

Parkinson Hastalığı Tespiti

1) Özet

Parkinson hastalığı, beyindeki dopamin seviyesindeki azalma ile ilişkili, motor beceri bozukluklarına ve el yazısındaki değişikliklere yol açabilen nörolojik bir bozukluktur. Bu hastalığın erken teşhisinde, hastaların yazılı veya çizili aktiviteleri üzerinde yapılan analizler önemli bir yer tutmaktadır. Bu çalışmada, Parkinson hastalığının teşhisinde kullanılmak üzere bir el çizimi veri seti kullanılmıştır [1]. Veri seti, 102 spiral ve 102 dalga çiziminden oluşmaktadır ve hastalığın motor beceriler üzerindeki etkileri değerlendirilmektedir [2]. Başlangıçta, derin öğrenme tabanlı bir CNN modeli ile bu veriler üzerinde bir sınıflandırma yapılmış, ancak doğruluk oranının %51.22 ile düşük kalması nedeniyle daha geleneksel makine öğrenmesi algoritmalarına yönelinmiştir. Sonuç olarak, Random Forest ve XGBoost algoritmaları kullanılarak, spiral çizimleri için %76.67 ve dalga çizimleri için %80.00 doğruluk elde edilmiştir.

Anahtar Kelimeler: Derin öğrenme (DL); El çizimi; Makine öğrenmesi (ML); Parkinson hastalığı; hibrit modelle.

2) Giriş

Parkinson hastalığı, dopamin üreten hücrelerin kaybı nedeniyle beyindeki motor kontrol merkezlerinin işlevlerini kaybetmesiyle karakterize bir nörolojik hastalıktır. Erken teşhis, hastalığın seyrini yavaşlatmak için oldukça önemlidir ve erken dönemdeki motor beceri bozukluklarını tanıyabilmek, tedavi sürecinin etkinliğini artırabilir. Genellikle, Parkinson hastalığının teşhisi klinik semptomlar ve nörolojik muayeneye dayalı olarak yapılmaktadır. Ancak, bu yöntemler genellikle subjektif olup zaman alıcı olabilir. Bu nedenle, Parkinson hastalığının tespiti için alternatif ve non-invaziv yöntemler geliştirilmesi önemlidir [1].

El yazısı ve çizim analizleri, Parkinson hastalığının tespiti için kullanılabilecek güçlü araçlardır. Çünkü Parkinson hastalığı, bireylerin motor becerilerini ve koordinasyonunu etkileyerek yazı yazma ve çizim yapma yetilerini bozar. Bu çalışma, Parkinson hastalığını tespit etmek için bir el çizimi veri seti kullanmaktadır. Veri setinde yer alan spiral ve dalga çizimleri, hastalığın motor beceriler üzerindeki etkilerini değerlendirmeye yönelik olarak toplanmıştır [3].

Makine öğrenmesi algoritmaları, büyük veri setleri üzerinde desen tanıma ve sınıflandırma işlemleri gerçekleştirebilir. Bu çalışmada, Random Forest ve XGBoost algoritmaları, Parkinson hastalığının tespiti için iki ana model olarak seçilmiş ve her iki modelin doğruluk oranları karşılaştırılmıştır.

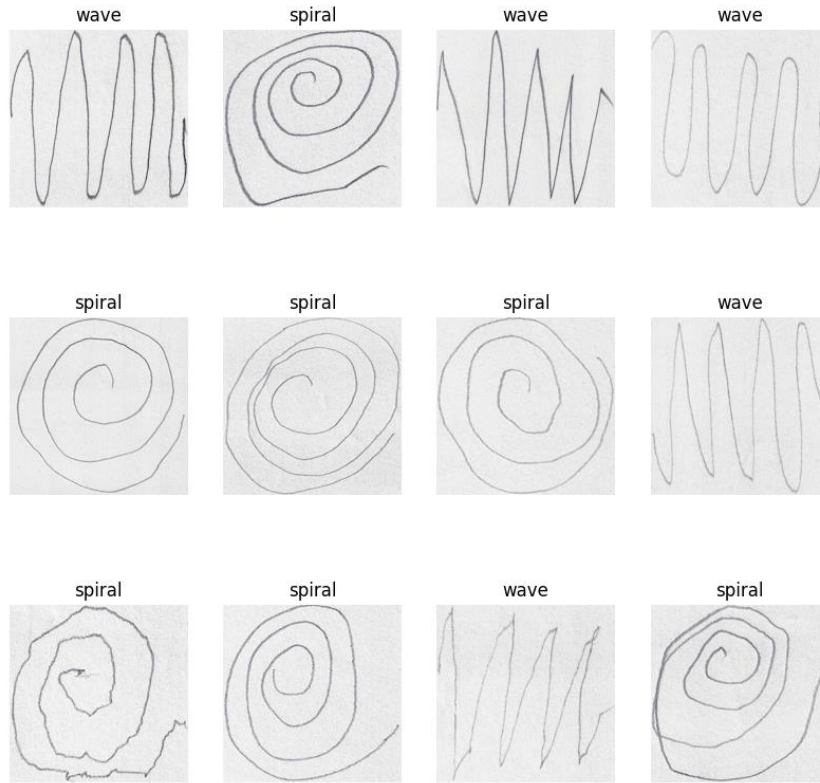
3) Metotlar ve Yöntemler

3.1 Veri Seti

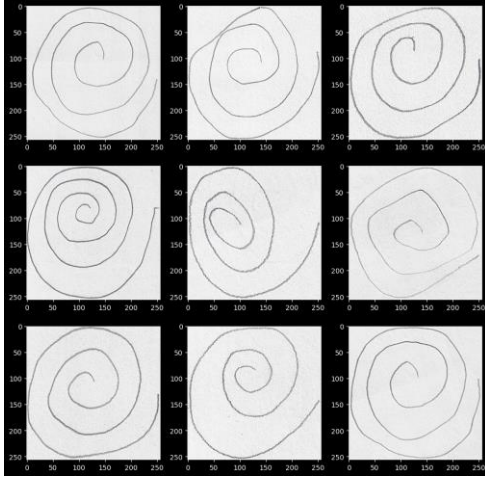
Bu çalışma, Parkinson hastalığının tespiti için, Parkinson hastaları ve sağlıklı bireyler tarafından yapılan spiral ve dalga çizimlerini içeren bir veri setini kullanmaktadır. Veri seti, Kaggle platformunda "Images of Healthy and Patients with Parkinson's Drawing Spirals and Waves" başlığıyla yayımlanan ve Mader (2017) tarafından derlenen bir kaynağa dayanmaktadır [2].

Veri seti, toplamda 204 el çizimi örneği içermektedir: 102 spiral çizimi ve 102 dalga çizimi. Bu çizimler, Parkinson hastalığının motor becerilerindeki bozulmaları belirlemek amacıyla toplanmıştır. Çizimler, spiral ve dalga olmak üzere iki ana kategoriye ayrılmıştır [2].

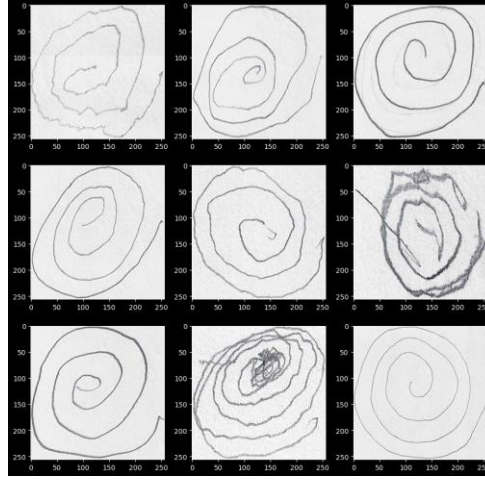
Her iki kategori de, sağlıklı bireyler ve Parkinson hastaları tarafından çizilen örnekleri içermektedir. Veri seti, her kategori için eğitim ve test aşamalarına ayrılacak şekilde düzenlenmiştir. Bu yapı, farklı birey grupları arasında karşılaştırmalar yaparak Parkinson hastalığının motor becerilerindeki etkileri değerlendirmeyi amaçlamaktadır [2].



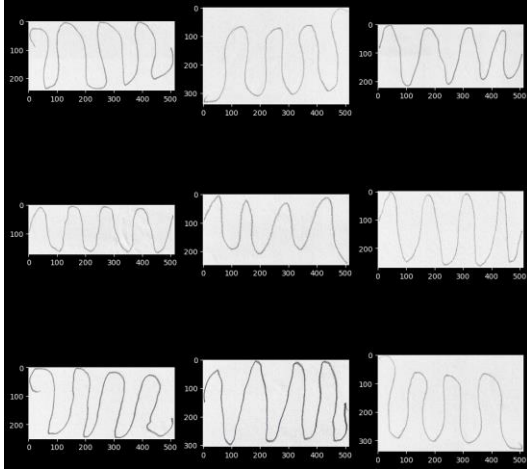
Şekil 1. Veri Setinin Sınıfları



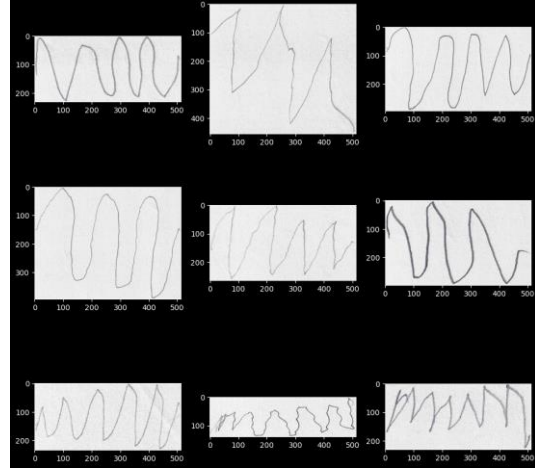
Şekil 2. Sağlıklı Spiral



Şekil 3. Parkinson Spiral



Şekil 4. Sağlıklı Dalga



Şekil 5. Parkinson Dalga

3.2 Özellik Çıkartımı

Görüntülerden anlamlı özelliklerin çıkarılması için Histogram of Oriented Gradients (HOG) yöntemi kullanılmıştır. HOG, görüntülerdeki kenarları ve şekilleri tanımlamak için etkili bir tekniktir. Bu özellikler, modelin öğrenmesini sağlayacak şekilde sınıflandırma algoritmalarına verilmiştir. HOG, görüntüdeki kenarların yönelimlerine göre histogramlar oluşturur ve bu histogramlar, modelin görüntülerin desenlerini anlamasına yardımcı olur.

3.3 Veri Ön İşleme

Veri ön işleme aşamasında, görüntüler gri tonlamaya dönüştürülmüş, boyutları 200x200 piksele küçültülmüş ve ardından eşikleme (thresholding) işlemi uygulanmıştır. Eşikleme,

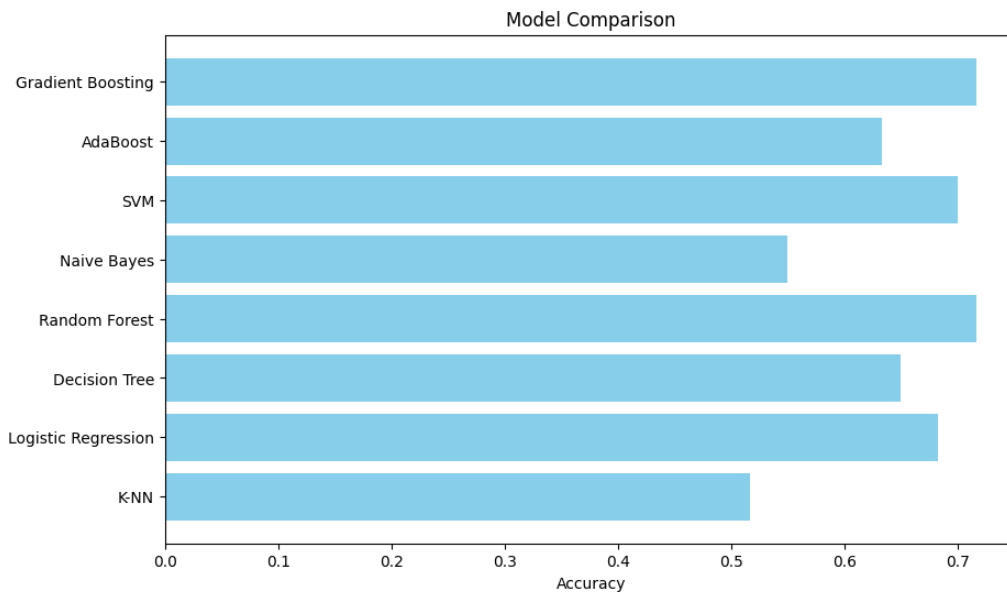
görüntüdeki arka planı ve çizimi net bir şekilde ayırt edebilmek amacıyla yapılmıştır. Bu adımlar, modelin doğru şekilde eğitilmesini sağlamak amacıyla yapılmıştır.

3.4 Model Seçimi ve Eğitim

Parkinson hastalığının tespiti için bu projede ilk olarak derin öğrenme tabanlı bir **Convolutional Neural Network (CNN)** modeli kullanılmıştır. Ancak, CNN modelinin doğruluk oranı beklenenin çok altında kalmıştır. Bu düşük doğruluk nedeniyle, daha geleneksel makine öğrenmesi algoritmalarına yönelilmiştir. Bu süreçte, farklı makine öğrenmesi algoritmaları denenmiş ve en yüksek başarıyı gösteren iki algoritma, **Random Forest** ve **XGBoost**, model seçiminde tercih edilmiştir.

- **Random Forest:** Birden fazla karar ağacından oluşan bir ensemble öğrenme yöntemidir. Her bir karar ağacı, farklı alt küme verisiyle eğitilir ve bu ağaçların sonuçları birleştirilerek nihai tahmin yapılır. Random Forest, yüksek doğruluk oranları ve düşük overfitting (aşırı uyum) riski ile tanınır. Bu yöntem, özellikle veri setlerinde çeşitlilik ve çokluk olduğunda etkili sonuçlar üretir [4].
- **XGBoost:** Gelişmiş bir gradient boosting yöntemidir ve büyük veri setlerinde oldukça başarılıdır. XGBoost, her iterasyonda hataları minimize ederek yeni modelin daha doğru hale gelmesini sağlar. Bu yöntem, düşük hata oranları ve hızlı eğitim süreleri ile özellikle büyük ve karmaşık veri setleri üzerinde güçlü performans gösterir [5].

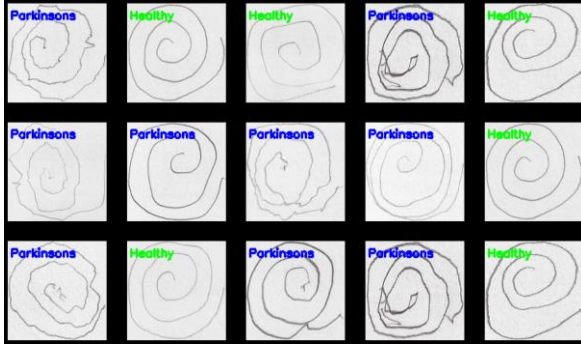
Eğitim verileri her iki modelle de eğitilmiş ve doğruluk oranları hesaplanmıştır. Modelin eğitilmesi sırasında, hiperparametreler de optimize edilmiştir. Başlangıçta farklı makine öğrenmesi algoritmaları denenmiş ve en yüksek doğruluk oranları **Random Forest** ve **XGBoost** algoritmalarında elde edilmiştir. Bu nedenle, Parkinson hastalığı tespiti için bu iki algoritma seçilmiştir.



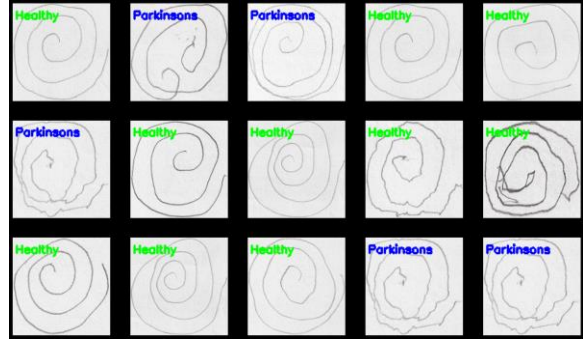
Şekil 6. Model Karşılaştırması

3.5 Model Eğitimi ve Değerlendirme

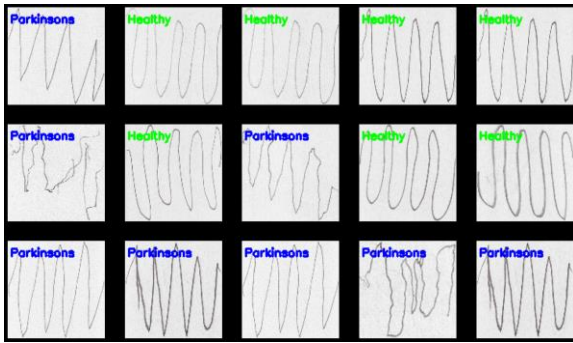
Veri seti, model eğitimi için eğitim ve test olarak ikiye ayrılmıştır. Eğitim verisi ile model eğitildikten sonra, test verisi üzerinde her iki algoritmanın performansları değerlendirilmiştir.



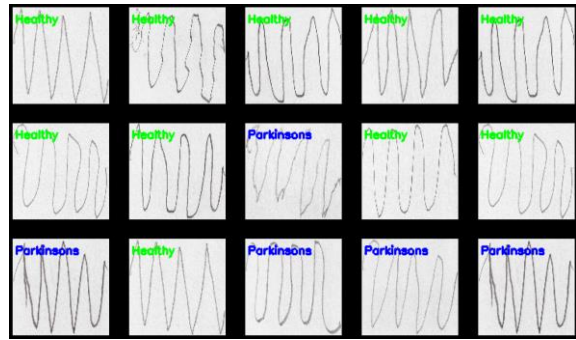
Şekil 7. Random Forest Kullanarak Test Etme 1



Şekil 8. XGB Kullanarak Test Etme 1



Şekil 9. Random Forest Kullanarak Test Etme 2



Şekil 10. XGB Kullanarak Test Etme 2

3.6 Masaüstü Uygulaması Geliştirme:

Şekil 11’de görüldüğü üzere, Tkinter çerçevesi kullanılarak bir masaüstü uygulaması oluşturulmuştur. Kullanıcılar, uygulama üzerinden el çizimi görüntülerini yükleyerek modelin Parkinson hastalığına dair tahmin sonuçlarını görebilirler. Uygulama, kullanıcıya seçim yapabileceği model seçenekleri sunarak, yüklenen görüntünün Parkinson hastalığına dair tahminini ve güven skorunu sağlar.



Şekil 11. Proje Arayüzü

4) Sonuç

Eğitim sürecinde spiral ve dalga çizimleri için modeller eğitilmiş ve doğruluk oranları hesaplanmıştır. Elde edilen sonuçlar aşağıda belirtilmiştir:

- **Spiral Çizimleri için Doğruluk:**
 - Random Forest: %76.67
 - XGBoost: %73.33
- **Dalga Çizimleri için Doğruluk:**
 - Random Forest: %80.00
 - XGBoost: %73.33

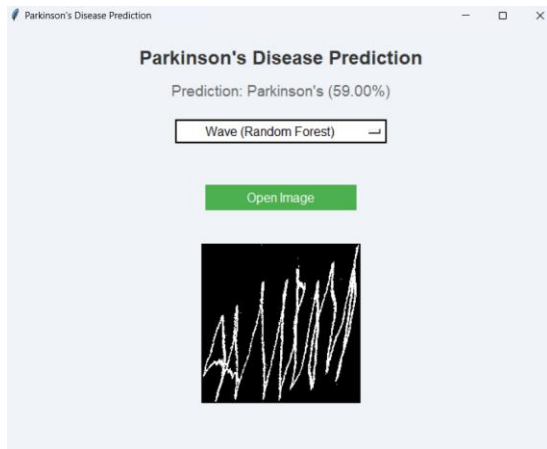
Bu doğruluk oranları, her iki modelin de genel olarak başarılı olduğunu göstermektedir. Ancak, Random Forest modeli her iki veri setinde de daha yüksek doğruluk oranlarına ulaşmıştır.



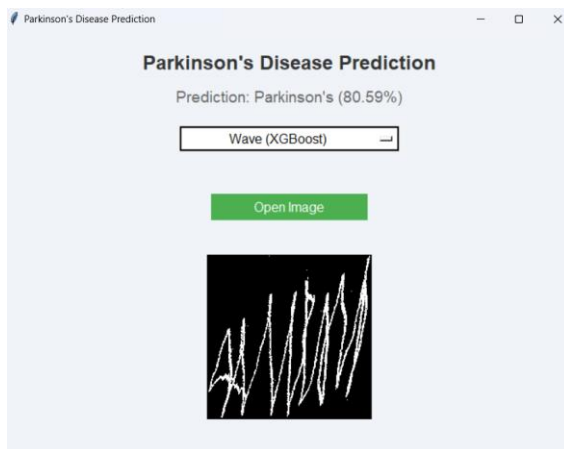
Şekil 12. Proje İncelemesi 1



Şekil 13. Proje İncelemesi 2



Şekil 14. Proje İncelemesi 3



Şekil 15. Proje İncelemesi 4

5) Bulgular ve Tartışma

Bu çalışmada, Parkinson hastalığının teşhisi için çeşitli makine öğrenmesi algoritmalarının etkinliği incelenmiştir. Başlangıçta derin öğrenme tabanlı bir model, özellikle **Convolutional Neural Network (CNN)**, kullanılarak el çizimi veri seti üzerinde sınıflandırma yapılmaya çalışılmıştır. Ancak, CNN modelinin doğruluk oranı %51,22 ile beklenenin oldukça altında kalmıştır. Bu düşük performans, veri setinin küçük boyutu, görüntülerin sınırlı çeşitliliği ve modelin gereksinim duyduğu daha büyük veri setine olan ihtiyaç gibi faktörlerle ilişkilendirilebilir. Bu durum, derin öğrenme modellerinin genellikle büyük veri setlerine ihtiyaç duyduğunu ve küçük veri setlerinde daha sınırlı performans gösterdiğini ortaya koymuştur [6].

Derin öğrenme modelinden elde edilen düşük doğruluk oranı, daha geleneksel makine öğrenmesi algoritmalarına yönelmek için bir gerekçe oluşturmıştır. **Random Forest** ve **XGBoost** gibi güçlü makine öğrenmesi algoritmaları kullanılarak yapılan ikinci model eğitimi, daha yüksek doğruluk oranları sağlamıştır. Spiral çizimleri için **Random Forest** modelinin doğruluk oranı %76,67, **XGBoost** modelinin ise %73,33 olmuştur. Wave çizimleri içinse **Random Forest** modelinin doğruluk oranı %80,00 olarak elde edilmiştir. Bu sonuçlar, küçük veri setlerinde **Random Forest** ve **XGBoost** gibi makine öğrenmesi yöntemlerinin derin öğrenmeye kıyasla daha etkili olabileceğini göstermektedir. Ayrıca, bu bulgular, klasik makine öğrenmesi yöntemlerinin veri setinin boyutundan bağımsız olarak tutarlı performans gösterdiğini ortaya koymaktadır [5].

Başka bir makalede, küçük veri setlerinin performansını artırmak amacıyla veri artırma (augmentation) tekniği kullanıldığı ve bu yöntemle elde edilen sonuçların daha başarılı olduğu gözlemlenmiştir. **RESNET-50** ve **SVM** birleşim modelinin doğruluk oranı %98,45, duyarlılık skoru 0,99 ve özgüllük skoru 0,98 olarak belirlenmiştir. Bu çalışma, derin öğrenme ve makine öğrenmesi tekniklerinin birleşiminden daha güçlü sonuçlar elde edilebileceğini göstermektedir. **RESNET-50**'nin özellikle derin öğrenme modelinin özellik çıkarımı için etkinliği vurgulanmıştır [1].

Bu karşılaştırma, bizim çalışmamızla elde edilen sonuçlarla benzerlikler ve farklılıklar göstermektedir. Özellikle, derin öğrenme tekniklerinin uygulanmasının doğruluk oranlarını artırabileceği gözlemlenmiştir. Bununla birlikte, mevcut çalışmamızda kullanılan **Random Forest** ve **XGBoost** modellerinin, küçük veri setleri ile de iyi sonuçlar verdiği ancak daha geniş veri setleri ve daha gelişmiş derin öğrenme teknikleri ile doğruluk oranlarının daha da iyileştirilebileceği düşünülmektedir. Ayrıca, hibrit modellerin ve derin öğrenme algoritmalarının, özellikle özellik çıkarımı ve sınıflandırma aşamalarında birlikte kullanılması, gelecekte daha iyi sonuçlar elde edilmesine olanak sağlayabilir.

6) Kaynaklar

1. Varalakshmi, P., Priya, B. T., Rithiga, B. A., Bhuvaneaswari, R., & Jaya Sundar, R. S. (2022). Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models. *Parkinsonism & Related Disorders*, 105, 24-31. <https://doi.org/10.1016/j.parkreldis.2022.10.020>
2. **Kaggle Parkinson's Drawings Dataset:**
<https://www.kaggle.com/datasets/kmader/parkinsons-drawings>
3. Thomas, M., Lenka, A., & Pal, P. K. (2017). Handwriting analysis in Parkinson's disease: Current status and future directions. *Movement Disorders Clinical Practice*, 4(6), 806-818. <https://doi.org/10.1002/mdc3.12552>
4. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).
6. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.