

The background is a dark blue gradient with a subtle pattern of small white dots. Overlaid on this are several concentric circles and arcs in a lighter blue color. Some of these arcs have degree markings, such as 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, and 260. There are also some dashed lines and arrows, suggesting a technical or scientific theme.

SAYISAL İŞARET İŞLEME FİNAL PROJESİ SUNUMU

ÖMER FARUK ORUÇ - SENTETİK SES VE DOĞAL SES AYRIMI

1-PROJENİN AMACI

- Teknolojideki gelişmeyle birlikte güvenlik ihtiyacı bulunan uygulamalarda kişisel erişimi sağlayabilmek amacıyla parmak izi, yüz, ses gibi biyometrik verilerin kullanımı gün geçtikçe yaygınlaşmaktadır. Bu biyometrik verilerden ses, yani konuşma sinyalinin hem kişiden kolaylıkla elde edilebilir olması hem de yüksek mobilite sağlaması otomatik konuşmacı doğrulama (ASV) sistemlerini popüler hale getirmektedir [1].
- Dünya genelinde yapay zeka tabanlı ses işleme uygulamalarının artarak büyük bir alan oluşturması ciddi güvenlik zaafiyetlerini de beraberinde getirmiştir. Özellikle konuşmacının temel ses özelliklerinden sentetik ses ya da daha kaliteli ses üretilebilmektedir [2].
- Sentetik ses oluşturma yöntemlerinin gelişmesi ile beraber artık doğal ses ile sentetik ses neredeyse ayırt edilemez hale gelmiştir. Projede amaç olarak ise son teknolojik yöntemler ile geliştirilen sentetik seslerin doğal hallerinden ayırt edilmesi amaç edinilmiştir. Bunun için özgün yöntemlerden yararlanılmıştır.

2- PROJENİN KULLANIM ALANLARI

- Sesin hemen hemen tüm alanlarda kullanımı yaygın olduğu için kullanım amacı oldukça yaygındır. Bazı önemli kullanım alanları şu şekildedir:
- **Sinyal İstihbaratı (SIGINT):** Sahtecilik ve manipölasyon tespiti, gizli iletişim ve kodlamaların çözülmesi, kimlik doğrulama ve yetkilendirme gibi çalışmaları ile kullanılır. Güvenlik ve savunma sektörü için de aynı veya benzer uygulama alanlarını kullanırlar.
- **Telekomünikasyon:** Telefon görüşmeleri ve sesli mesajlar üzerinde sahte ses algılama, telefon dolandırıcılığını önlemek ve iletişim güvenliğini artırmak için kullanılır.
- **Medya Ve Eğlence:** Podcast'ler, radio ve televizyon programları, müzik ve diğer medya içeriklerinde gerçek ve sentetik sesi ayırt etmek, içerik üreticileri ve dinleyiciler için kullanılır.
- **Sesli Asistanlar Ve Yapay Zeka:** Sesli asistanlar veya yapay zeka tabanlı sesli sistemler, gerçek bir kullanıcının komutlarını sentetik seslerden ayırt edebilmesi için güzel bir kullanım alanı oluşturmaktadır; fakat, kullanımı geliştirilme ve üzerinde proje yapılmaya oldukça açıktır.

3-LİTERATÜR TARAMASI

- Sahte konuşma tespiti için daha etkili ve verimli yöntemler geliştirmektir[3]. Bu çabalar, daha iyi performans gösteren parameter etkin evrişimli mimarilerin geliştirilmesi ve transfer öğrenme tekniklerinin kullanılması gibi konuları içermektedir. Sahte konuşma tespiti için EfficientCNN ve RES-EfficientCNN adlı dört parametre-etkin evrişimli mimariler kullanılmış ve bu yöntemler, sahte konuşma tespitinde yüksek F1 puanları elde etmiştir. Ayrıca, transfer öğrenme teknikleri kullanılarak yeni saldırı vektörlerine hızlı bir şekilde adapte olunabileceği ve düşük miktarda veri kullanarak yüksek tespit performansı elde edilebileceği gösterilmiştir. Bu yöntemler, sahte konuşma tespiti alanında önemli başarılar elde etmiştir.

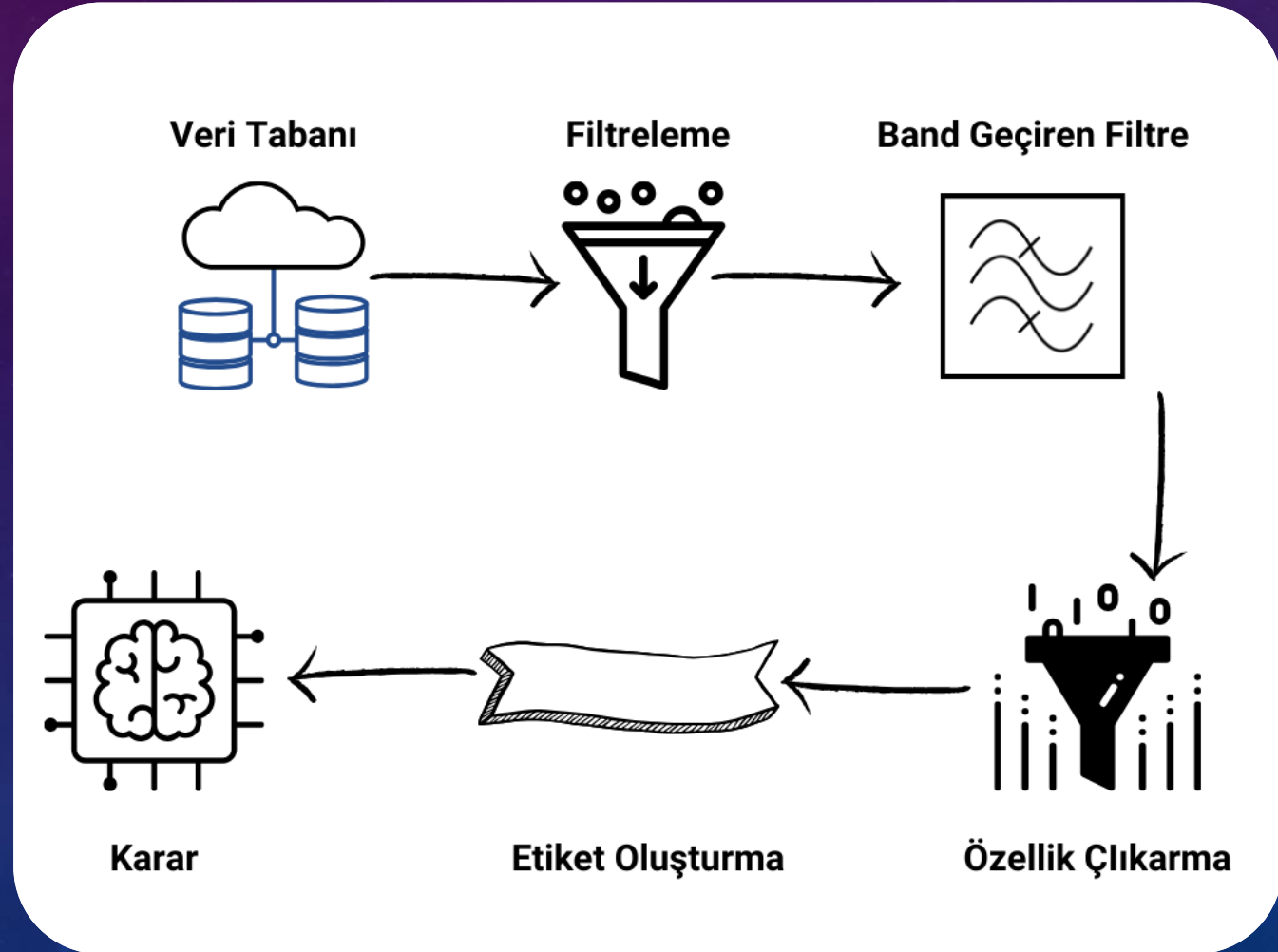
3-LİTERATÜR TARAMASI

- Başka bir çalışma [4], derin sinir ağları tarafından üretilen sentetik sesleri insan seslerinden ayırt etmek için bispektral analiz kullanmaktadır. Bispektral analiz, sentetik seslerin spektral ilişkilerini ölçerek insan seslerinden farklılık gösteren özellikleri belirlemeye yardımcı olmaktadır. Bispektral analiz, ses sinyallerinin Fourier dönüşümü kullanılarak yüksek dereceli spektral ilişkileri ölçen bir tekniktir. Elde edilen sonuçlar, bu bispektral özelliklerin insan sesinden sentetik sesleri ayırt etmede başarılı olduğunu göstermektedir. Ayrıca, sentetik seslerin farklı türlerini ayırt etme konusunda da başarılı olduğu belirtilmektedir. Sonuç olarak, bispektral analiz, derin sinir ağları tarafından üretilen sentetik sesleri tespit etme konusunda etkili bir yöntem olarak öne çıkmaktadır

4- PROJENİN GENEL AKIŞ DİYAGRAMI



4- PROJENİN AYRINTILI AKIŞ DİYAGRAMI



5-YÖNTEM

- Projeye başlarken, metodolojinin eksikliği nedeniyle Kaggle.com üzerinden elde edilen sentetik ve gerçek ses verileri, filtre ve algoritmalar kullanılarak işleme tabi tutulmuştur. İlk olarak, Python programının algoritma oluşturma becerisinin daha fazla olduğu düşünüldüğünden, bu programda scipy.signal, NumPy, matplotlib.pyplot, librosa kütüphaneleri kullanılmış ve gözlemlere dayanarak kıyaslama yapılmaya çalışılmıştır. Daha sonra nümerik işlemlerin daha kolay yapılması ve yanlış kodların düzenlenmesi amacı ile kodlar Matlab ortamında tekrar düzenlenmiştir.
- Daha önce eksik olan metodoloji düzenlendi ve tüm kodlar yeniden oluşturuldu. Ayrıca, algoritmanın çıkışına karar verme mekanizması eklendi.
- Elle sistematığe göre oluşturulan kodlar, sıfırdan Signal Labeler ve Classification Learner toolbox'ları ile düzenlenerek, doğruluk düzeyi çok daha yüksek sonuçlar elde edilmiştir.

5-YÖNTEM

- **VERİ TABANI:** Yoğun literatür taraması sonucunda, hazır olarak üretilmiş ve içerisinde ayırt etmesi oldukça zor olan sentetik ve gerçek ses kayıtlarını içeren 10 saniyelik örnekler bulunmuştur [5].
- Matlab ortamına herbir ses dosyası matris sütun vektörü olarak atanmıştır.

5-YÖNTEM

- **ÖN İŞLEME:** Bu aşamada, ileride uzay alanında uygulanacak algoritmalar nedeniyle veriler, 2 saniyelik parçalara ayrılmıştır.
- Daha sonra, sinyal işleme tekniklerinde kullanılan gürültü temizleme için Medyan Filtre kullanılmıştır.
- Son olarak, ses işleme alanında çalışıldığı için 300Hz-4KHz aralığındaki veriler seçilmiştir. Bu amaçla Band Geçiren Filtre kullanılmıştır.

Generate Frequency-Domain features

Select Features To Extract

<input checked="" type="checkbox"/> Mean Frequency	<input checked="" type="checkbox"/> Median Frequency	<input checked="" type="checkbox"/> Band Power
<input type="checkbox"/> Occupied Bandwidth	<input type="button" value="Parameters"/>	
<input type="checkbox"/> Power Bandwidth	<input type="button" value="Parameters"/>	

Specify parameters of Welch's PSD used for selected features

☐ Window Length(s)

Window Parameter

☐ Overlap Length(s)

☐ Specify FFT Length or Frequency Vector

☒ FFT Length

☐ Frequency Vector(Hz)

Min Max Length

Generate Time-Domain features

Select Features To Extract

Statistics

<input checked="" type="checkbox"/> Mean	<input checked="" type="checkbox"/> Standard Deviation	<input checked="" type="checkbox"/> RMS
<input type="checkbox"/> Shape Factor		

Pulse Metrics

<input type="checkbox"/> Crest Factor	<input type="checkbox"/> Peak Value	<input type="checkbox"/> Impulse Factor
<input type="checkbox"/> Clearance Factor		

Harmonics

<input type="checkbox"/> SNR	<input type="checkbox"/> SINAD	<input type="checkbox"/> THD
------------------------------	--------------------------------	------------------------------

5-YÖNTEM

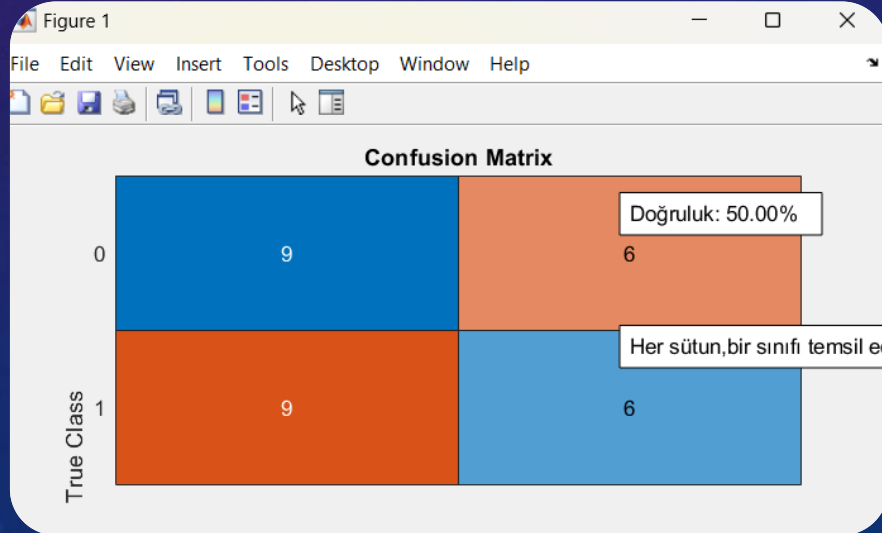
- **ÖZELLİK ÇIKARMA:** El ile yazılan kodlarda hem frekans hem zaman domenine ait 10 özellik çıkartılıp her verinin özellikleri 10 satırlık veri matrisi sütun boyutunda yeni bir matrise atanmıştır.
- Toolbox kısmında ise 3 adet zaman domeninde, 3 adet frekans domeninde şekilde gösterildiği gibi özellik çıkartılmıştır.

5-YÖNTEM

- **ÖZELLİK SEÇME:** Özellik çıkarma işlemi sırasında çok değişimi gözle gözlenemeyen veya çok ayırt edici özelliği olmayan algoritmalar çıkartılmıştır. Elle yapılan kodda 10 özellik kullanılırken Toolbox uygulamasında sadece 6 özellik kullanılmıştır.

5-YÖNTEM

- **ETİKET VE KARAR VERME:** Elde edilen özellikler matrisi ilk olarak 0 ve 1 olarak etiketlendirilerek hangi sesin gerçek hangi sesin sahte olduğuna dair etiket verilmiştir. Elle yapılan kodda doğruluk oranını arttırmak için randperm fonksiyonu kullanılıp doğruluk olasılığı artırılmıştır. Toolbox kısmında ise özellik matrisine yeni bir sütun eklenerek atama yapılmıştır.
- Karar verme mekanizması olarak elle yapılan metoddan 70% train, 30% test edilmiş olup KNN kullanılmıştır. Değişken doğruluk değerinden dolayı Gauss Çan Eğrisi'ne benzer doğruluk sonucu alınmıştır.



5-YÖNTEM

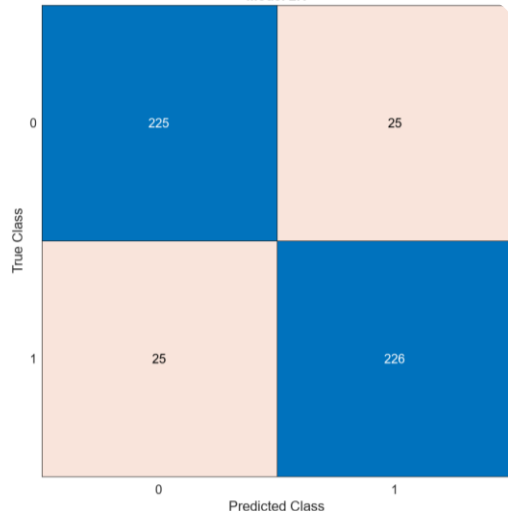
- Toolbox yönteminde ise test sonuçları oldukça yüksek çıkmıştır.

Models		
Sort by	Model Number	
1	Tree	Accuracy (Validation): 89.6%
Last change:	Fine Tree	6/6 features
5.1	Tree	Accuracy (Validation): 89.6%
Last change:	Fine Tree	6/6 features
5.2	Tree	Accuracy (Validation): 89.6%
Last change:	Medium Tree	6/6 features
5.3	Tree	Accuracy (Validation): 88.2%
Last change:	Coarse Tree	6/6 features
5.4	Linear Discriminant	Accuracy (Validation): 84.7%
Last change:	Linear Discriminant	6/6 features
5.5	Quadratic Discriminant	Accuracy (Validation): 90.9%
Last change:	Quadratic Discriminant	6/6 features
5.6	Binary GLM Logistic Reg...	Accuracy (Validation): 86.0%
Last change:	Binary GLM Logistic Regression	6/6 features
5.7	Efficient Logistic Regres...	Accuracy (Validation): 83.6%
Last change:	Efficient Logistic Regression	6/6 features
5.8	Efficient Linear SVM	Accuracy (Validation): 82.7%
Last change:	Efficient Linear SVM	6/6 features
5.9	Naive Bayes	Accuracy (Validation): 89.1%
Last change:	Gaussian Naive Bayes	6/6 features
5.10	Naive Bayes	Accuracy (Validation): 88.0%
Last change:	Kernel Naive Bayes	6/6 features
5.11	SVM	Accuracy (Validation): 81.8%
Last change:	Linear SVM	6/6 features

5.11	SVM	Accuracy (Validation): 81.8%
Last change:	Linear SVM	6/6 features
5.12	SVM	Accuracy (Validation): 85.8%
Last change:	Quadratic SVM	6/6 features
5.13	SVM	Accuracy (Validation): 86.0%
Last change:	Cubic SVM	6/6 features
5.14	SVM	Accuracy (Validation): 83.1%
Last change:	Fine Gaussian SVM	6/6 features
5.15	SVM	Accuracy (Validation): 83.1%
Last change:	Medium Gaussian SVM	6/6 features
5.16	SVM	Accuracy (Validation): 83.8%
Last change:	Coarse Gaussian SVM	6/6 features
5.17	KNN	Accuracy (Validation): 83.6%
Last change:	Fine KNN	6/6 features
5.18	KNN	Accuracy (Validation): 82.9%
Last change:	Medium KNN	6/6 features
5.19	KNN	Accuracy (Validation): 82.5%
Last change:	Coarse KNN	6/6 features
5.20	KNN	Accuracy (Validation): 85.1%
Last change:	Cosine KNN	6/6 features
5.21	KNN	Accuracy (Validation): 81.6%
Last change:	Cubic KNN	6/6 features
5.22	KNN	Accuracy (Validation): 83.8%
Last change:	Weighted KNN	6/6 features

5.23	Ensemble	Accuracy (Validation): 88.7%
Last change:	Boosted Trees	6/6 features
5.24	Ensemble	Accuracy (Validation): 90.2%
Last change:	Bagged Trees	6/6 features
5.25	Ensemble	Accuracy (Validation): 83.6%
Last change:	Subspace Discriminant	6/6 features
5.26	Ensemble	Accuracy (Validation): 76.7%
Last change:	Subspace KNN	6/6 features
5.27	Ensemble	Accuracy (Validation): 89.1%
Last change:	RUSBoosted Trees	6/6 features
5.28	Neural Network	Accuracy (Validation): 90.0%
Last change:	Narrow Neural Network	6/6 features
5.29	Neural Network	Accuracy (Validation): 89.1%
Last change:	Medium Neural Network	6/6 features
5.30	Neural Network	Accuracy (Validation): 91.1%
Last change:	Wide Neural Network	6/6 features
5.31	Neural Network	Accuracy (Validation): 89.6%
Last change:	Bilayered Neural Network	6/6 features
5.32	Neural Network	Accuracy (Validation): 86.5%
Last change:	Trilayered Neural Network	6/6 features
5.33	Kernel	Accuracy (Validation): 82.7%
Last change:	SVM Kernel	6/6 features
5.34	Kernel	Accuracy (Validation): 84.9%
Last change:	Logistic Regression Kernel	6/6 features

EN YÜKSEK



```
>> [yfit,scores] = trainedClassifier.predictFcn(or3)

yfit =

     1

scores =

     0     1
```

5-YÖNTEM

- Elle çözme metodunu kıyaslamak amacı ile KNN kullanılmıştır. Doğruluk tablosu gösterildiği gibidir. Yüklenen 3 örnek ses için doğru sonuç verdiği de gösterilmiştir.

```
Command Window

validationAccuracy =

    0.8523

>> [yfit,scores] = trainedClassifier.predictFcn(or1)

yfit =

     0

scores =

     1     0
```

```
Command Window

>> [yfit,scores] = trainedClassifier.predictFcn(or2)

yfit =

     0

scores =

     1     0
```


KAYNAKÇA

- [1] Kasapoğlu, B. ve Koç, T. (2020). Sentetik ve dönüştürülmüş konuşmaların tespitinde genlik ve faz tabanlı spektral özniteliklerin kullanılması. Avrupa Bilim ve Teknoloji Dergisi.
- [2] Türkmen, H. İ. ve Karslığil, M. E. (2009). Disfonik konuşmanın yeniden yapılandırılarak normal konuşma elde edilmesi reconstruction of dysphonic speech for synthesizing normally phonated speech. İeee Xplore.
- [3] Subramani, N., & Rao, D. (2020). Learning Efficient Representations for Fake Speech Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 34(04), 5859-5866.
<https://doi.org/10.1609/aaai.v34i04.6044>
- [4] Albadawy, E. A. ve Lyu, S. (2019). Detecting ai-synthesized speech using bispectral analysis. Research Gate.
- [5] Ballesteros, D. M., Rodríguez, Y. ve Renza, D. (2020). A dataset of histograms of original and fake voice recordings (h-voice). Science Direc.