması konusunda çok hassas bilgiler sağlamaktadır. Bu kuramların en başta geleni ve en yaygın olarak araştırılanı süpersimetrik modellerdir (Kane, Gordon). Süpersimetri, bilinen parçacıkların her biri için bir süpereş atayarak Higgs parçacığının kütlesini kuantum çalkanımlarına karşı kararlı hale getirir. Doğada yer alan en ağır kuark olan top kuark`ın süpereşi olan skaler top kuark`ın en hafif kuark veya lepton süpereşi olması, fotonun süpereşinin de Evren'in %24'ini oluşturan Karanlık Madde'yi oluşturması beklenmektedir (Planck 2015). Karanlık Madde'yi oluşturan bu parçacık adayı çarpışma süreçlerinde kendini "kayıp enerji" olarak göstermektedir ve günümüz araştırmalarında en önemli soru işaretlerinden biri olarak ön plana çıkmaktadır (Focus). Deneyin başladığı günden bugüne kadar keşfedilemeyen yeni parçacıklar, toplanan veri miktarının artması ile aramalarına devam etmekte aynı zamanda 2000'li yılların başlarında LEP deneyinde olduğu gibi SM hassaslık ölçümleri ile birlikte kuramsal modelin birçok etkileşim ve süreçler arası bağımlılıklar incelenmekte ve önemli çalışmalar yürütülmektedir [CMS Public, ATLAS Public, LHCb Public, ALICE Public]. Özellikle LHC içi ve dışı birçok deneyin aynı konu özelinde yürüttükleri çalışmalar büyük resmin anlaşılması konusunda çok önemli ipuçları sağlamaktadır (AMS, DUNE, Dark Energy Survey, Cosmos Experiment, CMB, LUX experiment).

Kuantum alan kuramı çerçevesinde genişletilen yeni fizik modellerinden biri olan Süpersimetri ve/veya efektif alan kuramlarında özellikle Top kuark kütlesine yakın noktalarda bulunan yeni parçacıklar, Süpersimetrik Top kuark, ardalan'dan gelen katkıların sistemi domine etmesinden dolayı büyük bir gürültü yaratmakta ve şu ana kadar hiçbir analizde tam anlamı ile anlaşılamamaktadır (CMS Top). Biz analizimizde bu ardalan ile farklı bir şekilde başetmek için süpersimetrik top çiftlerinin faz uzayından yer aldıkları zor bölgeleri algoritmik çözümleme teknikleri ile kurgulanmış bir ayıraç kurguladık (Bernreuther, W.). Bu noktada üç temel probleme çözüm önerdik: Top Kuark'a yakın, örneğin süpersimetrik skalar top kuark, yeni parçacıkların tesir kesiti birkaç femtobarn seviyesindedir ve analiz edilebilmesi için oldukça yüksek luminosite değerlerine çıkılması gerekmektedir (Guadagnoli, G.). Şu anda elimizde bu noktaların anlaşılması için yeterli veri miktarda veri toplanamamıştır veya kombine edilememiştir. İkincisi, tesir kesitinin düşük olmasından dolayı çok zor gözlemlenen bu durumların ölçümlenmesi için, detektörlerde oluşan radyasyona bağlı bozunumların (deneyin yaklaşık 10 seneden beri çalıştığını ele alırsak) yarattığı elektronik kayıplar, detektörün bakımı ve yüksek çözünürlüklü hale getirilmesi oldukça önemlidir ve önemli bir güncelleme çalışmalarının yürütülmesi gerekmektedir (CMS SUSY, CMS TDR). En son olarak, kuramsal ve deneysel olarak oldukça iyi kurgulanmış ve toplanmış verilerin güncel yapay zekâ teknikleri ile evrende yarattığı tüm yapısal durumların algoritmik çözümlenmesi ve büyük veri analizlerinin sürdürülmesi gerekmektedir (AI LHC). Bu çerçevede veri dosyalarımızı geleneksel çok değişkenli analiz metotlarını içeren TMVA paketi ile inceleyerek ve devamında en güncel olan TensorFlow ve H20.ai paketleri ile çalıştık (AI Tools). Böylece makine öğrenme metotları hem fizik analizinde hemde detektör verisinin akıllı işlenmesinde kullanma çalışmalarına başladık. Bu üç temel başlık altında projemiz oldukça heyecan verici ve her konu özelinde önemli çalışmalar içerecek, ulusal ve uluslararası önem arz eden bir yapıya bürünmüştür.

LHC Lumi, “LHC Luminosity public results” “https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults”

Kane, Gordon L., *Supersymmetry: Unveiling the Ultimate Laws of Nature*, Basic Books, New York (2001). ISBN 0-7382-0489-7.

Planck 2015, XIII. Cosmological parameters, 10.1051/0004-6361/201525830

Focus, “Focus areas”, https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy

CMS public, https://cms-results.web.cern.ch/cms-results/public-results/publications/

ATLAS public, https://twiki.cern.ch/twiki/bin/view/AtlasPublic

LHCb Public, https://lhcb.web.cern.ch/lhcb/Physics-Results/Recent-LHCb-Results.html

ALICE Public, https://twiki.cern.ch/twiki/bin/view/ALICEpublic/ALICEPublicResults

AMS Public, http://www.ams02.org

DUNE, “DUNE Experiment”, http://www.dunescience.org

Dark Energy Survey, “Dark Energy Survey”, https://www.darkenergysurvey.org

Cosmos experiment, “Cosmos Planck”, https://www.cosmos.esa.int/web/planck

CMB Experiment, https://cmb-s4.org

LUX Experiment, http://luxdarkmatter.org

CMS Top, “Public CMS Top Physics Experiment Measurements” https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsNewsTOP

W. Bernreuther, A. Brandenburg and P. Uwer, “Transverse polarization of top quark pairs at the Tevatron and the large hadron collider” Phys. Lett. B 368 (1996) 153 doi:10.1016/0370-2693(95)01475-6

D. Guadagnoli and C. B. Park, “$M_{T2}$-reconstructed invisible momenta as spin analizers, and an application to top polarization," JHEP 1401 (2014) 030, doi:10.1007/JHEP01(2014)030

CMS SUSY, CMS experiment SUSY Public results twiki, https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS

CMS TDR, “Public CMS Technical Design Report” https://cds.cern.ch/search/?cc=Design%20Reports&p=CMS-TDR*&f=reportnumber&sf=year&so=d

AI LHC, “Artifical Intelligence Called in to tackle LHC Data Deluge”, https://www.nature.com/news/artificial-intelligence-called-in-to-tackle-lhc-data-deluge-1.18922

AI Tools, “Projede kullanılan yapay zeka araçları” http://tmva.sourceforge.net, https://www.tensorflow.org, https://www.h2o.ai

# 3. GEREÇ ve YÖNTEM

Bu projede kullanılması planlanan çeşitli yazılım araçları ve teknikleri bulunmaktadır. Bilimsel olarak incelemek istediğimiz konu SM ve SM ötesi parçacıkların BHÇ'de etkileşimlerini incelemek ve sinyal/ardalan ayrımı olduğu için ilk başlanılması planlanan yöntem bu etkileşimlerin kuramsal ve deneysel verilerinin üretilmesi tamamlanmıştır.

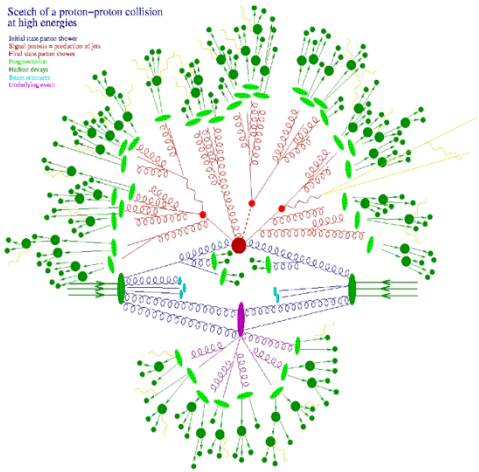
Kuramsal veriler genel anlamda SM'den bildiğimiz ve büyük hassaslık ile ölçebildiğimiz prosesleri ileri seviye MC hesaplayıcıları arasında, aralarındaki farklılıklarda göz önünde bulundurularak üretilmiştir. Bunlar sırası ile TTjets, W/Z Jets, VV+Jets, Single Top, ve farklı SUSY sinyal örnekleri olarak özetlenebilir. Bu sinyal olayları için bize SM ötesi modeller tarafından sağlanan fenomonolojik modeller irdelenerek çeşitli sinyal örnekleri üretilmiştir. Her bir ardalan ve sinyal prosesleri sahip oldukları tesir kesit değerlerine göre uygun parlaklık (luminosite) değerine göre toplamda milyonlar mertebesine kadar örnek üretimi yapılmıştır. Kuramsal anlamda tesir kesiti için hesaplanarak üretilen tüm olaylar daha sonra detektör simülasyonlarına aktarılarak tekrar canlandırma ve çarpışma esnasında olan olayları canlandırma aşamasına geçilmiştir. Bu durum kendi başına oldukça karmaşık olmasına rağmen elde edilen veriler arasında bağıntı denklemleri ve ayrıştırma metotları paralelinde yapay sinir ağları ve makine öğrenim metotları bu örneklere uygulanarak çeşitli sonlu durumlar irdelenmiştir.

Uygulanan temel metodoloji, araç gereç teknikleri ve yüksek enerji fiziği modeli aşağıdaki gösterilmiştir:

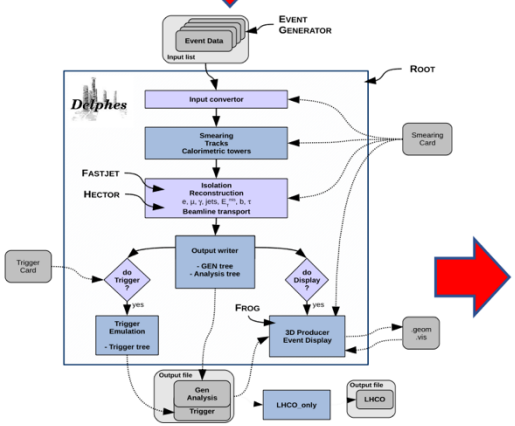
1. Parçacık fiziği kuram ve hesaplamalar, parçacık fiziği olaylarının simülasyonu (MC) – Kullanılan MC araçları: Sherpa, Madgraph (Pythia8 Parton Sağanağı simülasyonları hadron seviyesine getirilmiştir), Powheg ((Pythia8 ve Herwig Parton Sağanağı simülasyonları ile ayrı ayrı simüle edilmiştir), Pythia 8 ve Herwig simülasyonları ayrıca kullanılmıştır.

2. Detektör simülasyonu ve MC'dan üretilen kuramsal modellerin tekrar oluşturulması – Tüm ardalan ve sinyal simülasyonları Delphes hızlı detektör simülasyonundan geçirilerek Kuram-Deney karşılaştırmalarına uygun gale getirilmiştir.
3. İleri veri çözümleme tekniklerinin uygulanması için gerekli veri çözümleme programları – ROOT, TMVA ve TENSORFLOW gibi araçlar kullanılmıştır.
4. Sinyal/Ardalan veri çözümleme uygulamaları, makine öğrenimi ve yapar sinir ağları veri sonuçlarının farklı teknikler ile karşılaştırılması – Karar ağaçları (Decision Trees), Vektör Destek Makinesi (Support Vector Machine), Nöral Ağlar (NN) algoritmaları ağırlıklı olarak analizlerimizde kullanılmaktadır.
5. Yapay sinir ağları ve makine öğrenimi detaylıca incelenmesi ve araştırılması için ilgili alt-yapının elde edilmesi – CPU ve GPU özelinde yüksek hesaplama içeren yapılar özelinde NVidia ve Dell RT620 CPU işlemciler üzerinden kıyaslamalar yapılmıştır.

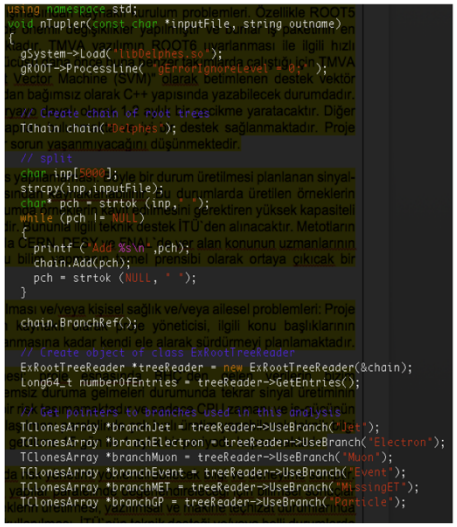
**Tablo 1.** Çizim olarak gösterilen methodolojik yaklaşım şemasında siyah kutu ile gösterilen kısım proje grubu tarafında geliştirilen nTuple kodunun küçük bir kısmı olmaktadır. Kendi nTuple'larımız ile birlikte oluşturduğumuz detaylı analiz kodumuz ile birlikte aldığımız grafik çıktıları en sağ tarafta gösterilmektedir. Bu grafiklerin elde edilmesi için oldukça kapsamlı ve detaylı yazılım teknikleri kullanılmıştır




1. Adım (1. ve 2. ve 6. iş paketleri)

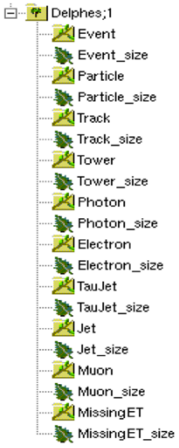


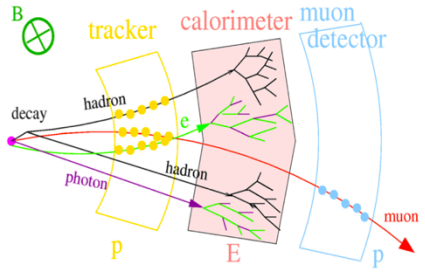

2. Adım (3. ve 6. iş paketleri)



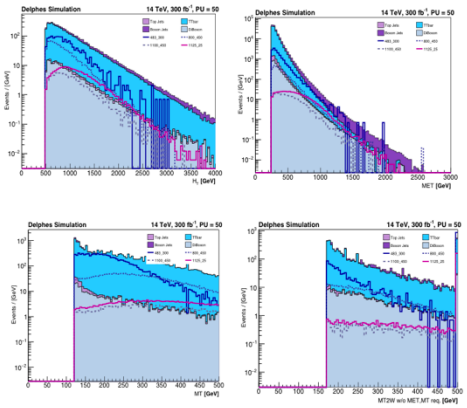
```
using namespace std;
void nTupler(const char *inputFile, string outname)
{
  gSystem->Load("libDelphes.so");
  gROOT->ProcessLine( "gErrorIgnoreLevel = 0;");
  // Create chain of root trees
  TChain chain("Delphes");
  // split
  char inp[5000];
  strcpy(inp,inputFile);
  char* pch = strtok (inp,",");
  while (pch != NULL)
  {
    printf ("Add %s\n",pch);
    chain.Add(pch);
    pch = strtok (NULL, ",");
  }
  chain.BranchRef();
  // Create object of class ExRootTreeReader
  ExRootTreeReader *treeReader = new ExRootTreeReader(&chain);
  Long64_t numberOfEntries = treeReader->GetEntries();
  // Get pointers to branches used in this analysis
  TClonesArray *branchJet = treeReader->UseBranch("Jet");
  TClonesArray *branchElectron = treeReader->UseBranch("Electron");
  TClonesArray *branchMuon = treeReader->UseBranch("Muon");
  TClonesArray *branchEvent = treeReader->UseBranch("Event");
  TClonesArray *branchMET = treeReader->UseBranch("MissingET");
  TClonesArray *branchGP = treeReader->UseBranch("Particle");
```

4. Adım (5. iş paketi)

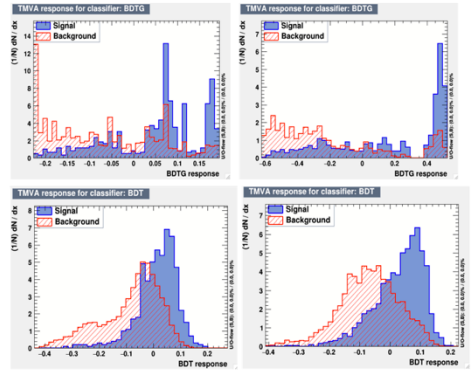



3. Adım ( 4. iş paketi)




5. Adım (7. iş paketi)




6. Adım (8. iş paketi)

### 3.1 Monte-Carlo (MC) simülasyonlarının (sinyal ve ardalan) ortak çalışma sunucusuna kurulumu ve kullanıma hazır hale getirilmesi ve kıyaslanması:

Günümüzde çok çeşitli MC üreteçleri birçok amaç için kullanılmaktadır. Bu nedenle MC üreteçleri yüksek enerji fiziğinde kuantum alan kuramı uygulaması olarak karışık ve hesaplanması güç etkileşim diyagramlarını bir paket dahilinde çeşitli seviyelerde bilimsel olarak kullanılmak üzere incelenmesi oldukça önemli bir başlık olarak ortaya çıkmaktadır. Kuantum alan kuramı parton modellemesi uyarınca BHÇ'de gözlemlenen durumlar MC üreteçlerinde çeşitli seviyelerde incelenmiş ve incelenen MC farklılıkları sırasıyla, başlangıç parton yağmuru (initial parton shower), ilgili sinyalin üretilmesi - parton etkileşmesi, son durum parton yağmuru (final state parton shower), parçalara ayrılma (fragmentation), hadron bozunumları (hadron decays), proton ışını kalıntıları (beam remnants) ve çoklu alt-etkileşimler (underlying events) bir veri seti olarak üretilmiştir. Her bir adım için ve aynı zamanda tümünü içeren çeşitli MC simülasyonları bugün geliştirilmekte ve yaygın olarak kullanılmaktadır. Bunlar arasında tüm bu adımları basitleştirilmiş ya da en kapsamlı şekilde inceleyen çok amaçlı Pythia6/Pythia8 (Sjostrand vd. 2008), Herwig/Herwig++ (Belm vd. 2015), Sherpa (Gleisberg vd. 2004) gibi MC simülasyonlarının yanı sıra, bu adımlardan birini ileri seviyede inceleyerek diğer adımlar için yukardaki çok amaçlı MC simülasyonlarını kullanan MC programları da çok kapsamlı olarak proje ekibi tarafında çalışılmıştır.

Proje ekibi detaylıca birçok MC üreteci ve fizik sonuç değerlendirmesi yapmıştır. Özellikle yeni fizik arayışları için anlaşılması elzem olan ardalan katkıları üzerinden kuramsal belirsizliğin deneysel simülasyonlara olan propagasyonu çalışılmıtır. Bu nedenle parton yağmuru, parçalara ayrılma, hadron bozunumları ve çok alt-etkileşimler konusunda oldukça kararlı ve literatür tarafından kabul görmüş Pythia 8 ile birlikte parton modellemesi konusunda ağaç seviyesi ve üzerinden hesaplama desteği veren ileri seviye genlik hesaplayacıları Softsusy (Allanach 2002), Susyhit (Djouadi), Spheno (Porod 2003), Isajet (Baer 1993), MadGraph/MadEvent (Maltoni vd. 2003), PowHeg (Frixione ve Nason, 2007) ve aMC@NLO (Frixione ve Webber, 2003) gibi matrix element hesaplayıcılarının Pythia 8 ile olan arayüzleri kullanılarak hesaplamaların ileri seviye değerleri elde edilmiştir. En son olarak bu MC üreteçleri arasında tüm fizik bilgilerini ileri seviye içeren ve helisite genliğini ve parton yağmurunu ağaç seviyesi ve daha üst seviyelerde (Next-to-leading-order (NLO) ve Next-to-leading-log (NLL)) detaylıca hesaplayan Sherpa MC üreteci bu projede kullanılmaktadır.

### 3.2 Detektör simülasyonu ve MC'dan üretilen kuramsal modellerin tekrar oluşturulması:

MC üreteçlerinde üretilen hadron seviyesindeki veriler, kuram-deney karşılaştırması için detektör simülasyonlarından tekrar canlandırılmıştır (reconstruction). Bu canlandırmalar bozunuma uğrayan parçacıkların detektör içerisinde yer alan alt-detektör ve detektör malzemesi ile nasıl etkileştiğini göstermesi, parçacıkların çarpışma anında itibaren takip etmesi olası yörünge ve enerji dağılımları konusunda bize bilgi vermesi sağlanmıştır. Detektör simülasyonları BHÇ'de en ileri seviyede CMS detektör yazılımı (CMSSW) ayrıca bir grup ile koordine edilmiş (makale çalışmasına katkıda bulunulmuştur). Hızlı detektör simülasyonları gerçek veri ile karşılaştırılarak ayarlanmış ve yaklaşık olarak temel proseslerin tekrar yaratılmasını pratik ve hızlı bir şekilde gerçekleştiren, basit bir şekilde CMSSW ve ATHENA benzeri programlara göre daha basitleştirilmiş geometri kullanan yazılımlardır ve projemizde biz bu şekilde hızlı detektör simülasyonlarını kullandık. Bunlar arasından en son yıllarda oldukça

önemli seviyede gelişme göstermiş ve CERN'de yer alan araştırmacılar tarafından geliştirilen DELPHES simülasyonu (Selvaggi 2014) bizim çalışmalarımızda aktif olarak kullanılmıştır. Delphes detektör simülasyonu bize ilgi duyduğumuz verinin çabuk üretilmesini (~10 kat) ve aynı zamanda derin öğrenim konusunda yapmak istediğimiz çalışmaları saygın dergilerde bağımsız olarak yayınlamamızı sağlamaktadır. Veri formatı CMSSW/ATHENA benzeri fakat basitleştirilmiş yapısı olduğu için CERN'de yer alan araştırmalarımız için başlangıç seviyesinde çalışma imkânı ve uygun bağımsız bir platform oluşumu sağlamıştır. Delphes çalışması methodolojisi ile ilgili adımlar yukarıda yer alan şemada gösterilmiştir. Kendi veri setlerimizi üretirken daha önce Snowmass platformu üzerinden üretilmiş ve valide edilmiş ver kartları ve konfigürasyonları PileUp'sız 50 PileUp ve 140 PileUp senaryoları özelinde kullandık. Burdan çıkan sonuçları birbiri arasında irdeleyerek kuramsal ve deneysel belirsizliklerinin derin öğrenme teknikleri aracılığı çözümlenmesini sağladık.

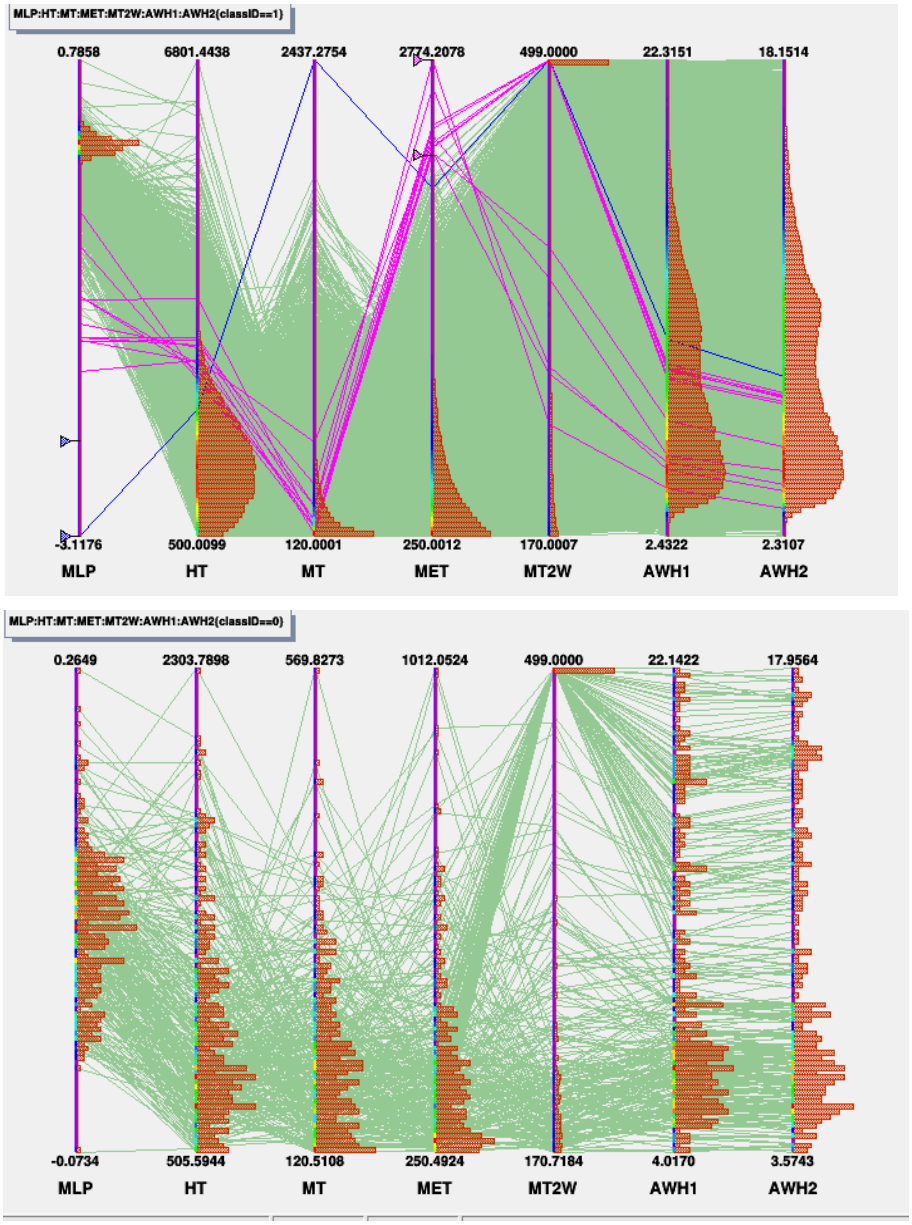
### 3.3 İleri veri çözümleme tekniklerinin uygulanması için gerekli veri çözümleme programları

Yukarıda bahsi geçen veri setlerinin üretiminden sonra ilgili bilgi kapsülleri-ntupler, bu aşamada detayına incelenerek verilerin derlenmesı/sınıflandırılması, incelenecek objelerin belirlenmesi, temel seçimlerin uygulanması, çeşitli değişken hesaplarının yapılması ve bunları bir alt-formata indirgenmelerini içermektedir. Sonuç olarak bu aşamada incelenen veriler bir bütün olarak filtrelendikten sonra yaklaşık birkaç 10 terabayt'lık büyüklüğünden gigabayt büyüklüğüne filtrelenerek ileri seviye hesaplamalara uygun formata dönüştürülecektir. Bu dönüşümün yapılması için CERN tarafından geliştirilen veri çözümleme programı ROOT kullanılmıştır (Bellenot vd. 2015). Bu veri çözümleme programı 1998 yılından beri dünyada yer alan birçok deneyde kullanılmakta olup oldukça detaylı ve ileri seviye çözümleme yetisine sahiptir. C++ tabanlı program, eldeki veri incelemek üzere çeşitli okuyucu kütüphaneleri sağlamakta ve araştırmacılar inceledikleri konu çerçevesinde kendi veri çözümleme makrolarını yazmaktadırlar. Bu makrolar yada veri çözümleme programları ntuple-reader olarak adlandırılmaktadır. Biz projemizde oldukça önem arz eden ve bizim elimizdeki tüm veriyi aynı anda çözümleyip yapay sinir ağları, karar ağaçları gibi ileri seviye istatiksel metotları uygulayacağımız bir sonraki adım için uygun formatlara indirgememizi sağlayan bir yapı oluşturduk ve tüm çalışmalarımızı bu yazılım ile koordine ettik. Projenin ana gövdesi durumundaki bu platformun bu proje ile geliştirilmiş ve devam eden yıllarda ve başka projelerde bu yapının kullanılması planlanmaktadır. ROOT veri çözümleme programı DELPHES kurulumunda da kullanılmakta olup, Delphes ROOT formatından başa analiz araçlarında kullanılmak üzere uygun ntuplelar üretmektedir.

### 3.4 Sinyal/Ardalan veri çözümleme uygulamaları makine öğrenimi ve yapar sinir ağları kullanımı

Tüm elde edilen veri setleri TMVA (The Toolkit for Multivariate Data Analysis with ROOT) paket programı (Hocker 2007) aracılığı ile bu proje için oluşturulan uygulama iskeletinde (yukarıda gösterilen şema uyarınca beş ana başlık) kullanılmıştır. Öncelikle SM ötesi Süpersimetri ve Ekzotik fizik olarak adlandırılan jenerik sinyallerin SM ardalan üzerinden ayrılması olasılıklarını hesaplamak ve belli eşik sinyalleri için çeşitli derin öğrenim ve makine öğrenim metotları kullanılmış ve projede parçacıkların sahip olduğu karakteristik özelliklerin üzerine araştırmalar yapılmıştır. Projemizde bu aşamada hedeflenen temel motivasyon: jenerik olarak seçilen birkaç sonlu duruma (hadronik, tekli lepton, ikili lepton çoklu lepton + kayıp enerji) sahip sinyal yapılarının çok küçük tesir kesiti ve bozunum yapılarına göre SM ardalan üzerinden keşfedilmesi için oluşturulması muhtemel derin öğrenim metotları ve regresyon çözümleme yapısının oluşturulmasıdır. Hali

hazırda proje kapsamında burs alan bir öğrencilerimle birlikte bu çalışmalara devam etmekte aynı zaman projenin en önemli çalışmasını tamamlamak üzere çalışmaktayız. Bu yapı yardımı ile CMS ve ATLAS deneyleri gibi deneysel kolaborasyonların kullandığı ileri seviye metotların kullanılarak geliştirilmesi, ayrıca LHC deneylerinden bağımsız bir şekilde kullanılması sağlanacaktır (MC simülasyonları ve Delphes detektör simülasyonları). Bu projeye başarısını pekiştiren, özgünlük ve görünürlük katan bir yapı oluşturmaktadır. **Projenin temel hedefi ilgili uygulama iskeletinin kurulduğu analiz platformu öğrenim, eğitim ve fizik uygulamaları için ilgili platformun oluşturulması sonrasında çok çeşitlilik gösteren bilişim ve sanayi uygulanmaları ile birlikte bilim-bilişim-sanayi etkileşimleri için yetişmiş eleman ve projelendirme çalışmalarına destek olmuştur, olmaktadır.**




**Şekil 1.** TMVA aracında farklı değişkenlerin birbirleri arasında olan korelasyonlarının her bir bin için çıkartılması. 1 boyutta grafikleri yatay ve dikey paralel koordinat dağılımları. MLP TMVA tarafında üretilen ayıraç grafiği olmak üzere diğer değişkenler analizimizde kullanılan bir boyutta dağılımlardır.

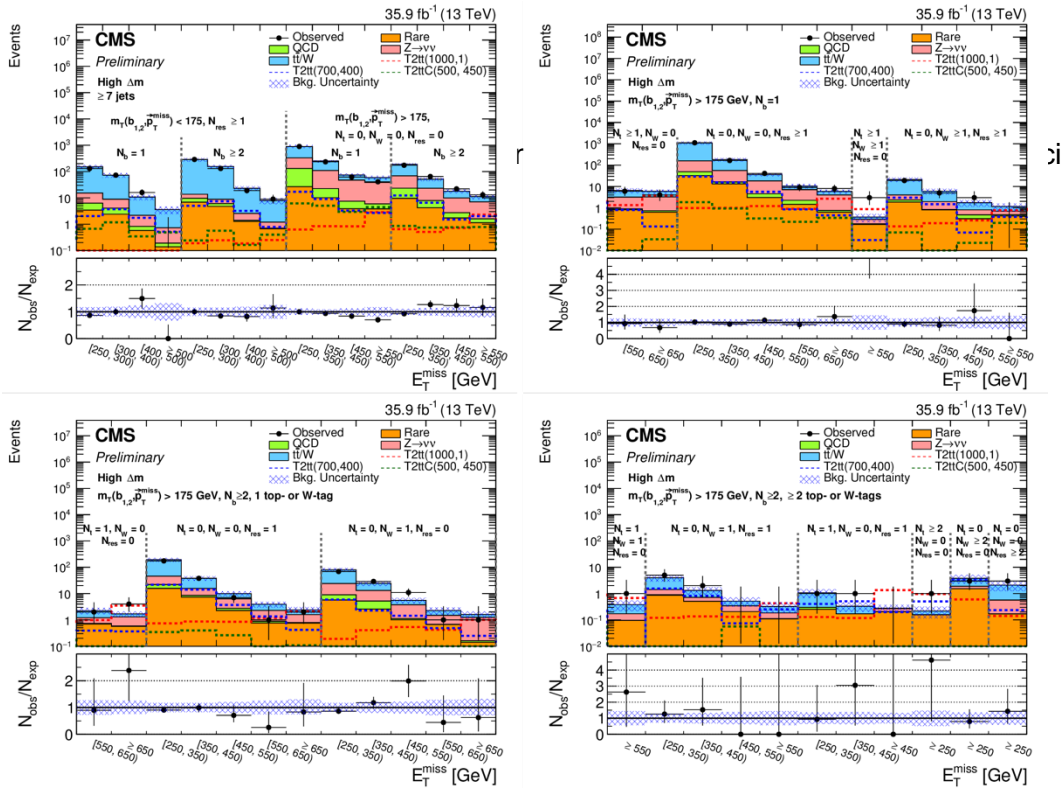
MLP metodu ile SM ardalan üzerinden tanımlanan değişkenleri kesmelerden önceki ve sonraki durumlar içerisinden birbirlerine olan bağımlılık analizileri. Paralel koordinatlar olarak adlandırılan bu yapı TMVA-2.4.0 ile yapılmıştır.

Projenin bu ve bir önceki aşamaları içeren araştırmalarımızda, parçacık fiziğinde yeni fizik arayışlarında kullanılan birçok teknik ve inceleme, yazılım teknolojileri olarak lisans öğrencileri ile koordine edilmiş ve proje iş paketlerine ek olarak çok fazla çalışma üretilmiştir. Özellikle bilimsel olarak proje kapsamında öğrenilen tekniklerin akademi-sanayi kapsamında konu özelinde kullanılması içeren yapılar içermesi, projenin küçük ama büyük katkısı olarak ortaya çıkmıştır.

# 4. BULGULAR

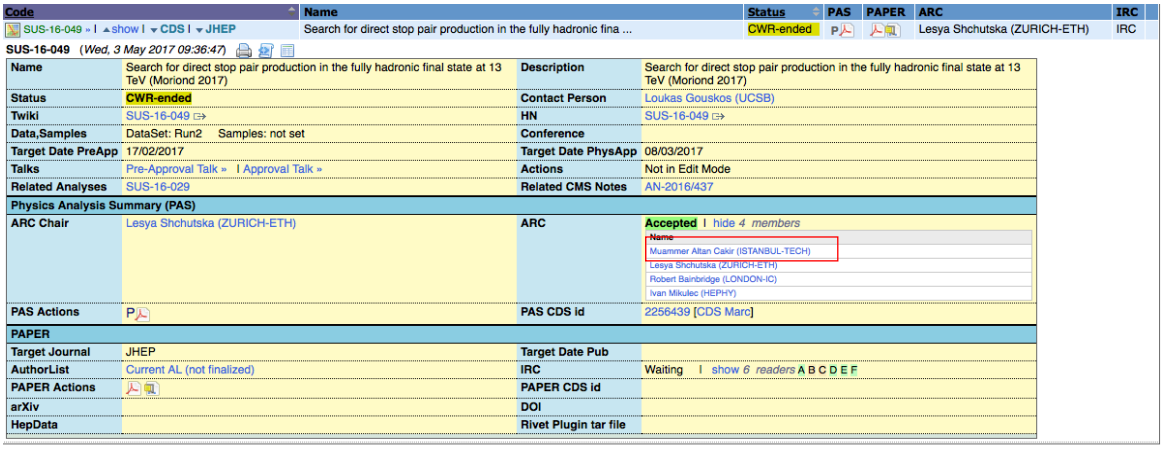
Proje çalışmamızda gereç ve yöntem çizelgesinde takip ettiğimiz metodoloji çerçevesinde gerek CMS deneyi içerisinde oldukça önem arz eden iki tane çalışmaya, gerek kendi imkanlarımız ile yarattığımız temel seviyede bir parçacık detektöründe, aynı zamanda proje kapsamında yaptığımız simülasyonlar sonucu üretilen veri setleri ile birlikte çok geniş spektrumda sonuçlar elde edilmiştir. Bu sonuçlar gelişme raporlarında açıklanmış olmasına rağmen sonuç raporunda temel anlamda ii sayfasında verilen özete ve gelişme raporlarına açıklanan **gelişmelere paralel sırasıyla tekrar** açıklanmıştır. Projenin temelini oluşturan ve hali hazırda makale çalışmaları süren detaylar 4.1, 4.4 ve 4.5 numaralı maddeler de açıklanmaktadır. Diğer maddeler bu proje çerçevesinde oluşturulan yapıdan yararlanılarak gerek bilimsel gerek akademi-sanayi işbirliği konularında başarı ile uygulanan çalışmalardır.

**4.1. CMS Kolaborasyonun da Direk Stop Araştırmaları**: Proje kapsamında yapılan çalışmalar paralelinde CMS kolaborasyonu içerisinde proje yürütücüsünün derin öğrenim araştırmalarının LHC uygulamaları konusunda önemli bir yayına katkılar yapılmıştır. Proje yürütücüsü CERN'de yer almakta olan LHC-CMS deneyinde çalışmakta ve bu proje kapsamında çalıştığı konuları gerçek veri analizlerine uygulamaktadır. Bu kapsamda SUSY araştırmalarında önemli bir sonuç arz eden ve İlknur Köseoğlu'nun tezi çerçevesinde incelenen sinyal yapısına benzer "Search for direct stop pair production in the fully hadronic final state at 13 TeV" makalesi yayınlamış ve ilgili yayın Moriond 2017 konferansında sunulmuştur. [1] nolu çalışmamda 13-22 figürler arasında BDT (Ötelenmiş karar ağaçları – boosted decision trees) efektiv olarak kullanılmış ve MVA teknikleri çalışılmıştır. İlgili çalışma JHEP'de yayınlamıştır. Projenin ilk zamanlarına gelmesinden dolayı bu çalışmaya hemen katkılarda bulunulmuş ve önemli bir başarı elde edilmiştir.




**Şekil 2.** Direk stop-stop üretimi kanalında SUSY sinyal arayışları kayıp momentum enerjisi ardalan farklı sinyal noktaları [*].

[*].http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SUS-16049/index.html



| Code | Name | Status | PAS | PAPER | ARC | IRC |
|---|---|---|---|---|---|---|
| SUS-16-049 » I show I CDS I JHEP | Search for direct stop pair production in the fully hadronic fina ... | CWR-ended | P | | Lesya Shchutska (ZURICH-ETH) | IRC |

SUS-16-049 (*Wed, 3 May 2017 09:36:47*)

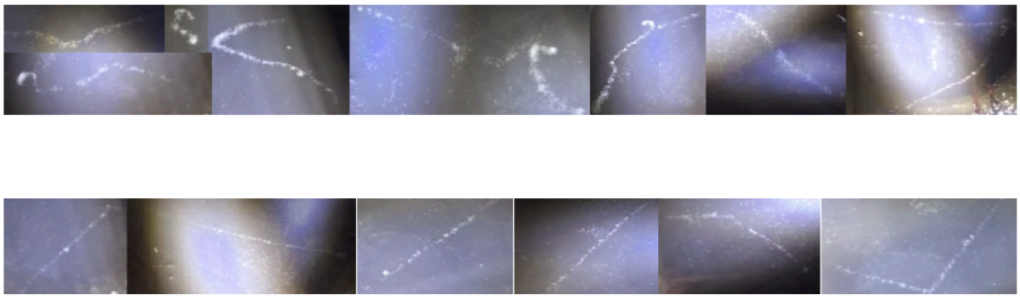
| | | | |
|---|---|---|---|
| **Name** | Search for direct stop pair production in the fully hadronic final state at 13 TeV (Moriond 2017) | **Description** | Search for direct stop pair production in the fully hadronic final state at 13 TeV (Moriond 2017) |
| **Status** | CWR-ended | **Contact Person** | Loukas Gouskos (UCSB) |
| **Twiki** | SUS-16-049 | **HN** | SUS-16-049 |
| **Data,Samples** | DataSet: Run2 Samples: not set | **Conference** | |
| **Target Date PreApp** | 17/02/2017 | **Target Date PhysApp** | 08/03/2017 |
| **Talks** | Pre-Approval Talk » I Approval Talk » | **Actions** | Not in Edit Mode |
| **Related Analyses** | SUS-16-029 | **Related CMS Notes** | AN-2016/437 |
| **Physics Analysis Summary (PAS)** | | | |
| **ARC Chair** | Lesya Shchutska (ZURICH-ETH) | **ARC** | Accepted I hide 4 members<br>Name<br>Muammer Altan Cakir (ISTANBUL-TECH)<br>Lesya Shchutska (ZURICH-ETH)<br>Robert Bainbridge (LONDON-IC)<br>Ivan Mikulec (HEPHY) |
| **PAS Actions** | P | **PAS CDS id** | 2256439 [CDS Marc] |
| **PAPER** | | | |
| **Target Journal** | JHEP | **Target Date Pub** | |
| **AuthorList** | Current AL (not finalized) | **IRC** | Waiting I show 6 readers A B C D E F |
| **PAPER Actions** | | **PAPER CDS id** | |
| **arXiv** | | **DOI** | |
| **HepData** | | **Rivet Plugin tar file** | |

**4.2 Peltier Tabanlı Bulut Odası ve Parçacık İzleri Sınıflandırma Çalışmaları**: Bu zaman içerisinde kozmik olarak gelen parçacıkların davranışlarının anlaşılması ve buna paralel olarak rastgele oluşan durumların yapay zeka teknikleri aracılığı yapısal olarak anlaşılabileceğine dair, o zamanlar lisans öğrencim olan Sevda Yeliz Nar aracılığı ile Peltier Tabanlı Bulut Odası (Peltier Based Cloud Chamber") çalışmaları yürütülmüş ve kendimiz tarafından hazırlanmış bir deney düzeneği üzerinden kozmik olarak dünyamıza ulaşan parçacıkların sınıflandırılması ve analiz edilmesi çalışılmıştır. Bu çalışma öncesinde TFD17 konferasında poster sunum arkasından da American Institute of Physics Conference Proceeding olarak lisans öğrencimle ile birlikte uluslararası bir ortamda yayınlamıştır. Bu çalışma konsept olarak basit olmak ile birlikte, veri toplamamıza yardımcı olması ve parçacıkların sınıflandırmalarının öğretilmesi ile birlikte şu anda derin öğrenme tekniklerini uyguladığımız ve gelen kozmik ışınların otomatik olarak öğretildiği bir yapay zekâ algoritması kullanılmaktadır. **Bu deney düzeneğinden toplanan verinin incelenmesinde, proje kapsamında yazdığımız analiz kodu ve çözümleme araç gereçleri kullanılmaktadır ve grup elemanlarının ortaklaşa çalıştıkları bir platform yaratılmıştır.**

Peltier Temelli Bulut Odası Deney Düzeneği

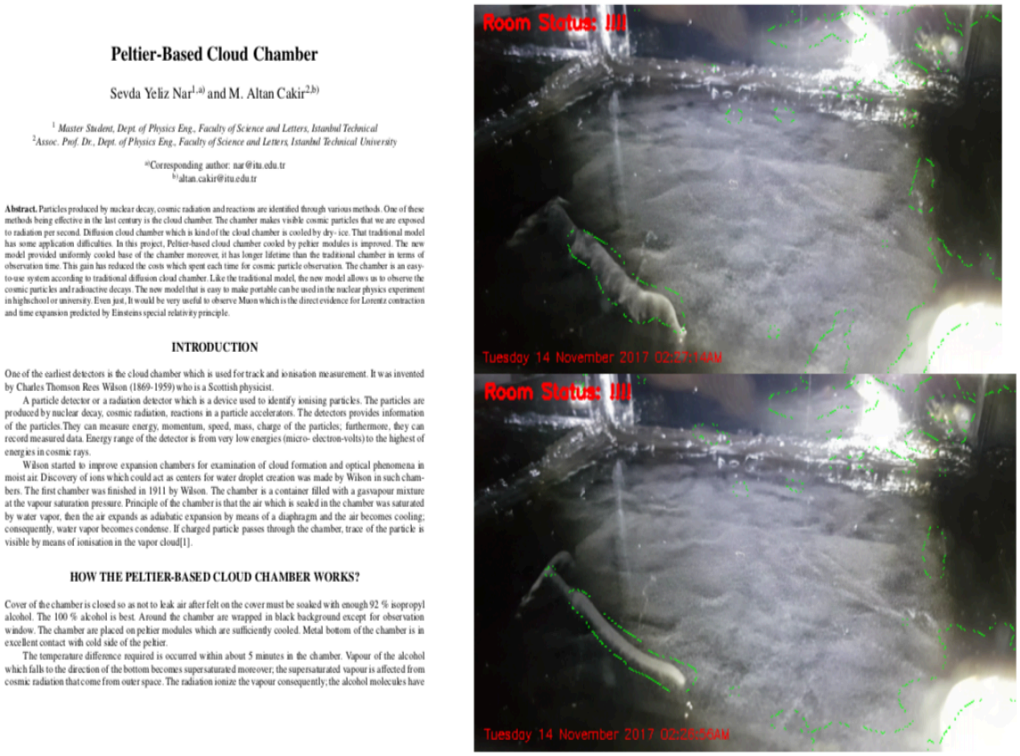


**Şekil 3.** Kozmik ışınların toplanması ve sınıflandırılması için bizim tarafımızdan geliştirilmiş Peltier Tabanlı Bulut Odası deney düzeneği. Düzenek İTU Data Science Laborotuvarında yer almaktadır.



**Şekil 4:** Peltier Tabanlı Bulut Odası deney düzeneği tarafında toplanan kozmik ışınların iz karakterlerine göre sınıflandırılmış fotoğrafları.

Kendi deney düzeneğimiz ile kamera aracılığı ile aldığımız görüntleri, proje kapsamında öğrendiğimiz derin öğrenme algoritmalarına uygulayarak sonuçlarımızı daha geniş bir bakış açısı ile derin öğrenme teknikleri ile daha önce uygulanmamış bir yapıya dönüştürdük. Böylece, gelen parçacıkların **izleri**:




**Şekil 4.** Kozmik ışınların toplanması ve sınıflandırılmasın da kullanılan Peltier Tabanlı Bulut Odası deney düzeneğinin iler seviye kamera görüntüleri ile anlık olarak gözlemlediği ve algoritmik olarak sınıflandırıldığı çalışmalar. Düzenek İTU Data Science Laborotuvarında yer almaktadır.

Takip edilerek kendi yaptığımız detektörümüz ile önemli sonuçlar elde ettik. Bu çalışma hala gelişme aşamasında olup, elde edilen sonuçlar proje kapsamında geliştirilen kod ile birlikte sonuçlandırılarak SCI-A indeksli bir dergiye bu yaz içinde gönderilecektir. Bu çalışma, parçacık fiziği kuramsal inceleme, detektör ar-ge ve

görüntü üzerinden derin öğrenme çalışmalarını içeren çok boyutlu bir yapı içermektedir.

**4.3 Derin Öğrenme Tekniklerinin Web Analitiğine Uygulanması:** Bir önceki çalışmaya benzer içerisinde MC simülasyonları ve derin öğrenme analizleri içeren bir diğer lisans projesi, bu proje çerçevesinde kurgulanarak akademi-sanayi iş birliği sonuçlarına göre yapılandırılmıştır. Daha önce gelişme raporlarında bahsedildiği üzere, bizim projemizde kullandığımız teknikleri, yüksek hacim içeren veri setleri üzerinden uygulanması ve burada kullanılan veri, veri indirgeme, derin öğrenme teknikleri başarı ile akademi dışında uygulanabilmiş ve bu çerçevede oldukça önemli (belki fizik bölümlerinde ilgili anabilim dalında ilk) TEYDEB projesi proje yürütücüsü tarafında alınmıştır. Basitçe yapılan çalışmalar ve elde edilen bulgular özetlenmiştir. Günümüzün en önemli konu başlıklarından biri olan makine öğrenimi uygulamalarının, e-ticaret ve web analitiği uygulamaları hakkında Hürriyet gazetesi ile birlikte "Makine Öğrenimi Teknikleri Aracılığı ile Kullanıcı Davranış, Hareketlerini Analitik Ölçme ve Tahminleme Motorunun Geliştirilmesi" isimli Teydep 1503 projesi (No: 3170089) Şubat 2017'de gönderilmiş (3001 projesinin birinci yılında) ve projem başarı ile kabul edilmiştir. Bu projede 9 numaralı iş paketi (en son) olarak belirtilen akademi-sanayi iş birliği Hürriyet ar-ge çalışmaları sağlanarak önemli bir başarıya imza atılmıştır. Benzer projenin devamında kapsamda projenin TÜBİTAK üzerinden devamlılığı ve öğrenci başarısının devamı için 1001 projesi yazılmaya başlanmıştır. Proje yürütücüsü bu projede danışman olarak çalışmaktadır.




**Şekil 6.** Parçacık fiziğinde yer alan çözüm ve uygulamaların, Hürriyet gazetesi özelinde, bilimsel bir konu dışında benzer uygulamarını içeren günlük operasyonel bir görüntü. Site içerisinde yer alan trafik, trafiğin yaratıldığı bağlantı noktalarına göre sınıflandırılarak algoritmik olarak izlenmektedir.

Aynı şekilde lisans öğrencim Mert Candar ile birlikte, akademi sanayi işbirliği kapsamında Endüstriyel uygulamalar kapsamında yürütülen iş-zekâsı problemlerinde derin öğrenim metotlarının kullanımı, gerçek şirket verileri ile uygulanarak çalışılmıştır. Bu konu başlığı özelinde bilgisayar mühendisliği ve endüstri mühendisliği bölümleri kapsamında yayınlanmak üzere makale çalışmaları yürütülmektedir. İlgili makalenin yazımına başlanmıştır önümüzdeki

yaz ayı içerisinde tamamlanacaktır. Bu aşamada öğrendiğimiz uygulamaları Hürriyet gazetesinden alınan veriler ile web analitiği üzerine uygulanarak veri üzerinden yapısal benzerlikler çıkartma analizleri yapılırken ve Newegg isimli elektronik e-ticaret platformunun satış verileri (Amerika Birleşik Devletleri) üzerinden zaman serilerine bağlı zamansal korelasyon verilerinin derin öğrenim metotları aracılığı ile uygulanması incelenmiş ve ilk çözümler önerilmiştir.

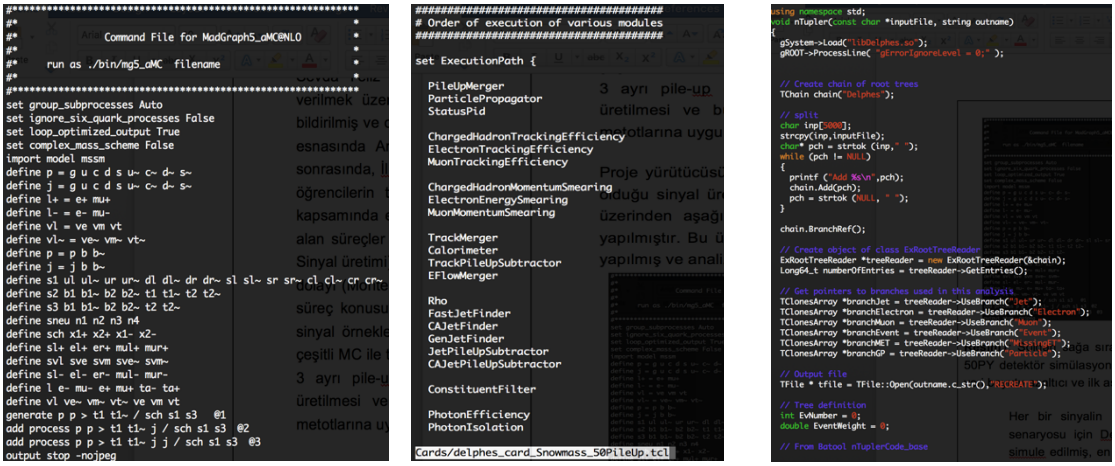



**Şekil 7.** Proje kapsamında öğrenilen MVA ve derin öğrenim tekniklerinin Amerika Birleşik Devletleri'nde yer alan e-ticaret şirket verisine uyarlanarak ürünler arası korelasyon haritaları, ürünlerin sipariş verilmeden önce tahminlenmesi ve stok kontrol yönetimlerininin iler seviye istatiksel çözümleme metotları ile tahmin grafikleri. En üst – örnek iki ürün arasında korelasyon, çapraz korelasyon ilişikileri. Orta-kısım gerçek veri (yeşil) ve analiz tahminlerinin (mavi) Amerika birlişk devletlerin içersinde satış kategorisi harita dağılımları (R-TMVA kullanıldı). En son derin öğrenim için kullanılan Eğitme/Validasyon ve Test (gerçek veri) ile tahminleme karşılaştırılması (mavi renk).

**4.4 LHC'de Tek Lepton Kanalında Direk Stop Araştırmaları ve Derin Öğrenme Tekniklerinin Uygulanması:** Proje kapsamında ele alınan başlıkların öğrenilmesi ve uygulanması oldukça zahmetli ve zaman alan süreçler olmaktadır. Özellikle makale için kullanılması planlanan 6 numaralı (Ardalan-Sinyal üretimi) iş paketinde belirtilen yeni fizik sinyallerinin teknik üretilme süreci çeşitli aksamalar yaşamasından dolayı (Monte-Carlo olay üretimi + detektör simülasyonu + analiz + derin öğrenim analizi) bu süreç konusunda beklenmedik bir zaman kaybı söz konusu olmuştur. Özellikle ardalan-sinyal örneklerinin Monte-Carlo (MC) olay üreteci aşamasından başlanılarak üretim süreci çeşitli MC ile test edilmesi (Pythia, Madgraph, Powheg ve/veya Sherpa), hizlı detektör simülasyonunun 3 ayrı pile-up (PU) senaryosu değerlendirilmesi (NoPU/50PU ve 140PU) kullanılarak üretilmesi ve bunların doğru analiz edilerek derin öğrenim ve çok değişkenli analiz metotlarına uygulanması zaman almaktadır.

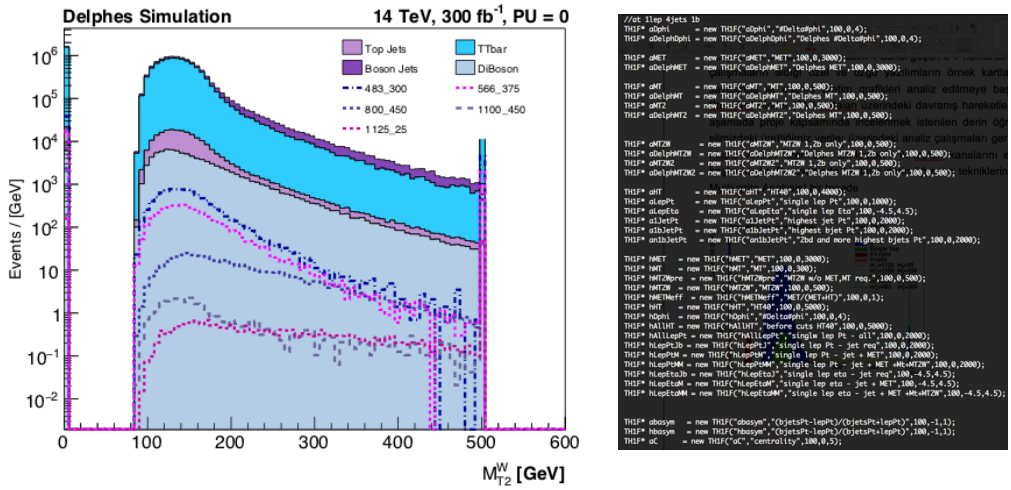
Proje yürütücüsü bu ve benzeri durumlar karşısında araştırmaların devamlılığı için uzmanı olduğu sinyal üretimi ve hesaplanması çalışmalarına başlamıştır. Aşağıda yer alana Şekil 1'de-sol, Madgraph MC olay üreteci ile basitleştirilmiş SUSY sinyalleri üretimi yapılmıştır. Basitleştirilmiş modeller (SMS) genel anlamda tesir-kesitini baskılayan eden kanalların ele alındığı üretim prosesleridir. Bu basitleştirilmiş modeller ilgili prosesin belli üretim kanallarını sadece ele alan ve bu kanalların her bir alt-kanal için %100 dallanma oranı ve/veya belli oranlarda azalarak indirgendiği durumları betimlemektedir. Bizim elimizdeki sinyaller sinyal değerlerine göre (birince sayı – stop kütlesi/ ikinci sayı – en hafif SUSY parçacık kütlesi) betimlemektedir. Bu çerçevede bilindiği üzere stop kütlesinin top ve LSP kütlesine bozunduğu ve %100 dallanma oranı ile bazı sinyallerin bu çerçevede ele alındıgını (özellikle yüksek değerli stop kütleleri >350 GeV) bölgeler, stop-LSP arasında kütle farkının azaldığı bölgeler ve aradaki farkın çok yüksek olduğu bölgelere bakarak genel bir senaryo süreci ele alınmaktadır. Bu nedenle farklı kütle değerleri içeren sinyal noktaları üretilmiştir.

Üretilen noktalar Süpersimetri modeli uyarınca çalışılan faz uzayında kalmış ve hala tam dışarlanmamış sinyal noktalarıdır. Bu noktalarının incelenmesi tesir kesitleri başta olmak üzere, dallanma oranlarının da %100 alınmasında dolayı çeşitli zorluklar içermektedir. Bu nedenle kesmeye dayalı analizlere ek olarak TMVA yapısı ile birlikte çok değişkenli algoritmik bir çözümleme optimizasyonu kurgulanması en uygun tekniklerden biri olarak ortaya çıkmaktadır.




**Şekil 8**. Soldan sağa sırasıyla, Madgraph basitleştirilmiş SUSY kanalı üretim kartı (sol), Delphes 50PY detektör simülasyonu üretim modülleri kartı ve ROOT ile yazılmış Delphes ROOT dosyaları üzerinden projeye özgü veri boyutu azaltıcı (ntuple) kodun örneği (sağ).

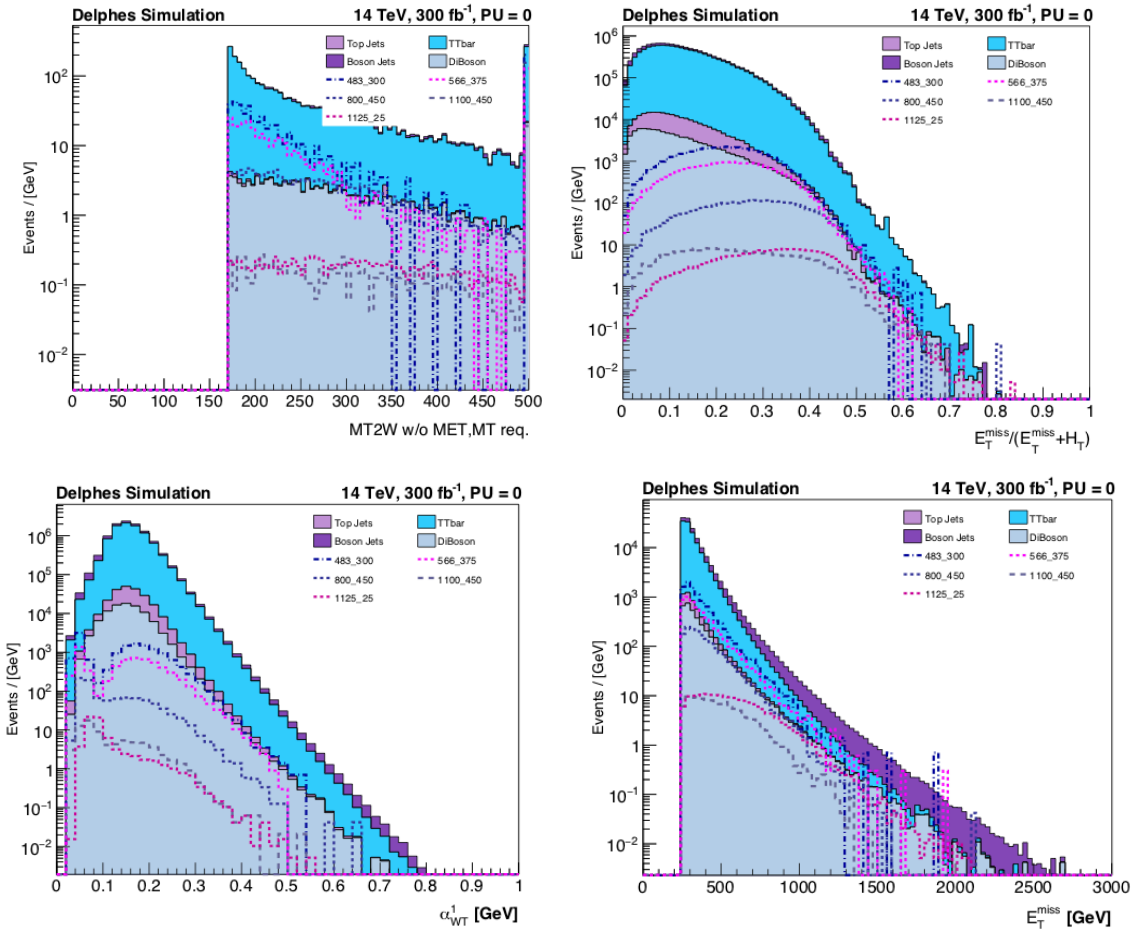
Her bir sinyalin kuramsal değerlerinin üretilmesinden sonra, ilgili üretim kanalı 50PU senaryosu için Delphes detektör simülasyonu ve üretim kartı (Şekil 1-orta) kullanılarak simule edilmiş, en son olarak üretilen ROOT dosyasının boyutunun küçültülmesi için ntuple yapılarak (Şekil 1-sağ) ve analizde ardalan-sinyal ayrımı için çalışılmaya başlanmıştır. Şekil 1'de bahsi geçen 6-7 numaralı iş paketleri kapsamında yapılan çalışmaların aldığı özel ve özgü yazılımların örnek kartları gösterilmiştir. Bu çalışmalar sayesinde ilk aşama üretim grafikleri analiz edilmeye başlanmış ve istenilen sinyallerin Standart Model (SM) ardalan üzerindeki davranış hareketleri detaylıca incelenmektedir. Bu aşamada proje kapsamında incelenmek istenilen derin öğrenim- yapay-zekâ tekniklerinin elimizdeki ürettiğimiz veriler üzerindeki analiz çalışmaları gerekmektedir. Bu nedenle üretilen ntuple verileri üzerinden tek-lepton / iki-lepton kanalarını ele alacak şekilde incelemelere başlanmıştır. Bu aşamada bir tane derin öğrenim tekniklerini uygulandığı TMVA (Toolkit for Multivarite Analysis) bir tanede analiz kodu yazılmıştır. MVA öncesi SM ardalan-sinyal çözümlemesinin başta MT2W olarak tanımlanan SUSY değişkenin grafiği çizilmiştir (arXiv: 1203.4813). Bunlara paralele Topness ve kendi yeni tanımladığız yayınlanmak üzere olan değişken tanımımız aşağıda gösterilmiştir.



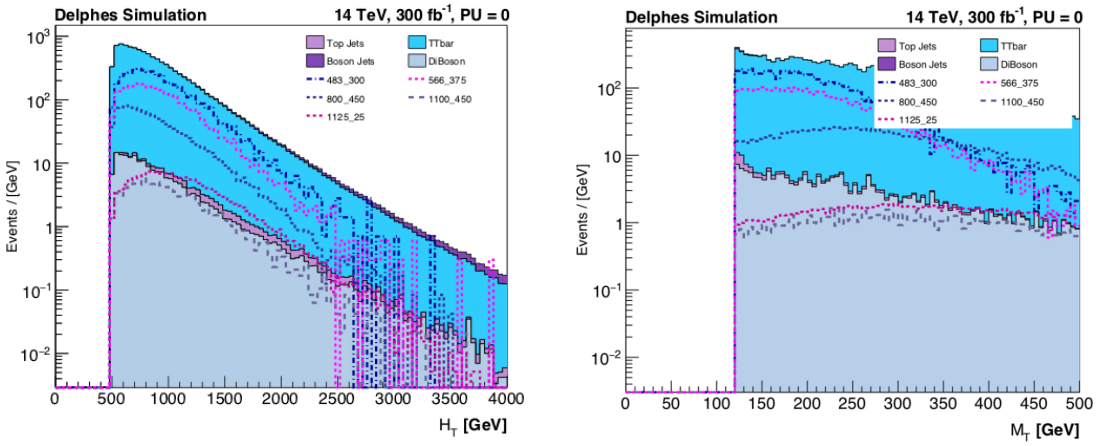

**Şekil 9**. Sol taraf, MT2W değişkeni yardımı ile üretilen farklı SUSY sinyallerinin ardalan üzerinden ayrıştırma çalışmaları. Sağ taraf, ntuple dosyaların üzerinden yazılan özgün analiz kodundan bir kesit. Kod içerisinde yazan incelenen histogramların basit bir listesini vermektedir.

MVA öncesi ve sonrası sinyal ayırma çalışmalarında burda 300/fb veri normalizasyonu kullanılarak bazı katı kesmeler uygulanmıştır fakat bu seviyede katı kesmelerin uygulanması çok sağlıklı ve keşfi kolaylaştıran durumlar değildir. Bu nedenle kesmeler daha hafif uygulanarak birbirine bağlı değişkenler üzerinde çok değişkenli analiz yapılmaktadır. Bununla ilgili sonuçlar ve sinyal üretimleri (özellikle 50PU/140 PU) sürmekte ve oldukça fazla CPU zamanı almaktadır. Örnek olarak rakamsal bir ölçek verilecek olunursa 10K olay için olay üretimi, NoPU/50PU ve 140 PU olaylarının üretilerek ntuple yapılması ortalama 1 gün sürmektedir. Toplamda her sinyal için en az 1 milyon mertebesinde istatistik gerektiğini için, hesaplamaların hata yapılmadan doğru bir şekilde yürütülmesi gerekmektedir ve şu anda aktif olarak yürütülmektedir. Bununla ilgili detaylar ve teknik verileri ilişikte referans aracılığı ile görebilirsiniz. Derin öğrenim kapsamında kullanılan algoritmik ayıraçlar arasındaki farklılıklar NO/50PU ve 140PU durumları arasında farklılık göstermekte (kayıp enerji – MET – kontrolü zorlaşmakta) ve aynı zamanda ayırma değerleri arasında senaryolara göre belirsizlikler ortaya çıkmaktadır. Bu nedenle karmaşık (birçok değişken ve obje bilgisi içeren) yeni değişkenler tanımlanmaktadır. MT2W olarak daha önce bahsedilen değişken SM ve ötesi analizlerde oldukça fazla olarak kullanılmaktadır. Bu referans listesinde yer alan kod (http://www.hep.phy.cam.ac.uk/~lester/mt2/) ile birlikte proje kapsamında geliştirilen analiz koduna girilerek ortak çalışması sağlanmış ve birçok sonlu LHC durumuna uygun hale getirilmiştir. Bu değişkenin bile proje kapsamında yazılan analiz koduna girilmesi kendi başına bir başarıdır. Daha önce gösterilen, BDT (ötelenmiş karar ağaçları – boosted decision trees), kNN (k-en yakın nöral network), Likelihood perseptron metotları HT, MT ve MET değişkenlerine ek olarak MT2W ile ve ayrıca MT2W/√MET, MT2W/√HT, /√HT+MET/ MT2W tanımlanmıştır. Tüm MVA teknikleri proje kapsamında SM ardalan ve herhangi bir sinyal için uygulanabilir ve araştırılabilinir bir alt-yapı haline dönüştürülmüştür en optimal çözüm (sinyal/ardalan oranı) aranmaktadır. İlgili detayların tamamlanmasında sonra ilgili makale çalışmaları tamamlanacaktır. 6-7 numaralı iş paketlerinin tamamlanmış, ilgili bilimsel çalışmaların dokümantasyonu, derin öğrenim

metotlarının uygulamalı karşılaştırmaları, makale ve sözlü bildiri sunum çalışmaları ile birlikte yürütülmektedir.




**Şekil 10.** Süpersimetri sinyallerini ardalan'dan ayırmak için çeşitli değişkenlerin tek-lepton kanalı sonlu durumunda davranışlarının NoPU (pile up'sız senaryo). Burada tanımlanan $\alpha_{WT}$ tanımlanan değişken √HT+MET / MT2W olarak tanımlanmaktadır ve proje kapsamında çalışılmaya başlanmıştır.




**Şekil 11.** HT ve MT dağılımlarının sinyal kesmelerinden sonraki sinyal-ardalan durumlarının tek-lepton sonlu durumu için NoPU (pile up'sız) senaryosunda gösterilmesi.

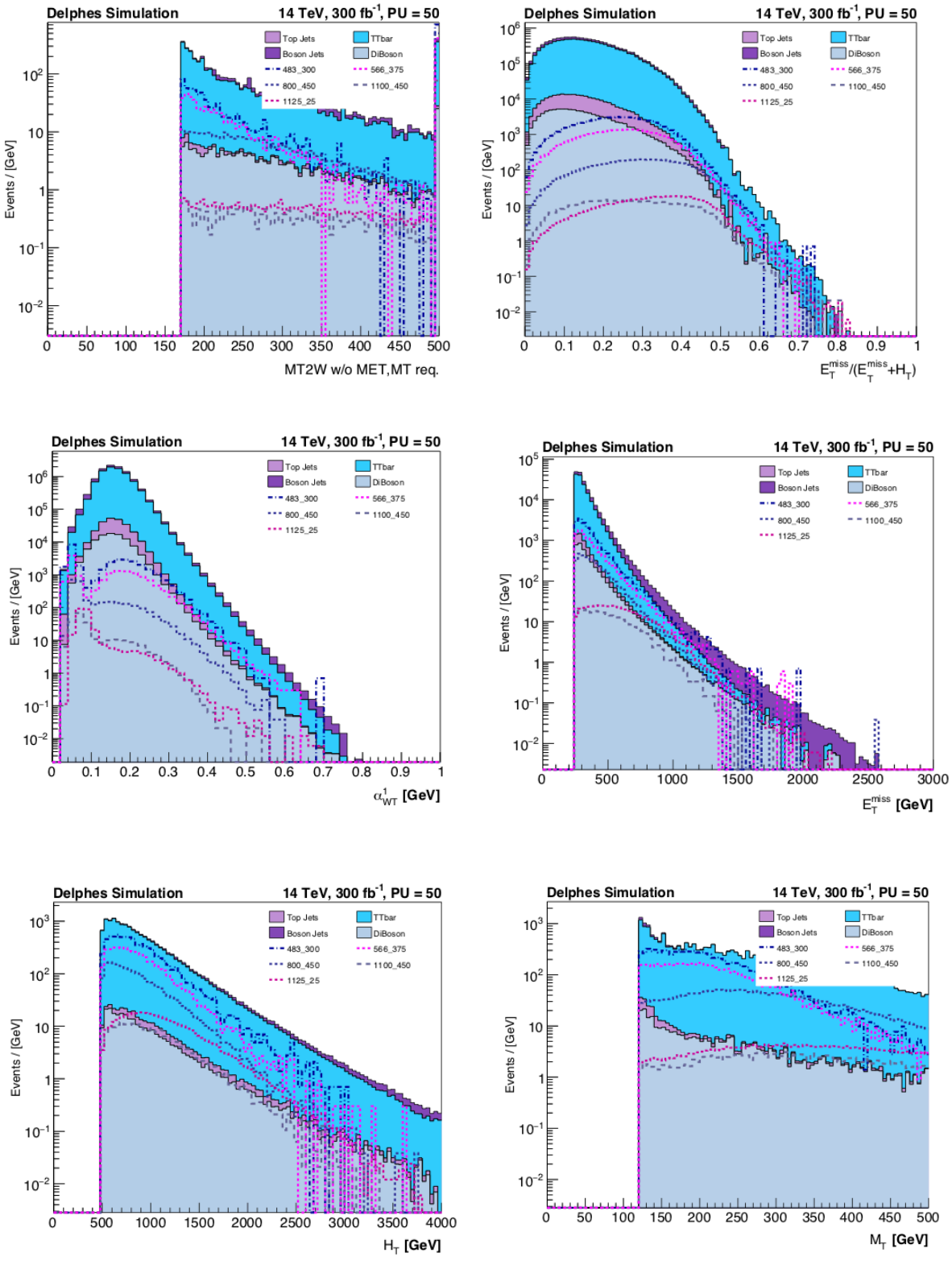
Yukarıda NoPU için gösterilen HT, MET ve MT gibi değişkenler hemen hemen tüm yeni fizik arayışlarında kullanılan değişkenlerdir. Son zamanlarda tanımlanan MT2W, Meff ve benzeri değişkenler HT, MET, MT ve objelerin Pt değerleri üzerinden tekrar yapılandırılarak birbirleri arasında açısal bağımlılıklara bağlı olarak optimize edilen bir yapıya sahiptir. Bizde analizimizde benzer bir şekilde $\sqrt{}$HT+MET / MT2W değişkeni çalışılan tüm değişkenlere nazaran toplamda iyi sonuçlar verdiğini gözlemledik. Hala çalışma aşamasında olan analizimiz, jenerik değişkenlere göre daha iyi oransal bir ayrım verdiğini, aynı zamanda sadece ttbar ardalan ile birlikte kıyaslanabilir bir yapıya taşıdığını göstermektedir. Ardalan dalgalanması farklı PU senaryoları ve luminosite değerleri için ayrıca çalışılmış ve Tablo X'de gösterilmiştir. Tablo en son optimal hale getirilmemiş olmakla birlikte $\alpha_{MT}$ değişkenin SUSY'nin bulunamayan zor faz uzayında diğer değişkenlere göre önemli davranışlar gösterdiği görülmektedir. Bu davranış şekli, farklı ardalan kompozisyonu üzerindeki davranış şekillerinin farklı olması, TTbar dışında diğer diğer katkıların baskılandığını, TTbar ardalan artışı olmasına rağmen, sinyal sayısında da bir artış olduğunu göstermektedir. İlk aşama olarak bu şekli ile TMVA analizine beslenecek dağılımlar için No, 50PU ve 140PU senaryoları için 300/fb ve 3000/fb luminosite değerleri için Tablo 2'de gösterilmektedir. Yukarıda tartışılan tüm değişkenlerin 50PU için çizilen dağılımları Şekil 12'de gösterilmektedir.

**Tablo 2.** No, 50PU ve 140PU senaryoları için 300/fb ve 3000/fb luminosite değerleri SM ardalana ve üretilen SUSY senaryoları için farklı değişkenler ile birlikte gösterilmiştir.

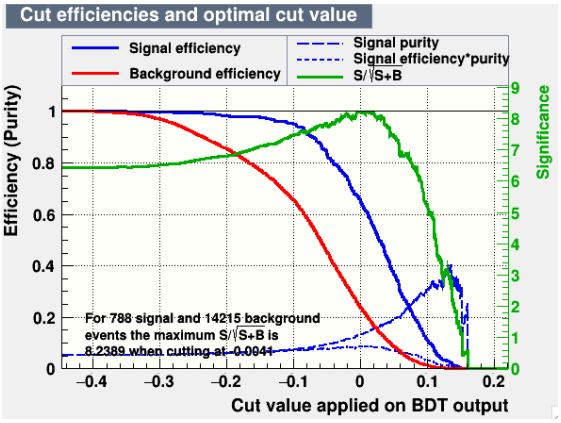
| Cutflow | VV+jets | $t\bar{t}$+jets | V+jets | Top+jets | Total SM | 483_800 | 566_375 | 800_450 | 1100_450 | 1125_25 |
|---|---|---|---|---|---|---|---|---|---|---|
| #Events | 110817000 | 215894000 | 16840400000 | 62062700 | 17229173700 | 840000 | 360000 | 42000 | 4200 | 3614 |
| Lepton(e,$\mu$) | 10848100 | 45875000 | 579264000 | 7938970 | 643926070 | 139648 | 60175 | 7170 | 661 | 467 |
| nJets $\geq$ 4 | 271691 | 11748100 | 4432360 | 167232 | 16619383 | 82612 | 36794 | 4935 | 400 | 327 |
| bjets $\geq$ 2 | 100583 | 10636400 | 1325830 | 151947 | 12214760 | 52419 | 23475 | 3293 | 266 | 227 |
| $E_T^{miss}$>120 | 3452 | 118366 | 27574 | 1853 | 151245 | 9733 | 5524 | 1733 | 121 | 187 |
| $M_T$>120 | 236 | 12462 | 129 | 26 | 12853 | 5087 | 3123 | 1358 | 102 | 166 |
| $H_T$>500 | 187 | 9896 | 129 | 26 | 10238 | 4731 | 2869 | 1145 | 96 | 159 |
| $\alpha_1 > 0.2$ | 48 | 7066 | 45 | 20 | 7179 | 4107 | 2463 | 861 | 69 | 113 |

| Cutflow | VV+jets | $t\bar{t}$+jets | V+jets | Top+jets | Total SM | 483_800 | 566_375 | 800_450 | 1100_450 | 1125_25 |
|---|---|---|---|---|---|---|---|---|---|---|
| #Events | 110822000 | 216124000 | 16842600000 | 62086200 | 17231632200 | 840000 | 360000 | 42000 | 4200 | 3614 |
| Lepton(e,$\mu$) | 10583400 | 45304000 | 563438000 | 7671830 | 626997230 | 191642 | 82646 | 10401 | 995 | 861 |
| nJets $\geq$ 4 | 298204 | 11059200 | 5276130 | 190355 | 16823889 | 95635 | 43132 | 6759 | 561 | 583 |
| bjets $\geq$ 2 | 106821 | 9991990 | 1500560 | 169733 | 11769104 | 80261 | 36424 | 6111 | 505 | 550 |
| $E_T^{miss}$>120 | 3843 | 140908 | 30865 | 2355 | 177971 | 17348 | 9777 | 3374 | 271 | 461 |
| $M_T$>120 | 340 | 17711 | 722 | 90 | 18863 | 8764 | 5377 | 2641 | 227 | 407 |
| $H_T$>500 | 262 | 13210 | 618 | 82 | 14172 | 7595 | 4564 | 2092 | 210 | 390 |
| $\alpha_1 > 0.2$ | 73 | 9165 | 147 | 55 | 9440 | 6155 | 3653 | 1295 | 115 | 209 |

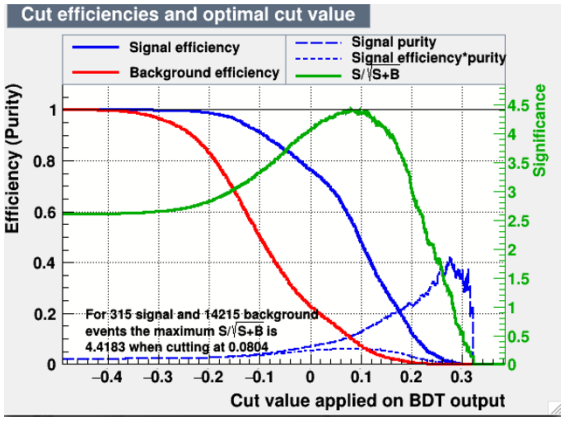
| Cutflow | VV+jets | $t\bar{t}$+jets | V+jets | Top+jets | Total SM | 483_800 | 566_375 | 800_450 | 1100_450 | 1125_25 |
|---|---|---|---|---|---|---|---|---|---|---|
| #Events | 1108210000 | 2156950000 | 168402000000 | 620665000 | 172287825000 | 8400000 | 3600000 | 420000 | 42000 | 0 |
| Lepton(e,$\mu$) | 104202000 | 460047000 | 5500730000 | 77172800 | 6142151800 | 1963150 | 847683 | 106600 | 10171 | 0 |
| nJets $\geq$ 4 | 5104540 | 119163000 | 144145000 | 4538450 | 272950990 | 876057 | 645645 | 66795 | 5593 | 0 |
| bjets $\geq$ 2 | 1561740 | 103615000 | 33608600 | 3634640 | 142419980 | 677068 | 468396 | 58038 | 4874 | 0 |
| $E_T^{miss}$>120 | 62324 | 2370920 | 611508 | 59745 | 3104497 | 168532 | 120735 | 32989 | 2789 | 0 |
| $M_T$>120 | 10126 | 504474 | 76000 | 10804 | 601404 | 86198 | 64926 | 25342 | 2312 | 0 |
| $H_T$>500 | 6621 | 274810 | 38917 | 5417 | 325765 | 69482 | 58767 | 19529 | 2116 | 0 |
| $\alpha_1 > 0.2$ | 2107 | 179569 | 12438 | 3549 | 197663 | 56658 | 41103 | 12559 | 1205 | 0 |



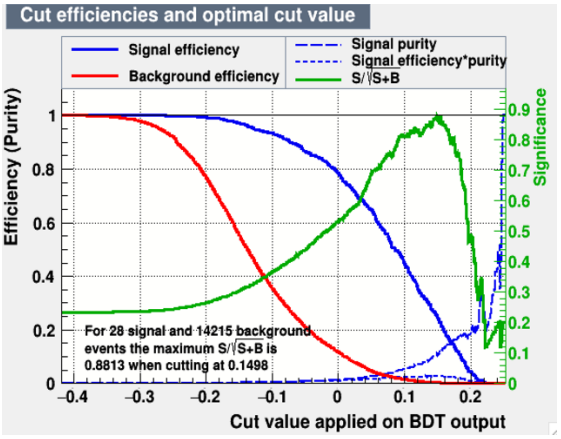

**Şekil 12.** Süpersimetri sinyallerini ardalan'dan ayırmak için çeşitli değişkenlerin tek-lepton kanalı sonlu durumunda davranışlarının 50PU. Burada tanımlanan $\alpha_{WT}$ tanımlanan değişken √HT+MET/MT2W olarak tanımlanmaktadır ve proje kapsamında çalışılmaya başlanmıştır.



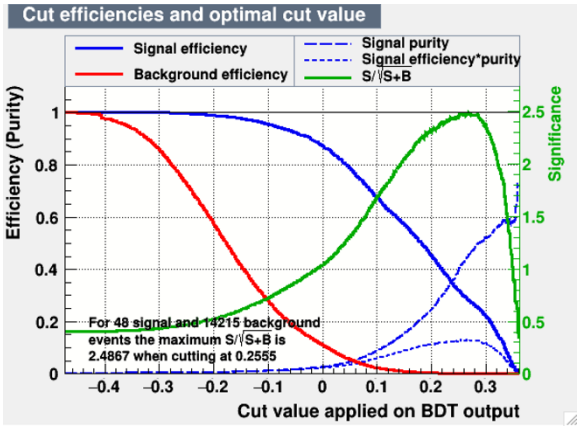

(a)




(c)

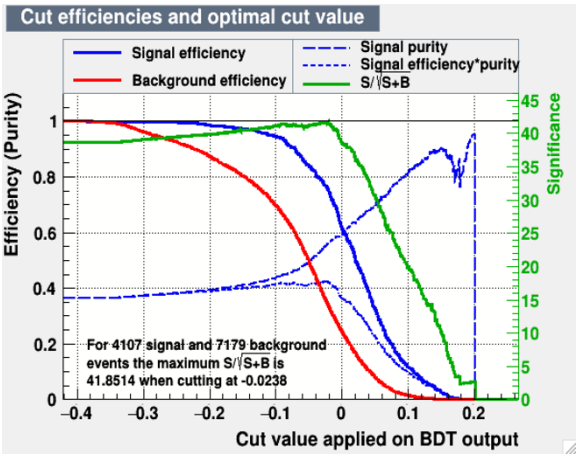



(b)

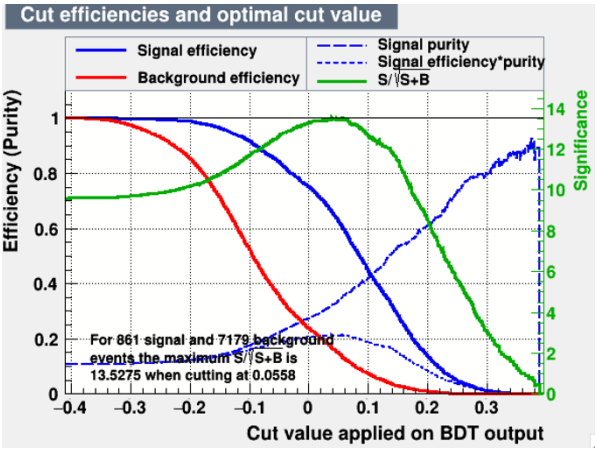



(d)

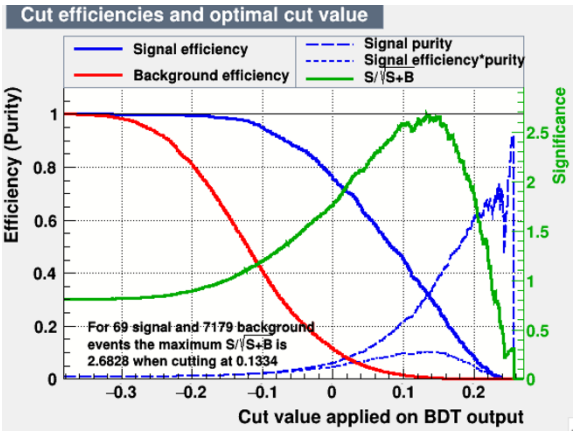
**Şekil 13.** NoPU senaryosu için üretilen Süpersimetrik top kütlesi 483 GeV – Nötralino kütlesi 300 GeV olan sinyal (a), Süpersimetrik top kütlesi 800 GeV – Nötralino kütlesi 450 GeV olan sinyal (b), Süpersimetrik top kütlesi 1150 GeV – Nötralino kütlesi 450 GeV olan sinyal (c), Süpersimetrik top kütlesi 1125 GeV – Nötralino kütlesi 15 GeV olan sinyal (d) senaryolarında TMVA üzerinden yapılmış ilk aşama ardalan-sinyal ayrımı istatiksel önem ($\sigma$) analizlerinin dağılım grafikleri.



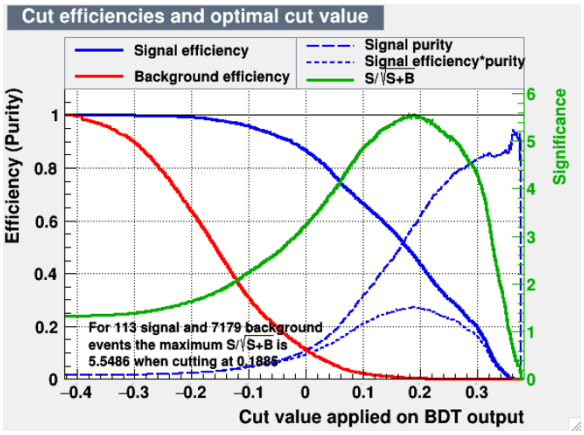

(a)




(c)




(b)

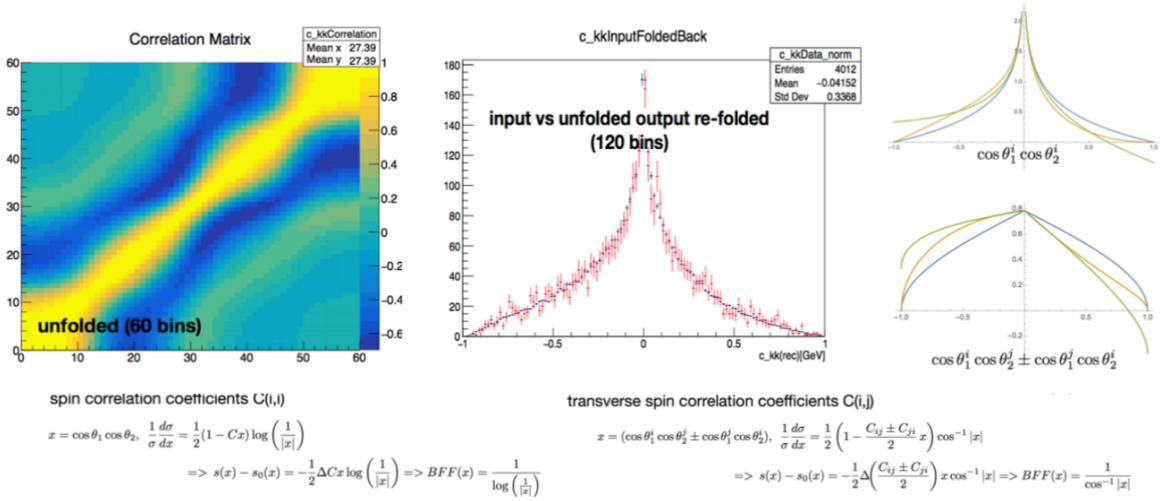



(d)

**Şekil 14.** 50PU senaryosu için üretilen Süpersimetrik top kütlesi 483 GeV – Nötralino kütlesi 300 GeV olan sinyal (a), Süpersimetrik top kütlesi 800 GeV – Nötralino kütlesi 450 GeV olan sinyal (b), Süpersimetrik top kütlesi 1150 GeV – Nötralino kütlesi 450 GeV olan sinyal (c), Süpersimetrik top kütlesi 1125 GeV – Nötralino kütlesi 15 GeV olan sinyal (d) senaryolarında TMVA üzerinden yapılmış ilk aşama ardalan-sinyal ayrımı istatiksel önem ($\sigma$) analizlerinin dağılım grafikleri.

**4.5 TTbar Çiftlerinin Spin Korelasyonu ve Polarizasyon bilgilerinin İki Lepton Sonlu Durumunda CMS Deneyinde Hassas Ölçümlenmesi ve Yeni Fizik Arayışları**: Proje kapsamında yapılan çalışmalara ek olarak CMS kolaborasyonu içerisinde proje yürütücüsünün derin öğrenim araştırmaları aktif olarak kullanmaktadır. CMS deneyindeki uygulamalar konusunda projenin ilk gelişme raporu döneminde döneminde yapılan "Search for direct stop pair production in the fully hadronic final state at 13 TeV" adlı çalışmanın dışında, Fermilab'dan alınan ödül kapsamında oldukça prestijli yeni bir çalışmaya başlanmıştır. En son rapor döneminde belirtildiği üzere, proje yürütücüsü CERN'de yer almakta olan LHC-CMS deneyinde çalışmakta ve bu proje kapsamında çalıştığı konuları gerçek veri analizlerine uygulamaktadır. Bu kapsamda SUSY araştırmalarında önemli bir sonuç arz eden derin öğrenme metotları, top kuark çiftlerinin spin korelasyonları ve polarizasyon gibi oldukça zor SM hassaslık ölçümü, bunun üzerinden SUSY'de yer alan süpersimetrik top kuark keşfi çalışmalarına uygulanmaya başlanmıştır. Bu

konu iki lepton sonlu durumu ele alınmış “Measurement of ttbar spin correlations and top quark polarization in dileptonic channel at 13 TeV” analiz notuna başlanmıştır. İlgili çalışma ile yapılan çalışmalar uluslararası mertebe iş birliği çerçevesinde yürütülmektedir.




**Şekil 15**. TUnfold tekniğini aracılığı ile spin korelasyon ölçümü ve Monte-Carlo kuram modelleme çalışmaları. CosTheata değişkeni korelasyon ayrımı en sağda gösterilmiştir bu şekilde geleneksel ölçümlerin ötesine geçilerek TTbar çiftlerinin açısal olarak ayrılması sağlanacaktır.

Şekil 15’de kuramsal olarak modellenmiş her bir top kuark çiftinin korelasyonlarının çıkartılması ve Monte-Carlo modellemelerine uygulama çalışmaları gösterilmektedir. Standart MC verisi ile TUnfold [arXiv: 1205.6201] ile düzeltilmiş, MC değişkenlerinin birbirleri olan bağımlıkları ayrıştırılmış, ve yeni değişkenler tanımlanmıştır. Bu değişkenler, açışla bağımlılıkları içeren yapılar ve bunlar üzerinden kinematik etkilerin derin öğrenim teknikleri ile optimize edilerek en uygun kesme değerleri tanımlanması ile oluşturulmaktadır. TTbar çiftlerinin spin korelasyon açısı (theta) olarak tanımlanarak beklenen (SM ardalan) ve beklenmedik sapmalar (SM ötesi) ayrılma çalışmaları yürütülecektir. Bilindiği üzere SUSY’den gelen top kuark çiftleri tamamen korele değilken, aynı kuplaj değerlerinden gelen top kuark çiftleri tamamen korele durumda olmalıdırlar. Böylece sıradan analizler aracılığı ile ayrıştırılamayan SUSY top kuark çiftlerinin en zor bölgede (Top kuark’a yakın) ayrıştırılması hedeflenmektedir. Sonuçlar CMS verisi ile incelenmeye başlanmış, ve ilk sonuçlar için Haziran 2018 ayı hedeflenmiştir. Bu kapsamda ITU-TECH olarak CERN’deki etkinliğimiz arttırılmıştır. Bu kısımda bu proje ilgili tek kısım, derin öğrenim teknikleri için oluşturulan alt-yapıdan yararlanılmasıdır. Problemin karmaşıklığının proje içeriği olarak değil, projenin bir uygulaması olarak değerlendirilmelidir.

# 5. SONUÇ

Projenin öncelikli amacı, derin öğrenim konularında alt-yapı çalışmaları yürütmek ve bu çerçevede yüksek lisans ve lisans öğrencilerinin TUBİTAK tarafında belirlenen burslar aracılığı ile eğitimlerinin sağlanması olarak hedeflenmiştir. Projenin başladığı tarihten itibaren (15.05.2016) çeşitli başta akademi olmak üzere tüm Türkiye'yi etkileyen üzüntü verici olayların yaşanması, aynı zamanda proje öğrencisi İlknur Köseoğlu'nun yurt dışında (bu proje kapsamında aldığı Fermilab burs desteği ve diğer gelişmeler) okumak istemesi ve bu yönde bölümden ayrılma kararı almasında dolayı, kendisi ile olan proje çalışmalarımızda bilgi aktarımı konusunda ve proje çerçevesinde daha önce lisans kursiyeri olarak çalışmış Sevda Yeliz Nar isimli yüksek lisans (daha önce ki lisans kursiyeri) öğrencisine destek verilmek üzere değişiklik yapılmıştır. Bu gelişmeler, kararın alındığı Eylül ayı itibari ile bildirilmiş ve değişim süreci, resmi olarak sorunsuz bir şekilde Ekim ve Kasım aylarında halledilmiştir. Ancak, bu süreç esnasında Amerika'da yer alan Fermilab'da yürüttüğümüz çalışmalarımız (Measurement of Spin Correlation and Polarization Top Quark Pairs at 13 TeV and Phase II Tracker Validation Studies) ve sonrasında, İlknur'un ayrılmasından kaynaklanan, bazı teknik işlerin yarıda kalması ve diğer öğrencilerin tekrar hazırlanması konularında bir zaman uzamasına yol açmıştır. Yukarıda yer alan Bulgular ve çalışmalar bölümünde anlatıldığı üzere, proje ana çerçevesi üzerinde akademi ve sanayi özelinde birkaç çalışma eş zamanlı olarak yürütülmüştür. Proje kapsamında ele alınan başlıkların öğrenilmesi ve uygulanması oldukça zahmetli ve zaman alan süreçler olmasına rağmen, 3001 özelinde verilen başlangıç ar-ge desteği oldukça verimli bir şekilde kullanılarak önemli çalışmalar yürütülmüş ve hala yürütülmektedir. Özellikle **4.1**, **4.4** ve **4.5** numaralı kısımlarda yürütülen makale çalışmaları için kullanılması planlanan 6 numaralı (Ardalan-Sinyal üretimi) iş paketinde belirtilen yeni fizik sinyallerinin teknik üretilme süreci durmasından dolayı (Monte-Carlo olay üretimi + detektör simülasyonu + analiz + derin öğrenim analizi) bu süreç konusunda beklenmedik bir zaman kaybı söz konusu olmasında dolayı proje bitiminde önce gönderilmesi planlanan makale çalışmaları tam istenildiği ölçüde olgunluk seviyesine ulaşmamıştır. Özellikle ardalan-sinyal örneklerinin Monte-Carlo (MC) olay üreteci aşamasından başlanılarak üretim süreci çeşitli MC ile test edilmesi (Pythia, Madgraph, Powheg ve/veya Sherpa), hızlı detektör simülasyonunun 3 ayrı pile-up (PU) senaryosu değerlendirilmesi (NoPU/50PU ve 140PU) kullanılarak üretilmesi ve bunların doğru analiz edilerek derin öğrenim ve çok değişkenli analiz metotlarına uygulanması zaman almaktadır. Sonuç raporunda sunulduğu üzere elde edilen tablo ve yeni değişken yapılarının LHC deneyinin önümüzdeki yıllarda toplayacağı veri miktarı ve PU senaryoları özelinde çalışılan yeni $\alpha_{MT}$ değişkeni, TMVA çalışmalarında diğer değişkenlere göre zor sinyal bölgelerinde daha iyi sonuçlar vermiştir. Bunlar yayınlanması planlanan makale için çalışılmaktadır.

Proje yürütücüsü bu ve benzeri durumlar karşısında araştırmaların devamlılığı için uzmanı olduğu sinyal üretimi ve hesaplanması çalışmalarına başlamıştır. Madgraph MC olay üreteci ile basitleştirilmiş SUSY sinyalleri üretimi yapılmıştır. Basitleştirilmiş modeller (SMS) genel anlamda tesir-kesitini baskılayan eden kanalların ele alındığı üretim prosesleridir. Bu basitleştirilmiş modeller ilgili prosesin belli üretim kanallarını sadece ele alan ve bu kanalların her bir alt-kanal için %100 dallanma oranı ve/veya belli oranlarda azalarak indirgendiği durumları betimlemektedir. Bizim elimizdeki sinyaller sinyal değerlerine göre (birince sayı – stop kütlesi/ ikinci sayı – en hafif SUSY parçacık kütlesi) betimlemektedir.

Bu çerçevede bilindiği üzere stop kütlesinin top ve LSP kütlesine bozulduğu ve %100 dallanma oranı ile bazı sinyallerin bu çerçevede ele alındığını (özellikle yüksek değerli stop kütleleri >350 GeV) bölgeler, stop-LSP arasında kütle farkının azaldığı bölgeler ve aradaki farkın çok yüksek olduğu bölgelere bakarak genel bir senaryo süreci ele alınmaktadır. Bu nedenle farklı kütle değerleri içeren sinyal noktaları üretilmiştir.

Proje kapsamında **4.2** numaralı kısımda anlatılan daha önce lisans kursiyeri olarak çalışan ve şu anda projenin tek yüksek lisans öğrencisi ile yürüttüğümüz çalışmalar kapsamında kozmik parçacıkların öncelikle gözlemlenebildiği bulut odası çalışmaları (Peltier temelli) sonrasında görüntü isleme teknikleri aracılığı ile parçacıkların ayrıştırılması ve sınıflandırılarak yapay zekâ tekniklerinin uygulamaları çalışılmaya başlanmıştır. Bu proje 3001 kapsamı ötesinde olmasına karşın görüntü̈ işleme teknikleri ile derin öğrenim metotlarının birleştirilmesi ve parçacıkların sınıflandırılmasını içermektedir ve oldukça önemli sonuçlar vermektedir. Lisans öğrencisi iken bitirme projesi kapsamında yaptığı çalışmayı “American Institute of Physics (AIP) Conference proceeding” isimli yayınlanmış olan çalışmamız bu projeden esinlenerek yapılmıştır. Sevda Yeliz Nar şu anda bu sistem özerinde derin öğrenim metotları ile parçacık sınıflandırma metotlarını çalışmaktadır ve 3001 çerçevesinde oluşturulan analiz kodu kullanılmaktadır.

Öğrendiğimiz derin öğrenme teknik ve araçlarının, sadece belli hazır sayısal veriler dışında kendi tasarladığımız enstrümanlar üzerinden çalıştırmak, bize gelecekte yapabileceklerimiz açısında çok önemli önemli bir bilgi paylaşımı, deneyim ve özgüven vermektedir. Hali hazırda çeşitli ortamlarda gözlemlenen kozmik parçacık geçişlerini kapalı bir sistem olarak işlemekte ve gelen kozmik ışınları sınıflandırarak farklı ortamlarda (irtifa, sıcaklık, manyetik alan vb.) incelemekteyiz. Sonuçlarımızı farklı algoritmalar, sınıflandırmalar temelinde karşılaştırmalı olarak sürdürülmekte ve yayınlanması planlanmaktadır.

Endüstriyel uygulamalar kapsamında lisans öğrencilerimden biri ile yürüttüğüm iş-zekâsı problemlerinde derin öğrenim metotlarının kullanımı, gerçek şirket verileri ile uygulanarak **4.3** kısımda anlatıldığı üzere çalışılmıştır ve çalışılmaktadır. Bu konu başlığı özelinde bilgisayar mühendisliği ve endüstri mühendisliği bölümleri kapsamında yayınlanmak üzere makale çalışmaları yürütülmektedir. İlgili makalenin yazımı hala hazırda devam etmektedir. Bu aşamada öğrendiğimiz uygulamaları Hürriyet gazetesinden alınan veriler ile web analitiği üzerine uygulanarak veri üzerinden yapısal benzerlikler çıkartma analizleri yapılırken ve Newegg isimli elektronik e-ticaret platformunun satış verileri (Amerika Birleşik Devletleri) üzerinden zaman serilerine bağlı zamansal korelasyon verilerinin derin öğrenim metotları aracılığı ile uygulanması incelenmiş ve ilk çözümler önerilmiştir. Bu konu başlıkları ve yapılan çözümlemeler proje ile direk ilintili olmaması rağmen aynı yazılım alt-yapısını kullanmaktadır. Sonuç olarak akademi-sanayi iş-birliği kapsamında önemli bağlantıların kurulmasını sağlamıştır. Bu durum projenin en son iş paketi tüm diğer iş paketlerini de içerecek şekilde akademik alt-yapı ve uygulanabilir çalışmalar içerecek şekilde tamamlandığını göstermektedir.

Proje kapsamında verilen burslar aracılığı ile başladığımız projemiz çok önemli sonuçlar barındıran bir yapıda tamamlanmıştır ve bu çerçevede projenin yeni bir 1001 önerisini vermesi gerektiğini göstermektedir. Bu konuda çalışmalara başlanmış projenin devamı niteliğinde özellikle Top-Kuark çiftlerinin hassaslık ölçümlerini içerecek şekilde benzer bir yapıda verilmesi planlanmaktadır.

Yapılan projeler kapsamında İTÜ Teknoloji Transfer Ofisi (TTO), Hürriyet gazetesi, Intel ve Tarentum yapay zeka şirketleri ile ortak projeler yürütülmeye başlanmış ve bu çerçevede Tarentum A.Ş ve Intel tarafından proje yürütücüsüne İTÜ Fizik mühendisliğinde kurulmak üzere 60.000 dolarlık makine-teçhizat yardımı yapılmıştır. 3001 kapsamında yürütülen çalışmalar sponsorluk desteği alınarak makine ve derin öğrenim çalışmalarının yürütüldüğü "Data Science Laboratory – Veri Araştırmaları Laboratuvarı" kurulmasına sağlamıştır. Projenin devamında verilmesi planlanan 1001 projesi araştırma grubu, bu araştırma laboratuvarı kullanmaya devam edecek, önemli bilimsel ve akademi-sanayi iş birlikleri yürütülmesini sağlayacaktır. Bu nedenle projenin devamının desteklenmesi bilimsel başarının sürekliliği için oldukça önem arz etmektedir. Aşağıda proje dahilinde projeden bağımsız alınan destek kapsamında makine-teçhizat listesi verilmiştir:

Veri Bilimi ve Yapay Zekâ Araştırma Laboratuvarı

- Çalışma sunucusu ve ekipmanları: Intel Power Edge R620 Rock Server, Raid SSD disk hafıza sunucusu, Rock server kabini, HP ağ anahtarları. Veri seti eğitimleri iş koşturmaları, paralel programlama, GPU tabanlı resim yapısal inceleme ve ileri sayısal hesaplayıcılar
- 8 adet dizüstü Lenovo G50-i7 çalışma bilgisayarları
- D-link gözetleme (surveillance) kamerası: 1 adet ek profesyonel kamera cihazı, 3 adet Intel akıllı kızılötesi kamera
- 4 adet Intel Galileo geliştirme kartı ve IoT çalışma kitleri: Özellikle sensörlerin bir araya getirilmesi ve uygulama çalışmalarında kullanılmaktadır
- 1 Adet NVidia Jetson TX2 GPU Gömülü Geliştirme kiti: Yapay zeka uygulamaları kapsamında görüntü ve ses işleme, gözetleme kamerasında gelen görüntülerin işlenmesi ve öğrenilmesi, sensör çözümlemeleri ve büyük veri akışını sağlama modülleri çalışmaları yürütülmektedir.

Günümüzün en önemli konu başlıklarından biri olan makine öğrenimi uygulamalarının, e-ticaret ve web analitiği uygulamaları hakkında Hürriyet gazetesi ile birlikte "Makine Öğrenimi Teknikleri Aracılığı ile Kullanıcı Davranış Hareketlerini Analitik Ölçme ve Tahminleme Motorunun Geliştirilmesi" isimli Teydep 1503 projesi (No: 3170089) Şubat 2017'de gönderilmiş ve projem başarı ile kabul edilmesi, bu proje kapsamında yürütüğümüz ar-ge çalışmaların çerçevesinde başarılmıştır. Bu projede 9 numaralı iş paketi (en son) olarak belirtilen akademi-sanayi iş birliği Hürriyet ar-ge çalışmaları sağlanarak önemli bir başarıya imza atılmıştır.

Proje yürütücüsü ilgili başlıklar kapsamında yürüttüğü çalışmalardan dolayı Amerika Birleşik Devleti Chicago şehrinde yer alan Fermi Araştırma Laboravatuvarından (Fermilab) 6 aylık tam burslu misafir araştırmacı öğretim üyesi ödülü almıştır. Bu ödül 2017 yılı içersinden 8 aylık bir zamanı kapsamaktadır. Bu nedenle projenin ayrılan öğrencisi ile İlknur Köseoğlu (Almanya'da doktora kabul almıştır) ile birlikte tüm yaz aylarını kapsayacak şekilde çalışmalar yürütülmüştür. Alınan ödül ile bu proje direk ilinti olmamasına rağmen, bu proje kapsamında yapılan çalışmalarda burada yer alan araştırmalarda kullanılmış ve hala aktif olarak kullanılmaktadır.

# KAYNAKÇA

Abers, E.S. Lee, B.W. 1973. Gauge Theories Phys.Rept. 10.1016/0370-1573(73)90027-6 1-141 9.

Arkani-Hamed, Nima Dimopoulos, Savas Dvali, G.R. 1998. The Hierarchy problem and new dimensions at a millimeter Phys.Lett. 10.1016/S0370-2693(98)00466-3 263-272 B429.

Allanach, B.C. 2002, SOFTSUSY: a program for calculating supersymmetric spectra Comput.Phys.Commun. 10.1016/S0010-4655(01)00460-X 305-331 143.

Baer, Howard Paige, Frank E. Protopopescu, Serban D. Tata, Xerxes 1993. Simulating Supersymmetry with ISAJET.

Bayatian, G.L. et all. 2006. CMS Collaboration, CMS physics: Technical design report, CERN-LHCC-2006-001, CMS-TDR-008-1.

Bellenot, B., Ph Couet, O. Ganis, G. Mato, P. Moneta, L. Naumann, A. Piparo, D., 2015. ROOT 6 and beyond: TObject, C++14 and many cores. J.Phys.Conf.Ser. 10.1088/1742-6596/664/6/062006 062006 664 6.

Bellm J. et al., 2015. Herwig 7.0 / Herwig++ 3.0 Rel. Note, CERN-PH-TH-2015.

C. M. Bishop, 2006. Pattern Recognition and Machine Learning. Springer. ISBN 0-387-31073-8

Cakir, A. et. all CMS SUSY Public page https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS

Chatrchyan S. and et. all, 2012. Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s}$ = 8 TeV Eur.Phys.J. 10.1140/epjc/s10052-013-2677-2 2677 C73 12 2013 2013-12-2.

Dobbs, M. Jorgen B., 2001. The HepMC C++ Monte Carlo event record for High Energy Physics Comput.Phys.Commun. 10.1016/S0010-4655(00)00189-2 41-46 134.

Djouadi, A. Muhlleitner, M.M. Spira, M., 2007. Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface) Acta Phys.Polon. 635-644 B38.

F.A. Berends, H. Kuijf, 1991. Jets at the LHC, Volume 353, Issue 1, Pages 59-86.

G. F. Giudice, 2010. A Zeptospace Odyssey: A Journey into the Physics of the LHC, Oxford University Press, Oxford, ISBN 978-0-19-958191-7

Gleisberg, Tanju Hoeche, Stefan Krauss, Frank Schalicke, Andreas Schumann, Steffen Winter, Jan-Christopher, 2004. SHERPA 1. alpha: A Proof of concept version JHEP 10.1088/1126-6708/2004/02/056 056 0402.

Frixione S., Webber B., 2003. The MC@NLO 2.0 event generator, CERN-TH-2003-159

Frixione, S. Nason, P., 2007. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method JHEP 10.1088/1126-6708/2007/11/070 070 0711.

Hocker, A. et al., 2007. TMVA- Toolkit for Multivariate Data Analysis PoS-ACAT (2007) 040 physics/0703039 [PHYSICS] CERN-OPEN-2007-007.

Khachatryan, Serguei and et all., 2012. CMS Coll., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716.

Maltoni, Fabio Stelzer, 2003. Tim MadEvent: Automatic event generation with MadGraph JHEP 10.1088/1126-6708/2003/02/027 027 0302.

Martin, Stephen P., 2010. A Supersymmetry primer Adv.Ser.Direct.High Energy Phys. 10.1142/9789812839657_0001 1-153 21.

Porod, W., 2003. SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e+ e- colliders Comput.Phys.Commun. 10.1016/S0010-4655(03)00222-4 275-315 153.

Rizzo, Thomas G., 2001. Probes of universal extra dimensions at colliders Phys.Rev. 10.1103/PhysRevD.64.095010 095010 D64.

Selvaggi, M., 2014. DELPHES 3: A modular framework for fast-simulation of generic collider experiments J.Phys.Conf.Ser. 10.1088/1742-6596/523/1/012033 012033 523.

Sjostrand, Torbjorn Mrenna, Stephen Skands, Peter Z., 2008. A Brief Introduction to PYTHIA 8.1 Comput.Phys.Commun. 10.1016/j.cpc.2008.01.036 852-867 178

Shiltsev, V., 2012. High energy particle colliders: past 20 years, next 20 years and beyond, Physics-Uspekhi 55.10 (2012) 965

Y. Gershtein et al., 2013. FERMILAB-CONF-13-584-T.

Wenzel, H. Yarba, J. Dotti, A., 2015. The Geant4 Physics Validation Repository J.Phys.Conf.Ser. 10.1088/1742-6596/664/6/062066 062066 664 6.

Wess, J. Zumino, B., 1974. A Lagrangian Model Invariant Under Supergauge Transformations Phys.Lett. 10.1016/0370-2693(74)90578-4 52 B49.

# TÜBİTAK
# PROJE ÖZET BİLGİ FORMU

| | |
|---|---|
| Proje Yürütücüsü: | Doç. Dr. MUAMMER ALTAN ÇAKIR |
| Proje No: | 116F026 |
| Proje Başlığı: | Büyük Hadron Çarpıştırıcısındaki İleri Veri Çözümleme Teknikleri İçin Derin Öğrenim Metotları |
| Proje Türü: | 3001 - Başlangıç AR-GE |
| Proje Süresi: | 24 |
| Araştırmacılar: | |
| Danışmanlar: | |
| Projenin Yürütüldüğü Kuruluş ve Adresi: | İSTANBUL TEKNİK Ü. FEN-EDEBİYAT F. FİZİK MÜHENDİSLİĞİ B. |
| Projenin Başlangıç ve Bitiş Tarihleri: | 15/05/2016 - 15/05/2018 |
| Onaylanan Bütçe: | 90000.0 |
| Harcanan Bütçe: | 48777.33 |
| Öz: | Proje süresince proje ekibi aşağıda yer alan yayın, sözlü sunum, bildiri, çalışta ve araştırmalara katılım gibi çalışmaların yanı sıra, araştırmalarda kullanılmak üzere TUBİTAK veya herhangi bir devlet kurumunun desteği dışında yapılandırılmış bir araştırma laboratuvarı ve proje kapsamında alınmış ödül ve burslar bulunmaktadır. 3001 başlangıç ar-ge desteği, bilimsel çalışmalar ile birlikte yürütülen akademi-sanayi iş-birliğini de içine alacak şekilde başarı ile tamamlanmıştır. |
| Anahtar Kelimeler: | LHC, SUSY, MC Generators, TMVA, Ntupler Analyzer, TMVA, Deep Learning, Machine Learning |
| Fikri Ürün Bildirim Formu Sunuldu Mu?: | Hayır |
| Projeden Yapılan Yayınlar: | 1- Peltier Based Cloud Chamber (Bildiri - Uluslararası Bildiri - Poster Sunum),<br>2- Fermilab LPC Visitor Scientist Fellowship (Ödül - Uluslararası Ödül - Üniversite, Kurum veya Kuruluşların Verdiği Ödüller),<br>3- Searches for Supersymmetry with Deep Learning (Tez (Araştırmacı Yetiştirilmesi) - Yüksek Lisans Tezi),<br>4- Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at 13 TeV (Makale - İndeksli Makale), |

TÜBİTAK