

## **DIYABET HASTALIĞININ TESPİTİ İÇİN DERİN ÖĞRENMEYE DAYALI BİR SINIFLANDIRICI MİMARİSİ**

**Dr. Öğr. Üyesi Volkan GÖREKE (ORCID: 0000-0002-2418-8373)**

Sivas Cumhuriyet Üniversitesi, Sivas TBMYO, Bilgisayar Teknolojileri Bölümü, Sivas-  
Türkiye

**Email:** [vgoreke@cumhuriyet.edu.tr](mailto:vgoreke@cumhuriyet.edu.tr)

### **ÖZET**

Diyabet, metabolik sorunların bir dizisine işaret eden ve uzun bir dönem boyunca yüksek kan şekeri seviyeleri ile karakterize edilen bir durumdur. Vücut, insülin üretimini yeterince gerçekleştiremediği veya pankreas tarafından üretilen insülini etkili bir şekilde kullanamadığı zaman, bu durum ortaya çıkar. Diyabetle ilgili başlıca sorun, zamanında teşhis eksikliği veya doktorların yanlış teşhis koymasındır. Bu sebeple, diyabetin erken evrelerinde doğru bir şekilde teşhis edilmesine katkı sağlayacak bir teknik uygulanması, hastalığın önlenmesi ve kontrol edilebilmesi için son derece kritik bir adım olacaktır. Bu çalışmada diyabet hastalığının tahmini için Pima Indians Diabetes Database veri seti kullanılarak derin öğrenmeye dayalı bir sınıflandırıcı mimarisi tanıtılmıştır. Elde edilen sonuçlara göre bu çalışmada önerilen yöntem literatürdeki çalışmalardan daha üstün bir başarımla ortaya koymuştur.

**Keywords:** Diyabet, veri seti balanslama, derin öğrenme

## **A DEEP LEARNING-BASED CLASSIFICATION ARCHITECTURE FOR DETECTION OF DIABETES**

### **ABSTRACT**

Diabetes is a set of metabolic problems defined by high blood sugar levels over a long period of time. It occurs when the body cannot produce enough insulin or cannot effectively use the insulin produced by the pancreas. If diabetes is not treated on time, it can lead to lifelong complications such as cardiovascular disease, cerebral palsy, kidney failure, and eye complications. The biggest problem with diabetes arises from not being diagnosed on time or misdiagnosed by physicians. Therefore, the application of a technique that contributes to the correct diagnosis of diabetes, especially in its early stages, will be a very important step for the prevention and control of the disease. To overcome this, researchers have tried to predict diabetes symptoms by applying different approaches such as machine learning and data mining. Pima Indians Diabetes Database (PIDDD) data set was used in this study. There are a total of 768 samples in the data set. 500 samples have diabetes diagnosis. The parameters for each sample in the data set are named as age, pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index and diabetes pedigree function and include numerical values. The proposed classifier architecture is implemented using Python language on Visual Studio Community 2017 platform. In the study, firstly, the Generative Adversarial Networks model was designed and the balance problem of the data set was solved, and data augmentation was carried out at the same time. Then, a deep convolutional neural network was designed and trained to predict diabetes. The classifier architecture based on deep learning proposed in this study has shown a superior performance against the studies in the literature. At the same time, with the proposed architecture, the problem of insufficient and unbalanced data set that researchers in the biomedical field often encounter can be eliminated.

**Keywords:** Deep Learning, Diabetes Disease, Data Augmentation, Convolutional Neural Network

## **GİRİŞ**

Diyabet, uzun bir süre boyunca yüksek kan şekeri seviyesi ile tanımlanan bir dizi metabolik sorundur. Vücut yeterli insülin üretmediğinde veya pankreas tarafından üretilen insülini etkili bir şekilde kullanamadığında ortaya çıkar (Egan ve Dinneen, 2004,s.679). Diyabetin nedenleri, türüne göre değişir. Tip 1 diyabet, bağışıklık sisteminin insülin üreten pankreas hücrelerine saldırması sonucu ortaya çıkar. Bu durumun nedeni tam olarak bilinmemektedir ancak genetik ve çevresel faktörler rol oynayabilir (Atkinson vd., 2014). Tip 2 diyabet, vücudun insüline karşı direnç geliştirmesi sonucu ortaya çıkar. Bu durumun nedenleri arasında obezite, yaşlılık, fiziksel aktivite eksikliği sayılabilir (Hu, 2011). Gebelik diyabeti, hamilelik sırasında hormonal değişiklikler nedeniyle ortaya çıkar. Bu durumun nedenleri arasında kilolu olmak, ileri yaşta hamile kalmak sayılabilir (Metzger vd., 2010). Yüksek kan şekeri belirtileri arasında aşırı idrara çıkma, her zaman susuzluk hissi ve artan açlık vardır (Ramachandran, A. 2014, s.579). Diyabet zamanında tedavi edilmezse kardiyovasküler rahatsızlık, beyin felci, böbrek yetmezliği, ve göz komplikasyonları gibi ömür boyu sürececek komplikasyonlara yol açabilir (Anjana vd., 2011,s.3022-3027). Aşağıda belirtilen sağlık özelliklerine sahip bireyler diyabet hastalığı için yüksek risk grubundadır. Pek çok farklı etnik gruplar için 45 yaş üstü, uzun süreli hipertansiyon, 25'ten büyük bir Vücut Kitle İndeksi, diyabet hastası olan aile üyelerine sahip olmak ve hareketsiz bir yaşam tarzını benimsemek (Tigga ve Garg, 2020, s.707).

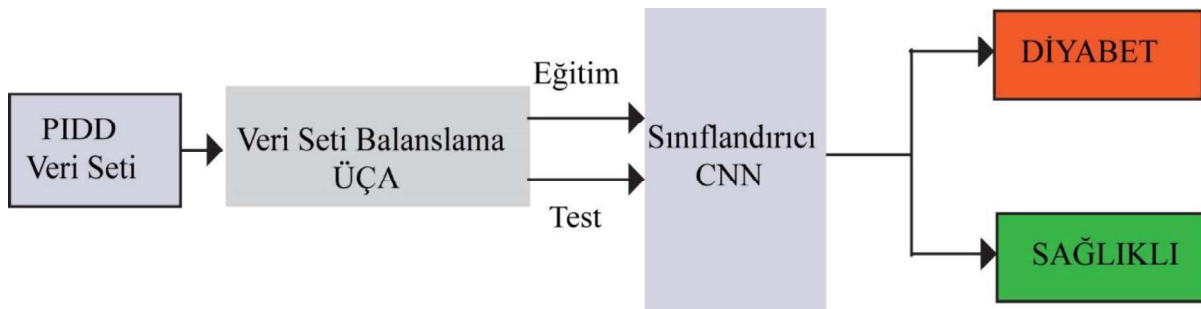
Diyabetle ilgili en büyük sorun, zamanında teşhis konulmaması veya hekimlerin yanlış teşhis koymasından kaynaklanmaktadır. Bu nedenle, diyabetin özellikle erken evrelerinde doğru teşhisine katkı sağlayan bir tekniğin uygulanması, hastalığın önlenmesi ve kontrolü için çok önemli bir adım olacaktır (Kamel ve Yaghoubzadeh, 2021, s.2 ). Bunun üstesinden gelebilmek için araştırmacılar makine öğrenimi ve veri madenciliği gibi farklı yaklaşımlar uygulayarak diyabet semptomlarını tahmin etmeye çalışmışlardır (Kavakiotis vd. 2017 s.104 ).

Sisodia ve Sisodia Pima Indians Diabetes Database (PIDD) veri setini kullanarak gerçekleştirdikleri çalışmada, diyabetin erken bir aşamada tespit edilmesi için Karar Ağacı, SVM ve Naive Bayes olmak üzere üç makine öğrenimi sınıflandırma algoritması kullanılmıştır. En yüksek sınıflandırma performansını Naive Bayes kullanarak 76,30% değerinde elde etmişlerdir (Sisodia ve Sisodia, 2018,s.1580). Yang ve arkadaşları, PIDD veri seti kullanarak gerçekleştirdikleri çalışmada K-ortalamalar algoritması ve Lojistik Regresyon (LR) algoritması olmak üzere iki bölümden oluşan bir model önermişlerdir. Yazarlar önerdikleri bu model ile diyabet tespitini 95.42% başarıyla yapmışlardır (Yang vd. 2018. s.100-107). Choubey ve Paul diyabet hastalığının tespiti için iki aşamalı bir yöntem önermişlerdir. Yazarlar Genetik

algoritma kullanarak öznelik seçme işlemi yapmışlar ardından Radyal Temel Fonksiyonlu Sinir Ağı (RTF SA) ile diyabet tespiti gerçekleştirmişlerdir. Yazarlar PIDD veri seti üzerindeki bu çalışma ile 76.087% başarımlar elde etmişlerdir (Choubey ve Paul, 2017,s.71-93). Zou ve arkadaşları Sinir Ağı (SA) sınıflandırıcı ile diyabet hastalığını tespit etmek için PIDD veri seti kullanarak 76,67% doğruluk başarımlarını elde etmişlerdir (Zou vd.2018 ). Tigga ve Garg PIDD veri seti kullanarak diyabet hastalığını tahmin etmek için çeşitli makine öğrenme tekniklerini sınıflandırıcı olarak kullanmışlardır. Rastgele Orman (RO) sınıflandırıcı ile en yüksek 75% doğruluk başarımlarını elde etmişlerdir (Tigga ve Garg, 2020, s.707). Kamel ve Yaghouzadeh PIDD veri seti kullanarak yaptıkları çalışmada Destek Vektör Makinesi (DVM) kullanmışlar ve diyabet hastalığını 97% doğrulukla tespit etmişlerdir (Kamel ve Yaghouzadeh, 2021, s.2).

## **MATERYAL ve YÖNTEM**

Bu çalışmada Pima Indians Diabetes Database (PIDD) veri seti kullanılmıştır. Veri setinde toplam 768 örnek yer almaktadır. 500 örnek diyabet tanısına sahiptir. Veri setinde her bir örnek için yer alan parametreler yaş, hamilelikler, glikoz, kan basıncı, cilt kalınlığı, insülin, vücut kitle indeksi ve diyabet soyağacı fonksiyonu şeklinde isimlendirilmiş ve nümerik değerler içermektedir. Önerilen sınıflandırıcı mimarisi Visual Studio Community 2017 platformunda Python dili kullanılarak gerçekleştirilmiştir. Önerilen mimariye ilişkin blok diyagram Şekil 1 ile verilmiştir.



**Şekil 1.** Önerilen mimariye ait blok diyagram

Biyomedikal alanda yapılan çalışmalarda araştırmacıların sıklıkla karşılaştıkları bir olumsuzluk sınırlı sayıda veya yetersiz örnek içeren veri setleridir (Chen ve Cao, 2019).

Üretken Çekişmeli Ağ (ÜÇA), sentetik veri üretmek için kullanılan bir derin öğrenme yöntemidir (Goodfellow, vd. 204:27). Sentetik veri üretme teknikleri kullanılarak dengesiz sınıflı (balanssız) veri setlerinde azınlık sınıfını artırmak için kullanılan bir yöntemdir (Chawla vd. 2002:16). Bu çalışmada 268 adet sağlıklı sınıfına ait örnek ÜÇA tekniği ile artırılarak 500 adete yükseltilmiştir. Bu sayede hem veri seti büyütülmüş hem de veri seti balanslanmıştır.

Evrişimsel Sinir Ağı (Convolutional Neural Network, CNN) özellikle görüntü işleme uygulamaları için tasarlanmış derin öğrenme modeli türüdür. Evrişimli katmanları, havuzlama katmanlarını, tam bağlantılı katmanları ve bir sınıflandırma katmanını içerir (Alakuş ve Türkoglu, 2020 s.2). Önerilen mimari Şekil 2 ile verilmiştir.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 8, 512)	1024
activation (Activation)	(None, 8, 512)	0
max_pooling1d (MaxPooling1D)	(None, 4, 512)	0
dropout (Dropout)	(None, 4, 512)	0
batch_normalization (Batch Normalization)	(None, 4, 512)	2048
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 2048)	4196352
dense_1 (Dense)	(None, 1024)	2098176
dense_2 (Dense)	(None, 512)	524800
dense_3 (Dense)	(None, 2)	1026

**Şekil 2.** Önerilen CNN mimarisi

Medikal klinik uygulamalarda çapraz doğrulama tekniği daha az net sonuçlar vermektedir (Alakuş ve Türkoglu, 2020). Bu nedenle bu çalışmada eğitim-test bölme yaklaşımı kullanılmıştır. Bölme oranı 20 olarak belirlenmiştir.

#### Bulgular ve Tartışma

Önerilen mimarinin başarımı aşağıda Eşitlik 1 ve Eşitlik 4 arasında matematiksel ifadeleri verilen metrikler kullanılarak değerlendirilmiştir. Eşitliklerdeki TP doğru olarak sınıflandırılan hasta sayısını, TN doğru olarak sınıflandırılan sağlıklı sayısını, FP sağlıklı ile hasta olarak sınıflandırılan sayısını FN ise hasta iken sağlıklı olarak sınıflandırılanların sayısını ifade etmektedir. AUC (eğri altındaki alan) değerini hesaplamak için alıcı işletim karakteristiği (ROC) eğrisi kullanılır. AUC, rastgele bir hasta ve rastgele bir sağlıklı birey arasında yapılan bir tıbbi test sonucunun doğruluğunu değerlendiren bir ölçüdür. Bu, test sonuçlarının hasta ve sağlıklı bireyler arasında ayırma yeteneğini ölçer. AUC değeri, bu ayırma yeteneğinin ne kadar güçlü olduğunu gösterir; yani, ne kadar yüksekse, testin doğruluğu o kadar iyidir. (Safari,2016:111-113).

$$\text{Doğruluk} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Duyarlılık} = TP / (TP + FN) \quad (2)$$

$$\text{Hassasiyet} = TP / (TP + FP) \quad (3)$$

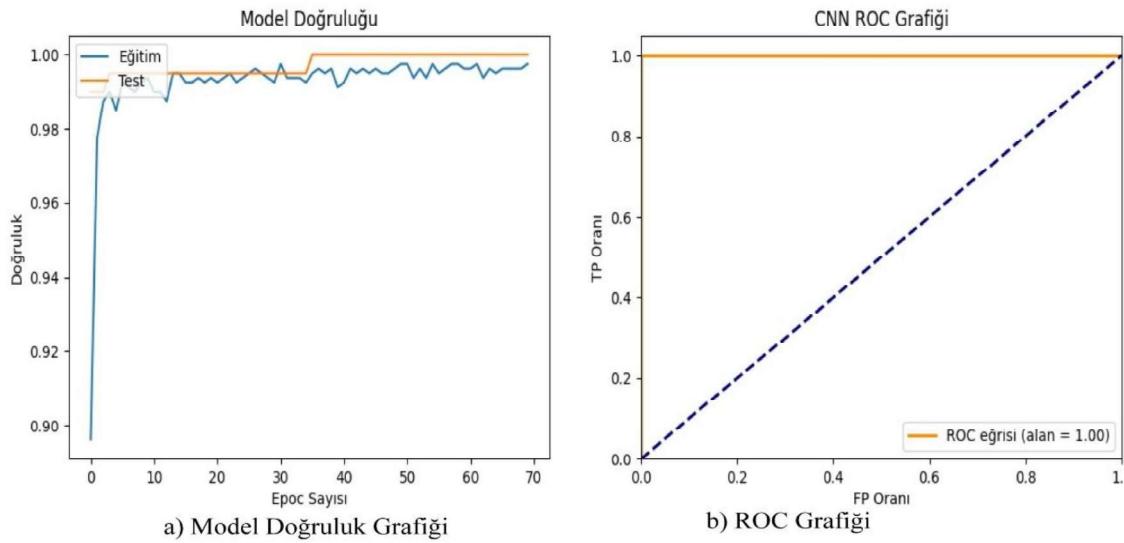
$$\text{F1-skor} = 2 \times \text{Hassasiyet} \times \text{Duyarlılık} / (\text{Duyarlılık} + \text{Hassasiyet}) \quad (4)$$

Bu metriklere göre önerilen derin öğrenmeye dayalı mimarinin ortalama sonuçları Tablo 1 ile verilmiştir.

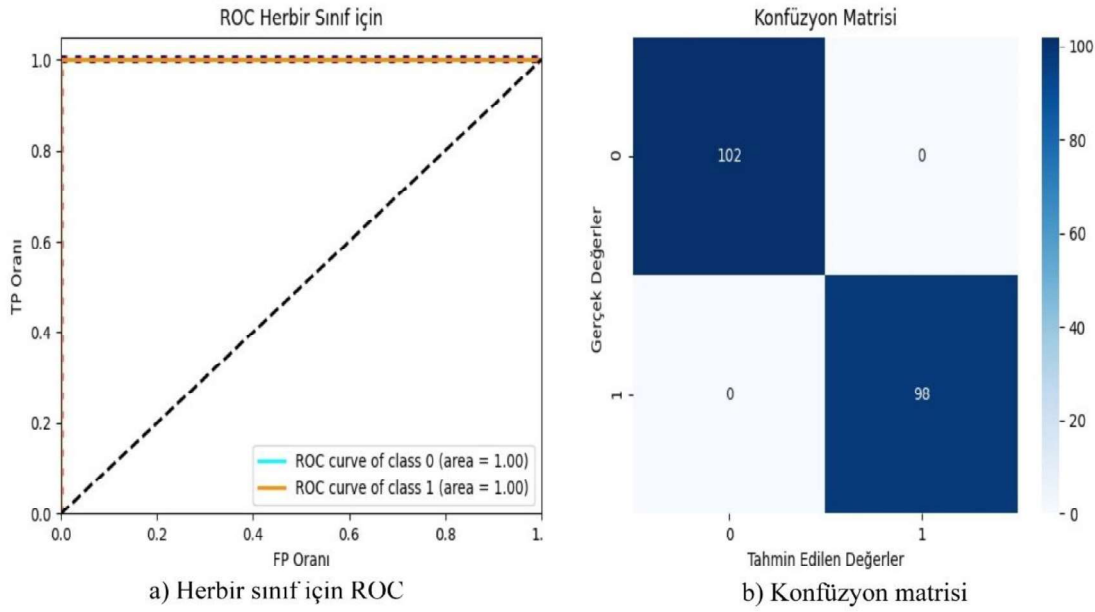
**Tablo 1.** Önerilen mimariden elde edilen performans sonuçları

Doğruluk	Duyarlılık	Hassasiyet	F1-skor	AUC
0,9972	0,9975	0,9975	0,9975	1,0000

Önerilen mimariye ait elde edilen grafiksel sonuçlar Şekil 2 ve Şekil 3 ile verilmiştir.



**Şekil 2.** a) Model doğruluk grafiği b) ROC grafiği



Şekil 3. a) Her bir sınıf için ROC b) Konfüzyon matrisi

Literatürdeki bu çalışmanın konusu ile ilgili yapılan çalışmalardan elde edilen performanslar bu çalışma sonucu ile karşılaştırılarak Tablo 2 ile verilmiştir.

Tablo 2. Literatür karşılaştırması

Referans	Yöntem	Doğruluk %
Sisodia	Naive Bayes	76,30
Yang	K-ortalama/ LR	95,42
Choubey	RTF SA	76,087
Zou	SA	76,67
Tigga	RO	75
Kamel	DVM	97
Bu Çalışma	CNN	99,72

## SONUÇ ve ÖNERİLER

Tablo 2 ile verilen sonuçlar göz önüne alındığında diyabet hastalığının tahmininde bu çalışmada önerilen derin öğrenmeye dayalı sınıflandırıcı mimarisi literatürdeki çalışmalara karşı daha üstün bir başarıya ortaya koymuştur. Önerilen mimari ile biyomedikal alandaki araştırmacıların sıklıkla karşılaştıkları yetersiz ve balansız veri seti problemi giderilebilir. Aynı zamanda derin öğrenmeye dayalı teknikleri kullanacak araştırmacılar için derin öğrenme sistemlerinin başarımının yükseltilmesinde de önemli bir rol oynayabilir. Diyabetle ilgili en büyük zorluk, erken teşhis eksikliği veya hekimlerin yanlış teşhis koyma riskidir. Bu yüzden, özellikle

diyabetin başlangıç aşamalarında doğru bir teşhis sağlayan bir yöntemin uygulanması, hastalığın kontrol altına alınması ve önlenmesi için kritik bir adım olarak öne çıkmaktadır. Bu çalışmada önerdiğimiz yöntemin yüksek doğruluk değerlerine sahip olması nedeniyle bu hastalığın erken teşhisi ve kontrol altına alınmasında hekimlere karar aşamasında yardımcı olabilir. Aynı zamanda diğer yöntemlere göre hem zaman hem de maliyet açısından bir avantaj sağlayabilir.



## **KAYNAKLAR**

- Atkinson MA, Eisenbarth GS, Michels AW. (2014). Type 1 diabetes. *The Lancet*. 383(9911):69-82.
- Egan A.M, Dinneen, S.F. (2004). Diabetes: basic facts What is diabetes?. *Medicine*. 42 (12) 697-681.
- Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. (2011). *Diabetes care*. 34(6):1249-1257.
- Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, Dyer AR, Leiva A, Hod M, Kitzmiller JL, Lowe LP. (2010). International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes care*. 33(3):676-82.
- Ramachandran, A. (2014). Know the signs and symptoms of diabetes. *Indian J Med Res*. 140 (5), 579-581.
- Tigga, N P, Garg S. (2020). Prediction of Type 2 Diabetes using Machine Learning Classification Methods. *Procedia Computer Science* 167: 706–716.
- Anjana, R. M., Pradeepa, R., Deepa, M., Datta, M., Sudha, V., Unnikrishnan, R., Bhansali, A., Joshi, S. R., Joshi, P. P., Yajnik, C. S., Dhandhania, V. K. (2011) . Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research–IndiaDIABetes (ICMR–INDIAB) study. *Diabetologia* 54 (12): 3022-3027.
- Kamel S.R., Yaghoubzadeh R. (2021). Feature selection using grasshopper optimization algorithm in diagnosis of diabetes disease. *Informatics in Medicine Unlocked*. 26:100707.
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I, Chouvarda, I. (2017) Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal* 15: 104-116.
- Sisodia D., Sisodia D.S. (2018). Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*. 132 : 1578–1585.
- Yang, H.W.S., Huang Z, He J., Wang X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*. 10 :100–107.
- Choubey D.K., Paul S. (2017). GA\_RBF NN: a classification system for diabetes. *Int. J. Biomedical Engineering and Technology*, Vol. 23, No. 1. 71-93.
- Zou,Q. Qu,K. Luo,Y.,Yin,D., Ju,Y.Tang,H. (2018). Predicting Diabetes Mellitus With Machine Learning Techniques. *Frontiers in Genetics*. 9: 515.

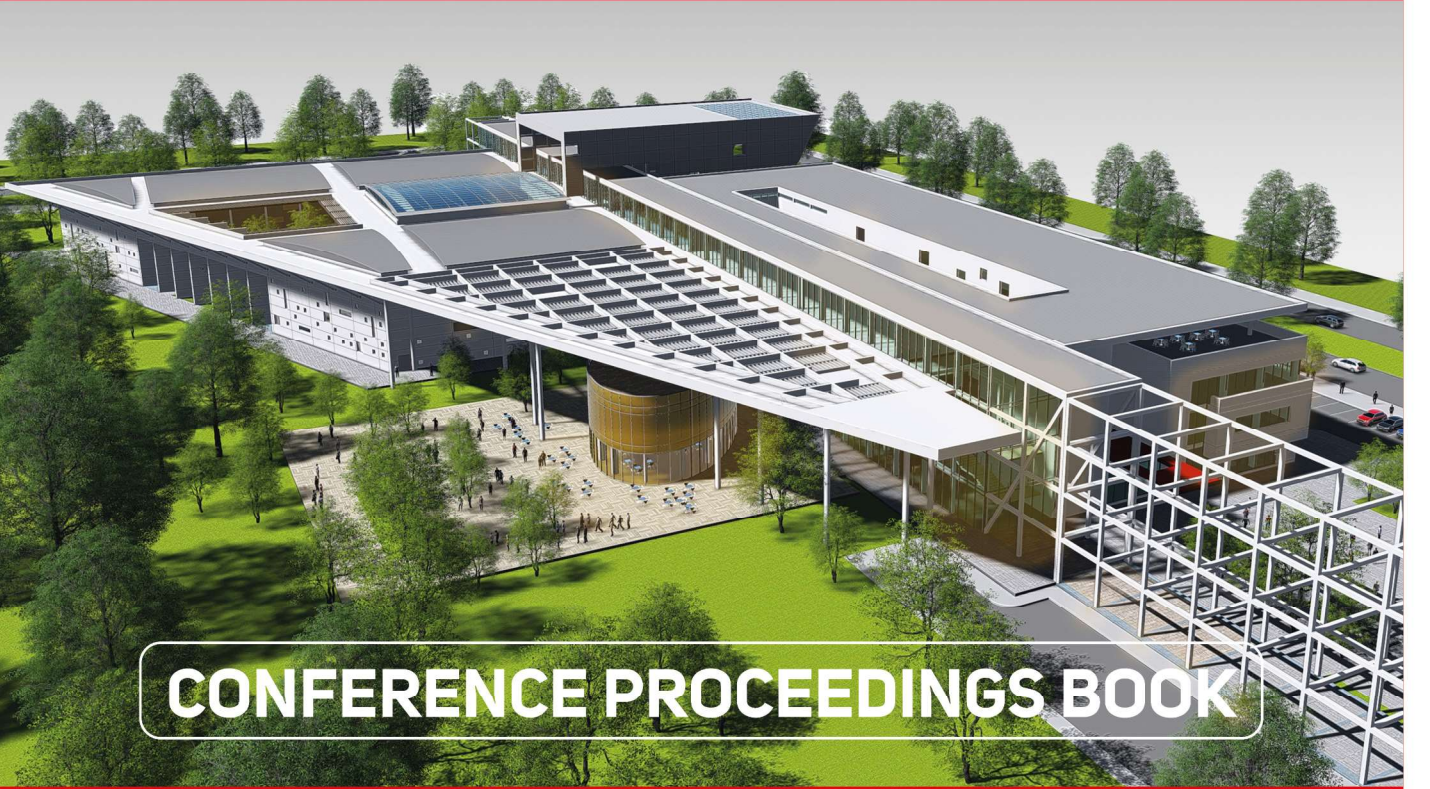
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. ve Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, 27, 2672–2680.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. ve Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16:321–357
- Chen, H., Cao, P. (2019). Deep Learning Based Data Augmentation and Classification for Limited Medical Data Learning. 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). 300-303. Shenyang, China. 12-14 July 2019.
- Alakuş T.B., Türkoglu İ. (2020). Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons and Fractals*. 140:110120.
- Safari, S., A. Baratloo, M. Elfil, A. Negida, Evidence based emergency medicine; part 5 receiver operating curve and area under the curve, *Emergency* 4:(2)111–113, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4893763>.



# SIVAS II. INTERNATIONAL CONFERENCE ON SCIENTIFIC AND INNOVATION RESEARCH

15-17 SEPTEMBER 2023

Sivas University of Science & Technology



**CONFERENCE PROCEEDINGS BOOK**

**Editors:**

Assist. Prof. Dr. Yeter ÇİLESİZ

Assoc. Prof. Dr. Seyithan SEYDOŞOĞLU

**SIVAS**

**II. INTERNATIONAL CONFERENCE ON SCIENTIFIC AND  
INNOVATION RESEARCH**

**DATE – PLACE**  
**September 15-17, 2023**  
**Sivas, Türkiye**

**CONFERENCE  
PROCEEDINGS BOOK**

**EDITORS**

**Assist. Prof. Dr. Yeter ÇİLEŞİZ**  
**Assoc. Prof. Dr. Seyithan SEYDOŞOĞLU**

**All rights of this book belongs to ISPEC Publishing House.**

**Without permission can't be duplicate or copied.**

**Authors of chapters are responsible both  
ethically and juridically.**

**ISSUED: 05/10/2023**

**ISBN: 978-625-367-323-9**

**DOI: <http://dx.doi.org/10.5281/zenodo.8409212>**

## **CONFERENCE ID**

### **CONGRESS TITLE**

**Sivas**

**II. International Conference on Scientific and  
Innovation Research**

### **DATE-PLACE**

**September 15-17, 2023**

**Sivas, Türkiye**

### **ORGANIZATION**

**Sivas University of Science and Technology**

### **ORGANIZING COMMITTEE**

**Prof. Dr. Tolga KARAKÖY - Sivas University of Science and Technology**

**Prof. Dr. Hüsnü Deniz BAŞDEMİR- Sivas University of Science and Technology**

**Prof. Dr. Kağan KÖKTEN - Sivas University of Science and Technology**

**Assoc. Prof. Dr. Emre EVLİCE - Sivas University of Science and Technology**

**Assoc. Prof. Dr. Şabit HOROZ - Sivas University of Science and Technology**

**Assist. Prof. Dr. Gökhan BAKTEMUR - Sivas University of Science and Technology**

**Assist. Prof. Dr. Asuman KAPLAN EVLİCE - Sivas University of Science and  
Technology**

**Assist. Prof. Dr. Yeter ÇİLEŞİZ - Sivas University of Science and Technology**

**Assist. Prof. Dr. Fuat ERDEN - Sivas University of Science and Technology**

**Assist. Prof. Dr. Fikret SAYGIN - Sivas University of Science and Technology**

**Research Assist. Meliha Feryal SARIKAYA - Sivas University of Science and Technology**

**Research Assist. Muhammed TATAR - Sivas University of Science and Technology**

**Research Assist. Ecem KARA - Sivas University of Science and Technology**

**Research Assist. İlker YÜCE - Sivas University of Science and Technology**

**Research Assist. İlhan DANACI - Sivas University of Science and Technology**

### **PARTICIPANTS COUNTRIES**

**Türkiye, Algeria, Azerbaijan, Bulgaria, Germany, India, Indonesia, Iran (Persia),  
Iraq, Kazakhstan, Malaysia, Morocco, Nigeria, Pakistan, Philippines, Romania,  
Saudi Arabia, Serbia, Ukraine, Vietnam**

### **TOTAL ACCEPTED ARTICLE**

**Türkiye: 195**

**Other Countries: 209**