



Türk Doğa ve Fen Dergisi

Turkish Journal of Nature and Science

www.dergipark.gov.tr/tdfd



LSTM Tabanlı Derin Ağlar Kullanılarak Diyabet Hastalığı Tahmini

Mehmet Bilal ER¹, İbrahim IŞIK^{2*}

¹ Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Şanlıurfa, Türkiye
² İnönü Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü, Malatya, Türkiye
 Mehmet Bilal ER ORCID No: 0000-0002-2074-1776
 İbrahim IŞIK ORCID No: 0000-0003-1355-9420

*Sorumlu yazar: ibrahim.isik@inonu.edu.tr

(Alınış: 30.10.2020, Kabul: 11.02.2021, Online Yayınlanma: 25.06.2021)

Anahtar Kelimeler
 Uzun kısa süreli bellek, Diyabet tahmini, Evrimsel sinir ağ

Öz: Diyabet, vücudun yeterli miktarda insülini üretmemesi veya iyi kullanmadığı durumda kan şekerinin normalin üstüne çıkması ile ortaya çıkan bir hastalıktır. Kan şekeri insanların ana enerji kaynağıdır ve bu enerji tüketilen yiyeceklerden gıdalardan gelir. Bu hastalık tedavi edilmez ise ölümcül olabilir. Ancak, erken tanı konulup tedaviye başlandığında tedavisi en olanaklı hastalıklardan biridir. Geleneksel diyabet teşhis süreci zorlu olduğundan, diyabetin klinik ve fiziksel verileri kullanılarak yapay sinir ağı, görüntü işleme ve derin öğrenme gibi sistemler kullanılarak hastalık teşhis edilebilmektedir. Bu çalışmada diyabet teşhisi için derin öğrenmeye dayalı bir model sunulmaktadır. Bu bağlamda Evrimsel Sinir Ağı (ESA), Uzun Kısa Süreli Bellek (Long-short Term Memory Networks- LSTM) modelinin hibrit kullanımı sınıflandırma için tercih edilmiştir. Ayrıca ESA ve LSTM modelleri deneylerde ayrı ayrı kullanılmıştır. Önerilen modelin performansını değerlendirmek için literatürde yaygın olarak kullanılan Pima Indians Diabetes veri seti kullanılmıştır. En yüksek sınıflandırma başarısı %86,45 olarak ESA+LSTM modelinden elde edilmiştir.

68

Prediction of the Diabetes Using LSTM-Based Deep Networks

Keywords
 Long-short Term Memory, Diabetes prediction, Convolutional neural network

Abstract: Diabetes is a metabolic disease that occurs when blood sugar is too high. Blood sugar is your main source of energy and comes from the food we eat. Sometimes our body doesn't produce enough or doesn't use insulin well. If this disease is not treated, it can be fatal. It is one of the most possible diseases when diagnosed early and treatment is started. Because the diabetes diagnosis process is challenging, an automated system is needed to identify diabetes from clinical and physical data. In this research, a deep learning-based model is presented for the diagnosis of diabetes. In this paper, the hybrid use of Convolutional Neural Network (CNN) and Long-short Term Memory Networks (LSTM) model was proposed for classification. In addition, CNN and LSTM models were used separately in the experiments. Pima Indians Diabetes data set, which is widely used in the literature, was used to evaluate the performance of the proposed model. The highest classification success was obtained from CNN + LSTM model with 86.45%.

1. GİRİŞ

Bazı hastalıkların teşhis süreci zorlu olduğundan, diyabet gibi çeşitli hastalıkların klinik ve fiziksel verileri kullanılarak yapay sinir ağı, görüntü işleme ve derin öğrenme gibi sistemler kullanılarak hastalık teşhis edilebilmektedir [1]. Diyabet, küçük ve büyük tüm yaş gruplarında görülebilen en yaygın kronik, hastalıklardan biridir. Pankreas, insan kan akışındaki şeker seviyesini düzenleyen çok önemli bir enzim olan insülin adlı hormonun üretilmesinden sorumlu bir organdır. İnsülin,

insan vücudunda metabolizmanın dengelenmesinde hayati ve önemli bir bileşendir. Diyabetin, insülinin yetersiz üretimi veya üretilen insülinin uygun olmayan kullanımı nedeniyle ortaya çıktığı bilinmektedir [2]. Ancak diyabetin altında yatan neden diyabetin Tip 1 ve Tip 2 çeşidi olmasına göre değişiklik göstermektedir.

Tip 1 ve tip 2 olmak üzere iki farklı diyabet türü vardır [3]. İnsülin eksikliği nedeniyle, Tip 1 diyabet meydana gelir. Normalde zararlı bakteriler veya virüslerle savaşan bağışıklık sisteminiz bu hastalık durumunda pankreastaki insülin üreten hücrelere saldırır ve tahrip

eder. Bu tip diyabet için tek çözüm, gerekli miktarda insülini hastanın vücuduna takviye olarak enjekte etmektir. Tip 2 diyabette, hücreler insülin etkisine dirençli hale gelir ve pankreas bu direncin üstesinden gelmek için yeterli insülin salınımı yapamaz. Tip 2 diyabet gelişiminde de genetik ve çevresel faktörlerin rol oynadığına inanılmaktadır. Aşırı kilolu olmak, tip 2 diyabet gelişimi ile güçlü bir şekilde bağlantılıdır. Diyabet yaygınlığı, orta ve düşük gelirli ülkelerde daha hızlı bir şekilde artmaktadır. Diyabetin, körlük, böbrek yetmezliği ve kalp krizinin ana nedenleri arasında olduğu bilinmektedir.

Bir diyabet hastası uygun tanı ve tedavi ile normal bir yaşam sürdürebilir. Diyabetin tanımlanması için A1c, rastgele kan şekeri, açlık kan şekeri ve oral glikoz tolerans testi gibi farklı ölçümler vardır [4]. Tek bir parametreye dayalı diyabet tanımlaması yanlış tanıya ve yanıltıcı karar vermeye yol açabilir. Bu nedenle, diyabeti etkili bir şekilde teşhis etmek için farklı parametreleri birleştirmeye ihtiyaç vardır. Glikoz, vücut kütle indeksi, diyabet soyağacı, kan basıncı, yaş, gebelik ve cilt kalınlığı gibi verilerle diyabeti tanımlamak ve tedavi etmek daha etkili olacaktır.

Hastalığın tahmini ve erken teşhisi bir doktorun bilgi ve tecrübesine bağlıdır, ancak bazı durumlarda yanlış teşhisler olabilir. Sağlık sektörü, sağlık hizmetleriyle ilgili çok miktarda veri toplamaktadır, ancak bu veriler tespit edilmemiş vakalarda etkin şekilde kullanılmamaktadır. Manuel kararlar, sağlık görevlisinin her zaman doğru olmayan gözlem ve yargısına dayandığından erken hastalık teşhisi için oldukça tehlikeli olabilir [5]. Sonuç olarak, hastalar düşük hizmet almaktadır; bu nedenle, otomatik tanı ve daha iyi doğruluk ile hastalıkların erken tespiti için farklı ileri mekanizmalar ve yazılım tabanlı çeşitli programların gerekli olduğu düşünülmektedir. Örneğin, makine öğrenimi algoritmalarının geliştirilmesi, büyük miktarda veriyi işlemeye ve karar verme sürecini kolaylaştıran temel veri modelini çıkarmaya yardımcı olur.

Bu çalışmada, diyabeti tahmin etmek için ESA ve LSTM modellerinin birleşimi kullanılmıştır. Modeller için en uygun konfigürasyonu bulmak için farklı parametreler denenerek modeller test edilmiştir. Ayrıca önerilen modelin performansı literatürdeki diğer yöntemler ile karşılaştırılmıştır.

Bu makalenin geri kalanı şu şekilde düzenlenmiştir. Bölüm 2'de literatürdeki çalışmalar gözden geçirilmiştir ve aralarındaki farklar ortaya koyulmuştur. Bölüm 3 ve 4'de materyal ve önerilen yöntem tanıtılmıştır. Bölüm 5'de araştırmada kullanılan veri seti ve diyabet tespit edilmesiyle ilgili deneysel uygulamalar verilmiştir. Bölüm 6'de ise araştırmanın bulguları tartışılmıştır.

2. İLGİLİ ÇALIŞMALAR

Diyabet teşhisi için oluşturmuş literatürde çok sayıda modern teknik vardır. Diyabet tahmini için makine öğrenme algoritmalarının performansının değerlendirildiği [6] çalışmada kullanılan algoritmalar

destek vektör makinesi (DVM), yapay sinir ağı (YSA), lojistik Regresyon ve K-en yakın komşu (K-NN) şeklindedir. Sistemin performansı, doğruluk, özgülük, duyarlılık, kesinlik açısından değerlendirilmiştir. Test ve eğitim verilerini ayırmak on kat çapraz doğrulama yöntemi tercih edilmiştir. En yüksek doğruluk %78 olarak lojistik Regresyon kullanılarak elde edilmiştir. [7] çalışmasında, diyabet sınıflandırması için çoklu sınıflandırma algoritmaları karşılaştırılmıştır. Kullanılan algoritmalar YSA, Bayes sınıflandırıcı, Karar ağacı, DVM ve K-NN'dir. En iyi performans olarak %97,44 olarak YSA'dan elde edilmiştir. DVM, K-NN, Karar ağacı ve Bayes sınıflandırıcılar ile sırasıyla %81,19, %90,85, %95,03 ve %91,60 doğruluk elde edilmiştir. Sistem için kullanılan veri seti, Tabriz Üniversitesi Tıp Bilimleri Üniversitesinden toplanmıştır ve tip 2 diyabet için taranan 2536 vaka içerir. Yazarlar [8] çalışmasında, ESA, LSTM ve giriş olarak kalp atım hızı değişkenliği (HRV) verilerinin bileşik kronolojik özelliklerini tespit etmek için bir model geliştirmişlerdir. Verileri sınıflandırmak için tespit edilen özelliklere DVM uygulanmıştır. DVM'ye dayalı çalışmalara kıyasla ESA ve ESA-LSTM mimarisinde sırasıyla % 0,03 ve % 0,06 performans artışı elde edilmiştir. Diyabet teşhisi için akıllı bir DVM modelinin sunulduğu [8] çalışmada, yazarlara göre diyabet, dünya çapında önemli bir sağlık sorunudur ve erken aşamada tespit edilirse tip 2 diyabet komplikasyonlarının %80'e kadar azalabilir. Önerilen senaryoda, diyabet tahmini için birçok veri madenciliği ve makine öğrenme algoritması analiz edilmiştir. Diyabet tahmini için hadoop tabanlı kümelerde makine öğrenme algoritmalarının yeni uygulanması önerildiği [9] çalışmasında, algoritmanın çalışmasını değerlendirmek için Ulusal Diyabet ve Sindirim Hastalıkları Enstitüsü'nden alınan Pima Indians Diyabet veri seti kullanılmıştır. Rastgele orman algoritması, 4 düğümlü hadoop küme ortamında karar ağacı ve bayes algoritmasından daha yüksek doğruluk sağlamıştır. [10] çalışmasında, K-ortalama ve lojistik regresyon algoritmalarını kullanarak diyabet tip-2'nin tespiti için yeni bir model önerilmiştir. Önerilen yöntemler hem küme hem de sınıf yöntemlerinden oluşmaktadır. Önerilen yöntemler, diyabeti tahmin etmede doğruluğu %3 artırmaktadır. Evrişimli Uzun Kısa Süreli Belleğe (Conv-LSTM) dayalı yeni bir diyabet sınıflandırma modeli geliştirildiği [10] çalışmasında, ESA, geleneksel LSTM ve Conv-LSTM gibi üç popüler model önerilmiş ve modellerin performansları değerlendirilmiştir. Uygulanan modeller için en uygun parametreleri bulmak amacıyla Grid arama algoritması kullanarak hiper parametre optimizasyonu gerçekleştirilmiştir. İlk deneyde veri seti test ve eğitim için farklı oranlarda bölünmüştür. İlk deneyde, Conv-LSTM tabanlı model diyabet hastalarını %91,38 ile en yüksek doğrulukla sınıflandırmıştır. İkinci deneyde ise çapraz doğrulama tekniği uygulanmıştır. Çapraz doğrulama tekniğini kullanarak Conv-LSTM modeli ile %97,26 ile en yüksek doğruluk elde edilmiştir. Pima Indians Diabetes'' veri seti ile farklı makine öğrenme algoritmaları kullanarak diyabet tahmini için bir metodolojinin sunulduğu [2] çalışmasında, YSA, Bayes sınıflandırıcı, karar ağacı ve derin öğrenme ile elde edilen doğruluklar genel olarak %90-98 aralığındadır. En iyi sınıflandırma

sonucu %98,07 olarak derin öğrenmeden elde edilmiştir. Çalışmada, derin öğrenme yaklaşımının doğruluğu omik veriler dahil edilerek daha da arttırılabileceği belirtilmiştir. [11]'de diyabet tahmini için LSTM sinir ağı uygulaması geliştirilmiştir. Önerilen modeli doğrulamak için literatürde kullanılan tipik bir veri kümesi kullanılmıştır. Çalışma, yapay kayıtlar kullanılarak doğru bir LSTM veri analizi için izlenecek prosedüre odaklanmıştır. Ayrıca geleneksel YSA ve LSTM yaklaşımları karşılaştırılmıştır. LSTM yaklaşımı sonuçları güncel YSA çalışmaları ile karşılaştırıldığında, test seti doğruluğunda %6,5'lik bir iyileşme gözlenmiştir. Etkin diyabet tahmini için hibrit bir yöntem önerildiği [12] çalışmasında Pima diyabetik veri kümesi kullanılarak önerilen model test edilmiştir. Önerilen hibrit modelin ilk aşamasında diyabet oluşma olasılığını yansıtan en ayırt edici belirleyicileri seçmek için özelliği seçimi işlemi uygulanmıştır. İkinci aşamada, modelin genel tanıma oranını arttırmak için DVM ve sinir ağı birleştirilerek filtrelenmiş verilere iki katmanlı bir sınıflandırma uygulanmıştır. Önerilen hibrit modelden %96,09 genel doğruluk elde edilmiştir.

3. MATERYAL VE METOT

3.1. Evrişimsel Sinir Ağı

Evrişimsel sinir ağları, özellikle görüntü nesnesi tanıma gibi görüntü analizi ve bilgisayarlı görmede başarıyla uygulanmış çok katmanlı, ileri beslemeli yapay sinir ağları gurubundadır. Evrişim ağları beyindeki biyolojik süreçlerden esinlenmiştir. Nöronlar arasındaki bağlantı paterni, görsel korteksün organizasyonuna benzer [13]. ESA'lar, görüntü tanıma ve sınıflandırma gibi alanlarda çok etkili olduğu kanıtlanmış bir sinir ağı modelidir [14]. ESA, tipik olarak üç tip katmandan oluşan matematiksel bir yapıdadır. Bu katmanlar konvolüsyon, havuzlama ve tamamen bağlı katmanlardır [15]. Konvolüsyon ve havuzlama katmanları, özellik çıkarma işlemini gerçekleştirirken, tamamen bağlı katmanlar ise çıkarılan özellikleri, sınıflandırmak için son çıktıya gönderir.

Bir sinir ağının görüntüdeki pikselleri sayısal değerler olarak yorumlayabilmesi için konvolüsyonlar gereklidir. Konvülyasyon katmanların işlevi, görüntüyü sinir ağının yorumlayabileceği sayısal değerlere dönüştürmek ve daha sonra ilgili kalıpları çıkarmaktır [15]. Konvolüsyon katmanında filtreler giriş görüntüsü üzerinde dolaştırılır. Konvolüsyon formülü denklem 1 de verilmiştir. Denklemde verilen 'M' özellik haritasını, 'w' ise (x,y) boyutunda konvolüsyon çekirdeğini ifade eder.

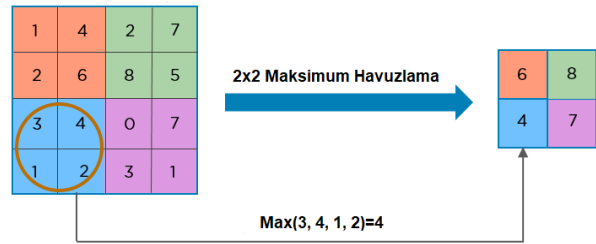
$$M(i,j)=(R*w)(i,j)=\sum_x \sum_y R(i-x,j-y)w(x,y) \quad (1)$$

Her konvolüsyon katmanından sonra, doğrusal olmayan bir katmanın veya aktivasyon katmanının uygulanması ESA'da yaygındır. Bu katmanın amacı, temelde doğrusal işlemleri hesaplayan bir sistemi doğrusal olmayan hale getirmektir. Doğrultulmuş Doğrusal Birim (ReLU) katmanı, gelen tüm değerlere $f(x) = \max(0, x)$ işlevini uygular. Temel olarak, bu katman tüm negatif aktivasyonları 0 olarak değiştirir. Bu katman, konvolüsyon katmanından gelen özellikleri etkilemeden

modelin ve genel ağı doğrusal olmayan özelliklerini arttırır [16]. ReLu için gerekli formül Denklem 2'de verilmiştir.

$$\text{ReLU}(x) = \max(x,0) \quad (2)$$

Havuzlama (pooling) katmanı bir ESA'nın başka bir yapı taşıdır ve konvolüsyon ve ReLu işleminden sonra uygulanır. İşlevi, ağıdaki parametre ve hesaplama miktarını azaltmaktır. Havuzlama katmanı her özellik haritasından bağımsız olarak çalışır. Havuzlamada kullanılan en yaygın yaklaşım maksimum havuzlamadır. Şekil 1'de maksimum havuzlama (maxpooling) örneği gösterilmektedir.



Şekil 1. Maksimum havuzlama

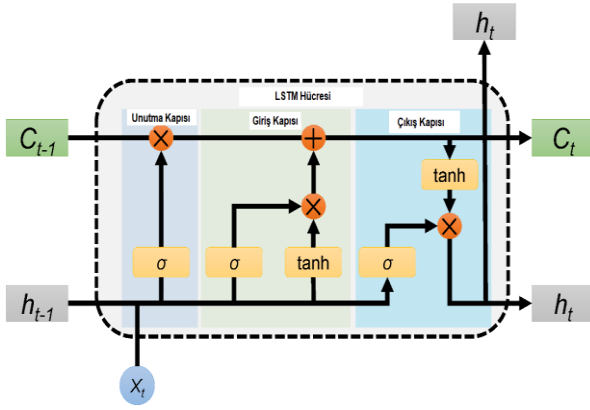
Konvolüsyon ve havuzlama katmanları tarafından alt örneklenen özellikler oluşturulduktan sonra bu özellikler tam bağlı katmana bağlanır. Özellikler bir veya daha fazla tamamen bağlı katmana bağlanır. Burada her girdi bir çıktıya bağlanır ve her bir nöron öğrenilebilir ağırlığa sahiptir [16]. Son tam bağlı katman tipik olarak sınıf sayısı ile aynı sayıda çıkış düğümüne sahiptir ve bu katmanda sınıflandırma işlemi yapılır. Son katmanda farklı sınıflandırıcılar kullanılabilir de genellikle Softmax kullanılır. Softmax denklemi 3'de verilmiştir. Softmax, nöronların çıkış değerlerinin (0,1) aralığında olmasını sağlar.

$$\text{softmax}(x)_j = \frac{e^{x_j}}{\sum_{n=1}^N e^{x_n}} \quad \text{for } j=1, \dots, N \quad (3)$$

3.2. Uzun Kısa Süreli Bellek

Uzun Kısa Süreli Bellek ağları genellikle "LSTM" olarak adlandırılır ve Tekrarlayan Yapay Sinir Ağlarının (RNN) özel bir türüdür. LSTM sinir ağı, gizli katmanında LSTM hücresi olarak adlandırılan karmaşık bir yapıya sahiptir. Şekil 2'de gösterilen LSTM hücresi, hücre ve sinir ağı boyunca bilgi akışını kontrol eden giriş kapısı, unutmaya kapısı ve çıkış kapısı olmak üzere üç kapıya sahiptir.

LSTM modeli bir zincir yapısı şeklinde düzenlenmiştir [17]. Ancak tekrar eden modül farklı bir yapıya sahiptir. Standart bir RNN gibi tek bir sinir ağı yerine, benzersiz bir iletişim yöntemine sahip dört etkileşimli katmana sahiptir. Tipik bir LSTM ağı, hücreler adı verilen bellek bloklarından oluşur. Hücre durumları bir sonraki hücreye aktarılır. Hücre durumu, verilerin temelde değişmeden ilerlemesine izin veren ana veri akış zinciridir.



Şekil 2. LSTM Modeli

Bununla birlikte, bazı doğrusal dönüşümler meydana gelebilir. Veriler, sigmoid kapılar aracılığıyla hücre durumuna eklenebilir veya buradan çıkarılabilir. Bir geçit, farklı bireysel ağırlıkları içeren bir katmana veya bir dizi matris işlemine benzer. Bir LSTM ağı oluşturmanın ilk adımı, gerekli olmayan ve hücreden çıkarılacak olan bilgileri belirlemektir. Bu veri tanımlama ve hariç tutma işlemine, $t-1$ zamanında son LSTM biriminin (h_{t-1}) çıktısını ve t zamanında mevcut girdiyi (X_t) alan sigmoid fonksiyonu karar verir. Ek olarak, sigmoid fonksiyonu eski çıktının hangi kısmının elenmesi gerektiğini belirler. Bu kapıya unutma kapısı (veya f_t) denir; burada h_{t-1} , 0 ile 1 arasında değişen ve hücre durumundaki her sayıya karşılık gelen C_{t-1} olan bir vektördür.

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (4)$$

Burada, σ sigmoid fonksiyonudur ve W_f ve b_f , sırasıyla, unutma kapısının ağırlık matrisleri ve bias'dır.

Aşağıdaki adım, yeni girişten (X_t) gelen bilgileri kararlaştırır, depolar ve ayrıca hücre durumunu günceller. Bu adım, sigmoid katman ve ikinci tanh katman olmak üzere iki bölümden oluşur. İlk olarak, sigmoid katmanı, yeni bilginin güncellenip güncellenmeyeceğine (0 veya 1) karar verir ve ikincisi, tanh fonksiyonu, önem düzeylerine (-1 ila 1) karar vererek, geçen değerlere ağırlık verir. Yeni hücre durumunu güncellemek için iki değer çarpılır. Bu yeni bellek daha sonra eski belleğe C_{t-1} eklenir ve C_t ile sonuçlanır.

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i), \quad (5)$$

$$N_t = \tanh(W_n[h_{t-1}, X_t] + b_n), \quad (6)$$

$$C_t = C_{t-1}f_t + N_t i_t \quad (7)$$

Burada, C_{t-1} , C_t , $t-1$ ve t zamanlarındaki hücre durumlarıdır. Sigmoid katman, hücre durumunun hangi kısımlarının çıktıya ulaşacağına karar verir. Daha sonra, sigmoid geçidin (O_t) çıktısı, tanh katmanı tarafından hücre durumundan (C_t) oluşturulan yeni değerle çarpılır.

$$O_t = \sigma(W_o[h_{t-1}, X_t] + b_o), \quad (8)$$

$$h_t = O_t \tanh(C_t) \quad (9)$$

Burada W_o ve b_o , çıkış kapısının sırasıyla ağırlık matrisleri ve bias'dır.

4. ÖNERİLEN YÖNTEM

Diyabet tespiti için önerilen yöntem üç adımdan oluşmaktadır. İlk olarak veri setindeki örneklerden elde edilen özellikler tasarlanan ESA'dan geçirilmiştir. Daha sonra ESA'dan elde edilen özellikler LSTM'e girdi olarak verilmiştir. Son olarak Softmax sınıflandırıcı kullanılarak veri setindeki örnekler sınıflandırılmıştır. Önerilen yöntem şekil 3'de verilmiştir. Ayrıca aynı deneyler ESA ve LSTM ayrı ayrı kullanılarak gerçekleştirilmiştir.

Tasarlanan ESA'nın topolojisi, 1 giriş katmanı, 2 konvolüsyon katmanı, 2 havuzlama katmanı, 2 tam bağlantılı katmanı ve 1 çıkış katmanından oluşur. Tasarlanan mimaride toplam 6 katman bulunmaktadır. Birinci konvolüsyon katmanında 128 adet konvolüsyon filtresi, ikinci konvolüsyon katmanında 64 adet konvolüsyon filtresi bulunmaktadır. Aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Ayrıca her konvolüsyon ve ReLu işleminden sonra 2x1 boyutunda maksimum havuzlama yapılmıştır. Konvolüsyon katmanlarından sonra veriler LSTM katmanından geçirilmiştir. Son olarak tamamen bağlantı katmanlar ve Softmax bağlanarak model tamamlanmıştır. Tasarlanan ESA mimarisi Tablo 1'de verilmiştir.

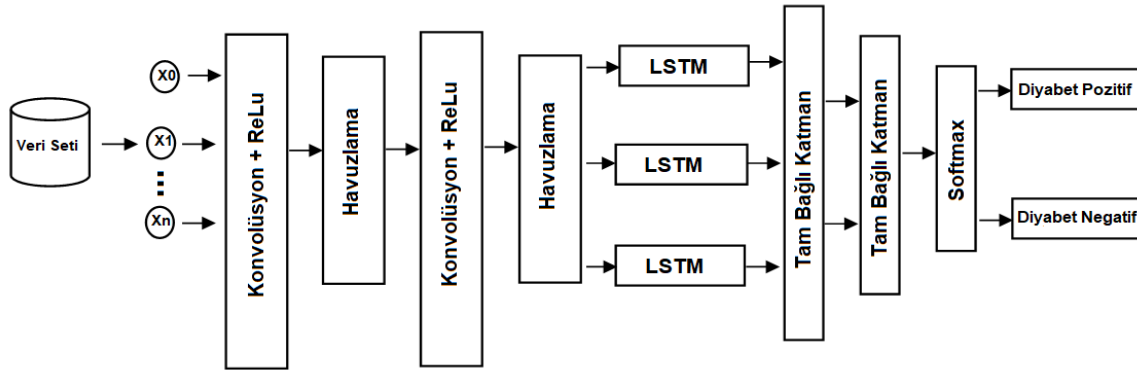
Tablo 1. Tasarlanan ESA mimarisi

No	Katman İsmi	Açıklama	Özellikler
1	Giriş	Giriş vektörü	-
2	'conv1'	Konvolüsyon	1 boyutunda 128 adet konvolüsyon filtresi
3	'relu1'	ReLu	-
4	'pool1'	Maksimum Havuzlama	2x1 havuzlama
5	'conv2'	Konvolüsyon	1 boyutunda 64 adet konvolüsyon filtresi
6	'relu2'	ReLu	-
7	'pool2'	Maksimum Havuzlama	2x1 havuzlama
8	'fc1'	Tam Bağlı Katman	1024 nöron
9	'relu7'	ReLu	-
10	'drop7'	Dropout	50% dropout
11	'fc2'	Tam Bağlı Katman	512 nöron
12	Çıkış	Softmax	2 sınıf

5. DENEYSEL UYGULAMALAR

5.1. Veri Seti

Modelimizin performansı, UCI makine öğrenimi havuzundan elde edilen popüler bir veri kümesi olan Pima Indians Diabetes kullanılarak değerlendirilmiştir [18]. Veri seti, 268 tanesi diyabet pozitif ve geri kalanı diyabet negatif olan en az 21 yaşındaki 768 kadın hastanın kayıtlarını içermektedir. Veri setinde, bir hastanın diyabetli olup olmadığını tanısal olarak tahmin



Şekil 3. Önerilen Yöntem

etmek için Gebelik sayısı, Glikoz, Kan Basıncı, Beden kitle indeksi, Deri Kalınlığı, İnsülin, Diyabet Soy ağacı ve Yaş gibi sekiz değişken ve sonuç olarak adlandırılan bir hedef değişken vardır. Veri setinin açıklaması Tablo 2’de verilmiştir.

Tablo 2. Pima Indian Diabetes Veri Kümesi

No	Özellikler	Açıklama	Aralık
1	Gebelik sayısı	Bir katılımcının hamile kalma sayısı	0-17
2	Glikoz	Oral glikoz tolerans testinde 2. saat plazma glikoz konsantrasyonu	0-199
3	Kan Basıncı	Diyastolik kan basıncı	0-122
4	Beden kitle indeksi	Vücut kitle indeksi (kg cinsinden ağırlık / (m cinsinden yükseklik)	0-67,1
5	Deri Kalınlığı	Deri kıvrım kalınlığı	0-99
6	İnsülin	2. Saatteki insülin değeri	0-846
7	Diyabet Soy ağacı	Diyabet prognozunda kullanılan bir özellik	0,078-2,42
8	Yaş	Katılımcıların yaşı	21-81
9	Sınıf	İkili değişken (0, 268 örneğin diyabetik olmadığını gösterir. 1 ise kalan 500 örneğin diyabetik olduğunu gösterir).	1/0

5.2. Deneysel Sonuçlar

Önerilen yöntemin başarımlarını kriterleri doğruluk, kesinlik ve F-skoru oranlarına dayalı olarak yapılmıştır.

- Yanlış pozitifler (YP): negatif sınıftan olan, pozitif olarak tahmin edilen örnekler.
- Yanlış negatifler (YN): gerçek sınıfı pozitif olan negatif olarak tahmin edilen örnekler.
- Doğru pozitifler (DP): pozitif sınıfa ait doğru tahmin edilen örnekler.
- Doğru negatifler (DN): negatif sınıfa ait olarak doğru tahmin edilen örnekler.

$$\text{Doğruluk} = \frac{|DN|+|DP|}{|YN|+|YP|+|DN|+|DP|} \quad (10)$$

Kesinlik (P), pozitif bir tahminin doğru olma olasılığını tahmin eden bir ölçüdür. Kesinlik ölçümü Denklem 11’de verilmiştir.

$$\text{Kesinlik(P)} = \frac{|DP|}{|DP|+|YP|} \quad (11)$$

F-skoru, pozitif kestirim oranı ve duyarlılık ölçülerinin uyumlu bir ortalaması olup Denklem12’de gösterildiği gibi hesaplanır.

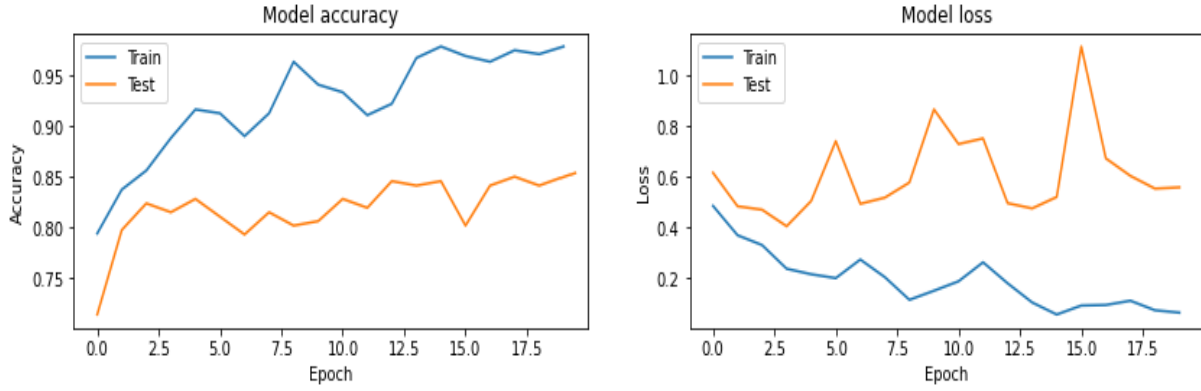
$$F - \text{skor} = \frac{2*|DP|}{2*|DP|+|YP|+|YN|} \quad (12)$$

Öğrenme oranı, epoch sayısı ve minibatch boyutu ESA’nın performansını etkileyen en önemli hiper parametrelerdir. Tasarlanan mimaride minibatch boyutu 64, maksimum epok sayısı 20 ve öğrenme oranı 1e-4 olarak ayarlanmıştır. Eğitim ve test için ayrılan verilerin boyutunun ağırlık performansını üzerindeki etkisini belirlemek amacıyla, bu veriler 2 farklı şekilde bölünmüştür. İlk deneyde, verilerin %70’i eğitim için, %30’i test için kullanılır; ikinci deneyde ise verilerin %80’i eğitim, %20’si test için kullanılmıştır.

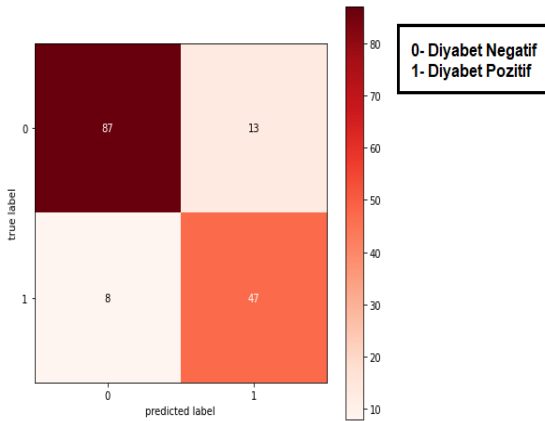
Tablo 3’de ESA, LSTM ve ESA+LSTM’den elde edilen sınıflandırma sonuçları verilmiştir. Verilerin eğitim ve test için sırasıyla %70-%30 olarak bölünmesiyle ESA’dan %82,47, LSTM’den %83,77 ve ESA+LSTM’den %85,21 doğruluk olarak elde edilmiştir. Verilerin eğitim ve test için sırasıyla %80-%20 olarak bölünmesiyle ESA’dan %83,25, LSTM’den %85,21 ve ESA+LSTM’den %86,45 doğruluk olarak elde edilmiştir. Eğitim ve test sıralarındaki en iyi sınıflandırma sonucuna ait doğruluk ve kayıp grafiği Şekil 4’de verilmiştir. En iyi sınıflandırma sonuçlarına ait karışıklık matrisleri Şekil 5’de verilmiştir.

Tablo 3. Sınıflandırma sonuçları

Verinin Eğitim ve Test için Farklı oranlarda Bölünmesi	Model	Doğruluk %	Kesinlik %	F-skoru %
%70- %30	ESA	82,47	83,14	83,56
	LSTM	83,77	84,46	84,23
	ESA+LSTM	85,21	84,94	58,14
%80- %20	ESA	83,25	83,56	83,69
	LSTM	85,21	85,33	85,41
	ESA+LSTM	86,45	87,00	88,23



Şekil 4. Öğrenme grafikleri



Şekil 5. En iyi sınıflandırma sonucuna ait karışıklık matrisi

Performans değerlendirmesi için elde ettiğimiz bulgularla literatürde kullanılan diğer yöntemlerden elde edilen sonuçlar karşılaştırılmıştır. Tablo 4'de önemli bazı çalışmalar verilmiştir.

Tablo 4. Performans karşılaştırılması

Yöntem	Sınıflandırıcı	Doğruluk %
Ashiqzaman ve ark. [19]	Derin Öğrenme Mimarisi	81,00
Rahman ve ark. [20]	Conv-LSTM	97,26
Massaro ve ark. [11]	LSTM	86,00
Dadgar ve ark.[21]	Genetik Algoritmali Sinir Ağı	87,46
Haritha ve ark. [22]	Ateşböceği Algoritmaları	81,00
Önerilen Yöntem	ESA-LSTM	86,45

Tablo 4'de görüldüğü gibi önerilen yöntem, literatürde kullanılan diğer üç yöntemden daha iyi bir performansa göstermiştir.

6. SONUÇLAR

Bu çalışmada, diyabetin ölçümü için bir tahmin modeli geliştirilmesi amaçlanmıştır. İnsan nüfusunun büyük bir kısmı diyabet hastalığından mustarıdır. Geliştirilen modelin uygulaması Pima Indian Diabetes veri seti üzerinde yapılmıştır. Geliştirilen modelde ESA ve LSTM mimarisi birlikte kullanılmıştır. Deney sırasında veri setindeki özellikler tasarlanan ESA+LSTM modeline girdi olarak verilmiştir. Sınıflandırma için tasarlanan mimarinin son katmanına Softmax yerleştirilmiştir. Ayrıca ESA ve LSTM mimarileri ayrı ayrı da kullanılarak aynı deneyler tekrar yapılmıştır. Veri seti ilk deneyde %70 eğitim %30 test, ikinci deneyde ise

%80 eğitim ve %20 test olarak bölünmüştür. Önerilen modelden %86,45 genel doğruluk elde edilmiştir. ESA ve LSTM'den ise sırasıyla %83,25 ve %85,21 doğruluk elde edilmiştir. Geliştirilen modelin son teknoloji modellerle karşılaştırılması, modelimizin diyabet tahmininde etkinliğini göstermektedir. Gelecekteki çalışmalarda, diyabetin erken teşhisinde sağlık uzmanlarına yardımcı olmak ve derin öğrenme algoritmalarını kullanarak uygulamalar geliştirilmesi planlanmaktadır.

KAYNAKLAR

- [1] Işık E, Toktamış H, Işık İ. Analysis of thermoluminescence characteristics of a lithium disilicate glass ceramic using a nonlinear autoregressive with exogenous input model. *Luminescence* 2020;35:827–34. <https://doi.org/10.1002/bio.3788>.
- [2] Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J Diabetes Metab Disord* 2020;19:391–403. <https://doi.org/10.1007/s40200-020-00520-5>.
- [3] Allam F, Nossai Z, Gomma H, Ibrahim I, Abdelsalam M. A Recurrent Neural Network Approach for Predicting Glucose Concentration in Type-1 Diabetic Patients BT - Engineering Applications of Neural Networks. In: Iliadis L, Jayne C, editors., Berlin, Heidelberg: Springer Berlin Heidelberg; 2011, p. 254–9.
- [4] Ramachandran A. Know the signs and symptoms of diabetes. *Indian J Med Res* 2014;140:579–81.
- [5] Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. 2008 IEEE/ACS Int Conf Comput Syst Appl 2008. <https://doi.org/10.1109/aiccsa.2008.4493524>.
- [6] Dwivedi AK. Analysis of computational intelligence techniques for diabetes mellitus prediction. *Neural Comput Appl* 2017;30:3837–45. <https://doi.org/10.1007/s00521-017-2969-9>.
- [7] Heydari M, Teimouri M, Heshmati Z, Alavinia SM. Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran. *Int J Diabetes Dev Ctries* 2015;36:167–73. <https://doi.org/10.1007/s13410-015-0374-4>.
- [8] G. S, R. V, K.P. S. Diabetes detection using deep learning algorithms. *ICT Express* 2018;4:243–6. <https://doi.org/10.1016/j.ict.2018.10.005>.
- [9] Yuvaraj N, SriPreetha KR. Diabetes prediction in

- healthcare systems using machine learning algorithms on Hadoop cluster. *Cluster Comput* 2017;22:1–9. <https://doi.org/10.1007/s10586-017-1532-x>.
- [10] Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. *Informatics Med Unlocked* 2018;10:100–7. <https://doi.org/10.1016/j.imu.2017.12.006>.
- [11] Massaro A, Maritati V, Giannone D, Convertini D, Galiano A. LSTM DSS Automatism and Dataset Optimization for Diabetes Prediction. *Appl Sci* 2019;9:3532. <https://doi.org/10.3390/app9173532>.
- [12] Gill N, Mittal P. A computational hybrid model with two level classification using SVM and neural network for predicting the diabetes disease 2016;87:1–10.
- [13] Fu Y, Aldrich C. Flotation froth image recognition with convolutional neural networks. *Miner Eng* 2019;132:183–90. <https://doi.org/10.1016/j.mineng.2018.12.011>.
- [14] GÜLCÜ A, KUŞ Z. Konvolüsyonel Sinir Ağlarında Hiper-Parametre Optimizasyonu Yöntemlerinin İncelenmesi. *Gazi Üniversitesi Fen Bilim Derg Part C Tasarım ve Teknol* 2019;7:503–22. <https://doi.org/10.29109/gujsc.514483>.
- [15] Cireundefinedan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J. Flexible, High Performance Convolutional Neural Networks for Image Classification. *Proc. Twenty-Second Int. Jt. Conf. Artif. Intell. - Vol. Vol. Two*, AAAI Press; 2011, p. 1237–1242.
- [16] ER MB. Akciğer Seslerinin Derin Öğrenme İle Sınıflandırılması. *Gazi Üniversitesi Fen Bilim Derg Part C Tasarım ve Teknol* 2020. <https://doi.org/10.29109/gujsc.758325>.
- [17] Olah C. Understanding LSTM Networks n.d.
- [18] Predict the Onset of Diabetes Based on Diagnostic Measures n.d.
- [19] Ashiquzzaman A, Tushar AK, Islam MR, Shon D, Im K, Park J-H, et al. Reduction of Overfitting in Diabetes Prediction Using Deep Learning Neural Network. *IT Converg Secur* 2017 2017:35–43. https://doi.org/10.1007/978-981-10-6451-7_5.
- [20] Rahman M, Islam D, Mukti RJ, Saha I. A deep learning approach based on convolutional LSTM for detecting diabetes. *Comput Biol Chem* 2020;88:107329. <https://doi.org/10.1016/j.compbiolchem.2020.107329>.
- [21] Dadgar SMH, Kaardaan M. A Hybrid Method of Feature Selection and Neural Network with Genetic Algorithm to Predict Diabetes, 2017.
- [22] Haritha R, Babu DS, Sammulal DP. A Hybrid Approach for Prediction of Type-1 and Type-2 Diabetes using Firefly and Cuckoo Search Algorithms, 2018.