# Parkinson Diseases Detection

Parkinson hastalığı, öncelikle motor sistemi etkileyen nörodejeneratif bir hastalıktır. Beynin substantia nigra adı verilen belirli bir bölgesindeki dopamin üreten hücrelerin ilerleyici kaybı ile karakterizedir. Dopamin, hareketi koordine etmede çok önemli bir rol oynayan bir nörotransmitterdir ve eksikliği Parkinson hastalığı ile ilişkili motor semptomlara yol açar.

Parkinson hastalığının kesin nedeni henüz tam olarak anlaşılamamıştır. Bununla birlikte, hem genetik hem de çevresel faktörlerin gelişimine katkıda bulunduğuna inanılmaktadır. Vakaların çoğu, bilinen bir neden olmadan meydana gelen sporadik olarak kabul edilirken, vakaların küçük bir yüzdesi spesifik genetik mutasyonlarla ilişkilidir.

Parkinson hastalığının birincil semptomları arasında titreme (istemsiz sallama), sertlik (kasların sertliği), bradikinezi (hareket yavaşlığı) ve postüral dengesizlik yer alır. Bu semptomlar genellikle kademeli olarak başlar ve zamanla kötüleşir. Diğer yaygın semptomlar arasında denge ve koordinasyon güçlüğü, hareketlerde donma, konuşma ve yazmada değişiklikler ve depresyon, anksiyete, uyku bozuklukları ve bilişsel bozukluk gibi motor olmayan semptomlar yer alabilir.

Parkinson hastalığı tipik olarak bir kişinin tıbbi geçmişine, klinik muayenesine ve semptomların gözlemlenmesine dayanarak teşhis edilir. Şu anda Parkinson hastalığının tedavisi yoktur, ancak semptomlarını yönetmek ve bu durumdaki bireylerin yaşam kalitesini iyileştirmek için tedavi seçenekleri mevcuttur. Levodopa gibi beyindeki dopamin seviyelerini artıran ilaçlar genellikle reçete edilir. Diğer yaklaşımlar arasında fizik tedavi, mesleki terapi, konuşma terapisi ve bazı durumlarda derin beyin stimülasyon cerrahisi yer alır.

Parkinson hastalığının altında yatan mekanizmaları daha iyi anlamak ve yeni tedavi stratejileri geliştirmek için araştırma çalışmaları devam etmektedir. Bu çalışmalar aşağıdaki gibi kısaca açıklanabilir.

Makine öğrenimi teknikleri, Parkinson hastalığının saptanmasına ve teşhis edilmesine yardımcı olmak için kullanılabilir. Araştırmacılar, erken teşhis ve doğru teşhis için tıbbi verileri analiz etmek ve yorumlamak amacıyla makine öğrenimi algoritmalarını kullanarak çeşitli yaklaşımlar keşfettiler. İşte birkaç örnek:

Tahmine dayalı modelleme: Makine öğrenimi algoritmaları, klinik veriler, genetik bilgiler ve görüntüleme verilerinin bir kombinasyonunu içeren veri kümeleri kullanılarak eğitilebilir. Bu algoritmalar, Parkinson hastalığına yakalanma riski taşıyan bireylerin belirlenmesine yardımcı olan tahmine dayalı modeller geliştirmek için verilerdeki kalıpları ve ilişkileri öğrenebilir.

Özellik çıkarımı: Makine öğrenimi, beyin taramaları veya işlevsel görüntüleme verileri gibi tıbbi görüntülerden ilgili özelliklerin çıkarılmasına yardımcı olarak Parkinson hastalığıyla ilişkili ince değişiklikleri belirlemeye yardımcı olabilir. Bu özellikler daha sonra, sağlıklı bireyler ile hastalığı olanlar arasında ayrım yapmak için sınıflandırma algoritmaları için girdi olarak kullanılabilir.

Destek vektör makineleri (SVM): SVM, Parkinson hastalığı tespitinde kullanılan popüler bir makine öğrenme tekniğidir. Belirli kriterlere göre bireyleri Parkinson hastaları veya sağlıklı kontroller olarak sınıflandırmak için motor semptom değerlendirmeleri gibi klinik veriler üzerinde eğitilebilir.

Derin öğrenme: Makine öğreniminin bir alt kümesi olan derin öğrenme, karmaşık veri kümelerinden karmaşık kalıpları otomatik olarak öğrenmek ve çıkarmak için çok katmanlı sinir ağlarını eğitmeyi içerir. Derin öğrenme modelleri, Parkinson hastalığı teşhisi ve ilerlemenin izlenmesi için tıbbi görüntülerin analizinde umut vaat ediyor.

Ses ve konuşma analizi: Parkinson hastalığı konuşma kalıplarını etkileyebilir ve hastalıkla ilişkili karakteristik değişiklikleri belirlemek için ses kayıtlarını veya konuşma örneklerini analiz etmek için makine öğrenimi algoritmaları kullanılabilir.

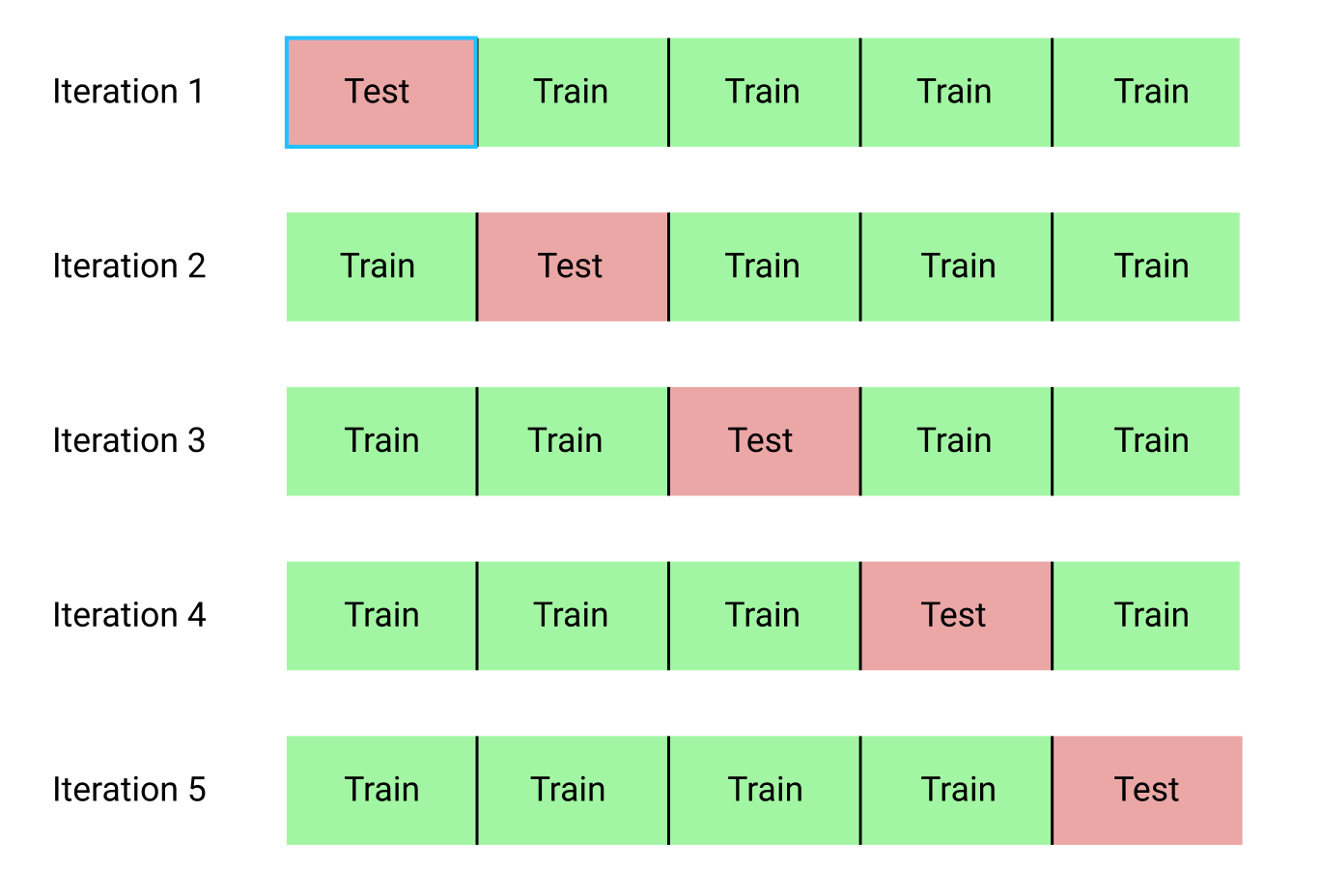
Parkinson hastalığının altında yatan mekanizmaları daha iyi anlamak için ve yeni tedavi stratejileri geliştirmek için makin öğrenmesi modelleri denerek bu modellerin kıyaslanması amaçlanmaktadır. Birden fazla eğitilen modelleri kullanarak en iyi sonuçları elde etmek için k-fold, Modellerin Birleştirilmesi (ensemble learning) yöntemleri kullanılması amaçlanmaktadır. Bu doğrultuda denenmesi planlanan makine öğrenmesi modelleri: Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), XG Boost, Gradient Boost, Logistic Regression.

K-Fold yöntemi:

Makine öğreniminde k-katlı çapraz doğrulamanın ana fikri, sınırlı veriler kullanarak bir modelin performansını ve genelleme yeteneğini değerlendirmektir. Mevcut verilerin yaklaşık olarak eşit boyutta k alt kümeye veya "bölümlere" bölünmesiyle meydana gelir. Model daha sonra k kez eğitilir ve değerlendirilir, her seferinde doğrulama seti olarak farklı bir kat ve eğitim seti olarak kalan katlar kullanılır. Fotoğraf 1 deki gibi gösterilir.

Cross validation bizim eğitmiş olduğumuz modelin aşırı öğrenmesine (overfitting) sahip olup olmadığına ve modelimin kalitesi hakkında fikir sağlayacaktır. Böylece yeni test verilerinde hata oranlarını önceden tahmine etmeye ve modelin performansının stabilitesini öngörebilmeyi sağlayacaktır. Kolay uygulanabilirliği sayesinde sıklıkla kullanılan bir yöntemdir.

Genellikle az veri setine sahip datalarda tam anlamıyla dataların verimli şekilde kullanılmasını sağlamaktadır.



Fotoğraf\_1: K-fold

Modellerin Birleştirilmesi (Ensemble Learning):

Modellerin birleştirilmesi yöntemini bölüm hocam olan Görkem Serbest hocamın derste verdiği örnek ile kısaca bir doktor tarafından kanser teşhisi konulan bir hastanın teşhisin kesinliği için istatistik olarak en az üç doktordan ikisinin kanser teşhisi koyması gerekmektedir. Böylece bir den fazla tek sayı da modelin tercih edilmesiyle birlikte elde edilen sonuçlar daha etkili olmaktadır. Bu yöntemde iki farklı yaklaşım vardır bunlar Hard Voting ve Soft Voting olarak adlandırılır. Hard Voting yaklaşımında makine öğrenimi modellerinden elde edilen sonuçların çoğunluğu ne derse sonucun da o doğrultuda olmasıdır. Kısaca modellerden çıkan sonuçların mod alınması olarak açıklayabiliriz. Soft Voting yaklaşımında makine öğrenimi modellerinde sınıfların çıktı ağırlıklarının model sayısı kadar eşit oranda toplanarak elde edilerek en çok ağırlığa sahip olan sınıfın çıktı olarak verilmesidir.

Data Set:

Bu veri seti, 23'ü Parkinson hastalığı (PH) olan 31 kişiden alınan çeşitli biyomedikal ses ölçümlerinden oluşmaktadır. Tablodaki her sütun belirli bir ses ölçüsüdür ve her satır bu kişilere ait 195 ses kaydından birine karşılık gelir ("isim" sütunu). Verilerin temel amacı, sağlıklı için 0 ve PH için 1 olarak ayarlanan "durum" sütununa göre sağlıklı insanları PH' liler den ayırmaktır.

Matris sütun girişleri (öznitelikleri):

isim - ASCII konu adı ve kayıt numarası

MDVP:Fo(Hz) - Ortalama vokal temel frekansı

MDVP:Fhi(Hz) - Maksimum vokal temel frekansı

MDVP:Flo(Hz) - Minimum vokal temel frekansı

MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP - Temel frekansta çeşitli varyasyon ölçümleri

MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Genlikte çeşitli varyasyon ölçüleri

NHR, HNR - Gürültünün sesteki tonal bileşenlere oranının iki ölçüsü

durum - Deneğin sağlık durumu (bir) - Parkinson, (sıfır) - sağlıklı

RPDE, D2 - İki doğrusal olmayan dinamik karmaşıklık ölçüsü

DFA - Sinyal fraktal ölçekleme üssü

spread1,spread2,PPE - Temel frekans değişiminin doğrusal olmayan ölçüsü

Veriler ASCII CSV biçimindedir. CSV dosyasının satırları, bir ses kaydına karşılık gelen bir örnek içerir. Ve bu dataset hakkında daha fazla bilgi için buradan yararlanılabilir.

(Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Parkinson hastalığının telemonitoringi için disfoni ölçümlerinin uygunluğu', IEEE Transactions on Biomedical Engineering)