

# YSA Uygulamaları İçin FPGA Tabanlı Softmax Transfer Fonksiyonunun Gerçeklenmesi

İsmail Koyuncu

Düzce Üniversitesi

Kontrol ve Otomasyon Teknolojisi Bölümü  
Uzunmustafa, 81100, Düzce  
e-posta: ismailkoyuncu@duzce.edu.tr

İbrahim Şahin

Düzce Üniversitesi

Teknik Eğitim Fakültesi, Elektronik ve Bilgisayar Eğitimi  
Bölümü, 81620, Konuralp, Düzce  
eposta: ibrahimsahin@duzce.edu.tr

**Özetçe—** Günümüzde yapay sinir ağları kontrol, optimizasyon, tıp, sinyal ve görüntü işleme gibi bir çok alanda kullanılmaktadır. Yapay sinir ağlarında genel olarak doğrusal ve doğrusal olmayan transfer fonksiyonları bulunmaktadır. Doğrusal olmayan transfer fonksiyonları üstel işlemler içerdiğinden bunların geniş değer aralıkları içerisinde donanımsal olarak gerçeklenmeleri oldukça zordur. Bu çalışmada doğrusal olmayan transfer fonksiyonlarından biri olan softmax transfer fonksiyonu FPGA tabanlı olarak tasarlanmıştır. Tasarım bir donanım tanımlama dili olan VHDL'de kodlanmış ve 32-bit IEEE 754-1985 kayan noktalı sayı standardı kullanılmıştır. Çalışmada 2, 3 ve 4 girişli olmak üzere 3 farklı softmax transfer fonksiyonu modellenmiştir. Modellenen transfer fonksiyonu Xilinx'in ISE 13.1 aracı kullanılarak Virtex-6 FPGA çipi için sentezlenmiş ve test edilmiştir. Test sonuçlarına göre tasarlanan transfer fonksiyonunun çalışma frekansı 266.429MHz'dir. Ayrıca tasarım bir milyon veri takımını 4.743ms gibi çok kısa bir sürede hesaplayabilmektedir.

## I. GİRİŞ

Günümüzde Yapay Sinir Ağları (YSA (Artificial Neural Networks)) pek çok alanda yaygın bir şekilde kullanılmaktadır. Bu alanlara sinyal ve görüntü işleme [1, 2] elektrik motorlarının kontrolü [3], optimizasyon [4,5] örnek olarak verilebilir. YSA'ların gerçeklenmesi birkaç farklı donanım ile yapılabilmektedir. Bunlar yazılım tabanlı çalışan bilgisayar programları ve donanımsal yapılardır. Yazılım gerçeklemeleri genellikle yeterli performans verememektedir. Donanımsal olarak YSA'ların gerçeklenmesi için literatürde Application Specific Integrated Circuits (ASIC) [6], Digital Signal Processing (DSP) [7], ve Field Programmable Gate Array (FPGA) [8-10] gibi farklı yapılar bulunmaktadır. ASIC tabanlı uygulamalarda oldukça yüksek performans elde edilmektedir. Ancak ASIC tabanlı uygulamaların en önemli dezavantajı tasarım aşamasının uzun sürmesi ve herhangi bir hata durumunda yapının tekrar kullanılamamasıdır. ASIC tabanlı seri üretimde tasarım sırasında yapılacak bir hata oldukça yüksek maliyet ve uzun zaman kaybına neden olmaktadır. YSA'lar genel yapısından dolayı paralel olarak çalışmakta ancak DSP çipleri ise seri olarak işlem yapmaktadırlar. FPGA sistemleri hem yeniden programlanabilir olmalarından dolayı esnek bir yapıya hem de paralel işlem yapabilme özelliğine sahiptirler. YSA'larda genel olarak doğrusal ve doğrusal olmayan transfer fonksiyonları olmak üzere iki çeşit transfer fonksiyonu bulunmaktadır. Doğrusal olmayan transfer fonksiyonları

üstel işlemler içerdiğinden bunların donanım tabanlı gerçeklenmeleri oldukça zordur. Bu nedenle bu fonksiyonların donanım tabanlı olarak modellenmelerine ihtiyaç duyulmaktadır. Bu amaçla çalışmada doğrusal olmayan transfer fonksiyonlarından biri olan softmax transfer fonksiyonu FPGA tabanlı olarak tasarlanmıştır. Tasarım 32-bit IEEE 754-1985 kayan noktalı sayı standardı kullanılarak çok yüksek hızlı entegre devreler için donanım tanımlama dili olan VHDL (Very High Speed Integrated Circuit (VHSIC) Hardware Description Language) dilinde kodlanmıştır. Çalışmada 2 girişli, 3 girişli ve 4 girişli olmak üzere 3 farklı girişe sahip softmax transfer fonksiyonu modellenmiştir. Tasarımı yapılan transfer fonksiyonu Xilinx'in ISE 13.1 aracı kullanılarak Virtex-6 FPGA çipi için sentezlenmiş ve test edilmiştir.

Bu makalenin ikinci bölümünde FPGA çipleri, YSA'larda kullanılan transfer fonksiyonları ve softmax transfer fonksiyonu hakkında genel bilgiler verilmiştir. Üçüncü bölümde tasarımı yapılan FPGA tabanlı softmax transfer fonksiyon modülü detaylarıyla anlatılmıştır. Dördüncü bölümde yapılan test çalışmaları ve bu çalışmalardan elde edilen veriler sunulmuştur. Son bölümde ise elde edilen sonuçların bir değerlendirilmesi yapılmıştır.

## II. GENEL BİLGİ

### A. FPGA Çipleri

Alanda programlanabilir kapı dizileri olarak tanımlanan FPGA çipleri programlanabilir tümdevrelerdir. Mantıksal fonksiyonları gerçekleştirilmesi amacıyla, kullanıldığı yerde programlanabilir olarak üretilirler. FPGA çipleri genel olarak mantıksal bloklar, giriş-çıkış blokları ve ara bağlantılar olmak üzere programlanabilir üç bileşenden oluşmaktadır. Kullanıcının tasarladığı mantıksal devreye göre, mantıksal bloklar, aralarındaki bağlantılar ve giriş/çıkış blokları programlanabilmektedir.

a. Yapılandırılabilir Mantıksal Bloklar (Configurable Logic Blocks (CLB)) mantıksal fonksiyonların oluşturulabildiği başvuru tablosu (Look-up table (LUT)) ve Flip-Flop'lardan oluşmaktadır. CLB'ler, kullanıcının oluşturmak istediği mantıksal devre için fonksiyonel elemanlar sağlarlar. CLB mimarisinin esnekliği ve simetrisi, uygulamaların kolaylıkla yerleştirilmesine ve gerçeklenmesine olanak tanır.

b. Giriş Çıkış Blokları (Input/Output Blocks (IOB)) çipin iç sinyal hatları ile çipin pinleri arasında programlanabilir arabirim görevini yerine getirirler. IOB'ler sayesinde FPGA

pinleri giriş, çıkış ya da çift yönlü olarak programlanabilir. FPGA çipinin türüne göre bir çipteki IOB sayısı (dolayısıyla pin sayısı) 1000'li sayılara ulaşabilmektedir.

c. Ara Bağlantılar (Interconnections) hem CLB'ler arasında hem de CLB'ler ile IOB'ler arasında bağlantıları yapılandırmada kullanılırlar. Programlanabilir olduklarından çok esnek bir yapıya sahiptirler [11,12].

### B. Softmax Transfer Fonksiyonu

Y.S.A.'larda kullanılan transfer fonksiyonları doğrusal ve doğrusal olmayan transfer fonksiyonları olmak üzere iki grupta toplanabilir. Doğrusal transfer fonksiyonlarına Pure Linear (PureLin), Positive Linear (PosLin), Hard Limiting (HardLim) ve Symmetric Hard Limiting (HardLims) fonksiyonları örnek olarak verilebilir. Doğrusal olmayan transfer fonksiyonlarından bazıları ise Radial Basis (RadBas), Log-Sigmoid (LogSig), Hyperbolic Tangent Sigmoid (Tansig) ve Softmax fonksiyonlarıdır. Aşağıda bu transfer fonksiyonlarının matematiksel denklemleri (1), (2) (3) verilmiştir.

$$RadBas(n) = e^{-x^2} \quad (1)$$

$$LogSig(n) = \frac{1}{1 + e^{-x}} \quad (2)$$

$$TanSig(n) = \frac{2}{(1 + e^{(-x)})} - 1 \quad (3)$$

Yukarıdaki eşitliklerden de görüleceği üzere bu transfer fonksiyonların tek bir giriş ve çıkışı vardır. Bu transfer fonksiyonlarından farklı olarak  $n$  tane giriş/çıkışa sahip olan doğrusal olmayan transfer fonksiyonlarından birisi softmax transfer fonksiyonudur. Softmax transfer fonksiyonun matematiksel denklemi (4) verilmiştir.

$$p_i = \frac{e^{\xi_i}}{\sum_{j=1}^n e^{\xi_j}} \quad (4)$$

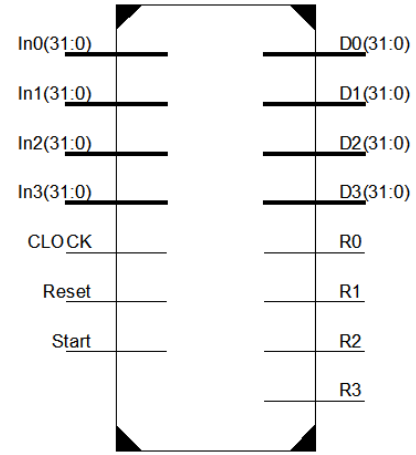
Burada  $p_i$  ağıdaki nöronun çıkışı ve  $j=1$ 'den  $n$ 'e kadar nöron giriş sayısıdır.  $\xi_i$  ise  $i$ 'nci nöronun giriş değeridir. Yukarıdaki matematiksel eşitliklerden de görüleceği üzere doğrusal olmayan transfer fonksiyonları genellikle üstel işlemler içermektedirler. Bu nedenle bu fonksiyonların geniş değer aralıklarında hassas bir şekilde donanımsal olarak hesaplanması oldukça zordur.

### III. SOFTMAX TRANSFER FONKSİYONU MODÜLÜ TASARIMI

Bu çalışmada 32-bit IEEE 754-1985 kayan noktalı sayı (floating-point) standardına uygun, YSA uygulamaları için FPGA sistemleri ile birlikte kullanılacak donanımsal softmax transfer fonksiyonu modülü tasarlanmıştır. Modül 32-bitlik 2, 3 ve 4 adet farklı giriş sayısına sahip olacak şekilde modellenmiştir. Modül bir donanım tanımlama dili

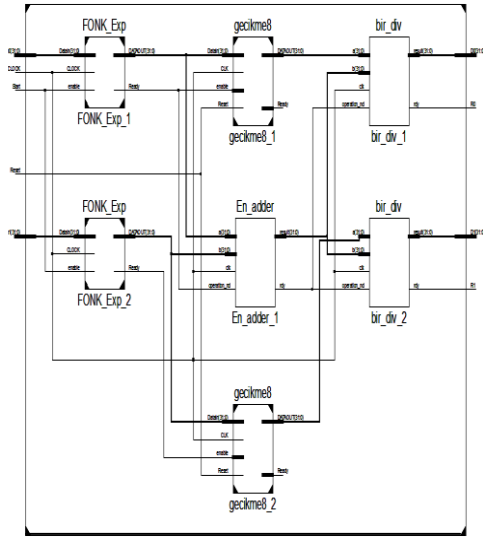
olan VHDL'de kodlanmış ve Xilinx ISE 13.1 aracı kullanılarak Virtx-6 FPGA çipi için sentezlenmiştir. Tasarımda kullanılan çarpıcı, toplayıcı, bölücü ve bazı diğer modüller Xilinx'in IP CORE Generator'ü kullanılarak oluşturulmuştur.

Şekil 1'de tasarımı yapılan 32-bitlik 4 girişli softmax transfer fonksiyonu modülünün en üst seviye blok diyagramı görülmektedir. 32-bitlik 2, 3 ve 4 girişli-çıkışlı olarak tasarımı yapılan modüllerin, karmaşıklığı engellemek amacıyla en üst seviye blok diyagramlarından sadece 4 girişli modülün blok diyagramı verilmiştir. *Reset*, *Basla*, *R0*, *R1*, *R2* ve *R3* sinyalleri modül zamanlaması ve modülün bağlı bulunduğu sistem ile arasındaki senkronizasyonu (hand-shaking) sağlamak için kullanılmaktadır. 32-bit kayan noktalı sayı formatında *In0*, *In1*, *In2* ve *In3* giriş sinyalleri ve *D0*....*D3* ise yine aynı formatta çıkış sinyalleridir.



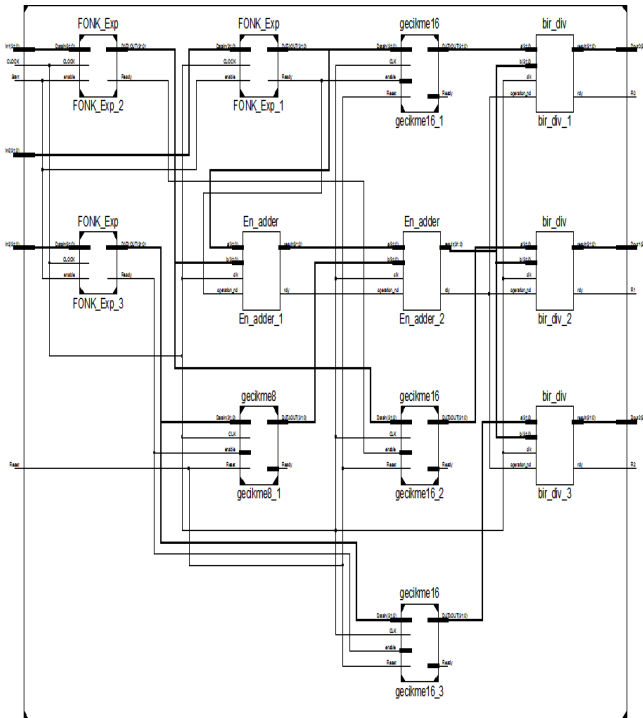
Şekil 1: 4 girişli Softmax transfer fonksiyonu ünitesi en üst seviye blok diyagramı.

Şekil 2'de 2 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı görülmektedir. Modülde kontrol sinyalleri, 2 adet 32-bit giriş sinyali ve 2 adet 32-bit çıkış sinyali bulunmaktadır. Tasarım pipeline olarak çalışabilmektedir. Bu amaçla her bir girişin *Fonk\_Exp* ünitesi ile üsteli alınmakta ve bu değerler *En\_adder* ünitesi ile toplanmaktadır. Sistemin ilk çıkışının elde edilebilmesi için tüm girişlerin üsteli alınarak toplanmış ve ilk girişin üsteli alınarak toplamına bölünmüş olması gerekmektedir. Bu iki işlem arasında zamanlama farkı bulunduğundan sistemin senkron bir şekilde çalışabilmesini sağlamak amacıyla 8 saat darbelik *gecikme8* üniteleri kullanılmıştır. Sistem çıkışları, *gecikme8* ünitesinden çıkan sinyallerin *En\_adder* ünitesinden çıkan sinyallere bölünmesi ile elde edilmektedir.



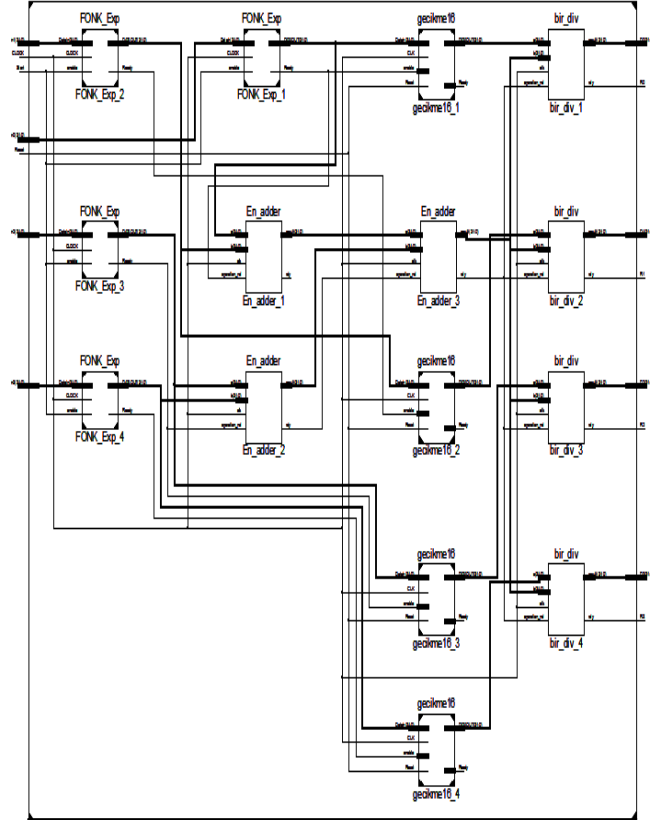
Şekil 2: 2 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı.

Şekil 3'te 3 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı görülmektedir. Modülde 3 adet 32-bit giriş/çıkış sinyalleri ve çeşitli kontrol sinyalleri bulunmaktadır. Tasarımda 3 tane *Fonk\_Exp*, 2 tane *toplama*, 4 tane gecikme ünitesi ve çıkışta 3 tane bölme ünitesi kullanılmıştır. Toplama ünitesi 8 saat darbesinde iki adet girişi toplayabilmektedir. Sistemde 3 giriş olduğundan tüm girişlerin toplanabilmesi için 2 tane toplama ünitesine ihtiyaç duyulmuştur. Bu işlemlerin yapılabilmesi için toplam 16 saat darbesi sürenin geçmesi gerekmektedir. Bu nedenle sistemin senkronizasyonu için bölme işleminden önce *Fonk\_Exp* ünitesi çıkışları 16 saat darbesi gecikme sağlayan *gecikme16* ünitesinden geçirilmiştir.



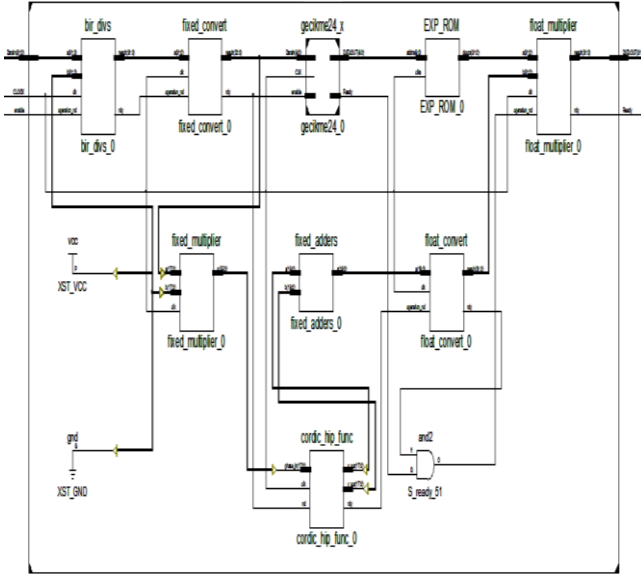
Şekil 3: 3 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı.

Şekil 4'te 4 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı görülmektedir. Modülde, 4 adet 32-bit giriş/çıkış ve çeşitli kontrol sinyalleri bulunmaktadır. Tasarımda 4 tane *Fonk\_Exp*, *gecikme* ve *bölme* ünitesi ve 3 tane *toplama* ünitesi kullanılmıştır. Yine sistemin senkronizasyonun sağlanabilmesi amacıyla 16 saat darbesi gecikme sağlayan *gecikme16* ünitesi kullanılmıştır.



Şekil 4: 4 girişli softmax transfer fonksiyonu ünitesi ikinci seviye blok diyagramı.

Şekil 5'te 4 girişli softmax transfer fonksiyonu ünitesi içerisinde üstel işlemini gerçekleştiren *Fonk\_Exp* ünitesi blok diyagramı görülmektedir. 2 ve 3 girişli softmax transfer fonksiyonu modüllerinde de aynı ünite kullanılmıştır. Ünite COordinate Rotation Digital Computer (CORDIC) ve başvuru tablosu temelli yaklaşımlar birleştirilerek hesaplama yapmaktadır. Bu yolla ünite -48.0 ile +47.25 arasındaki herhangi bir  $x$  reel sayısı için 4-5 hane hassasiyetinde  $e^x$  değerini hesaplayabilmektedir. Ayrıntılı bilgi için bakınız [13].

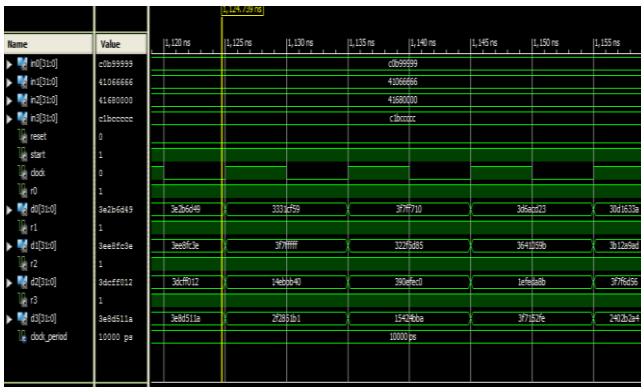


Şekil 5: 4 girişli softmax transfer fonksiyonu ünitesi üçüncü seviye blok diyagramı.

#### IV. TEST SONUÇLARI

Tasarlanan 2, 3 ve 4 girişli softmax transfer fonksiyonu ünitesi, Virtex-6 FPGA çipi için sentezlenerek FPGA çip istatistiklerine bakılmış ve ünitenin maksimum saat frekansları incelenmiştir. Ünitenin belirlenen veriyi işleme süresi, ISE simülasyon programı kullanılarak elde edilmiştir.

Aşağıda şekil 6'da 4 girişli softmax transfer fonksiyonu modülünün FPGA'de gerçekleşmesinden elde edilen Xilinx ISE Simülör ekran sonuçları verilmiştir. Tasarımda 32-bit kayan noktalı sayı standardı kullanılmıştır. Ancak Xilinx ISE Simülör sonuçlarının daha kolay incelenebilmesi amacıyla şekil 6'da hexadecimal notasyonunda gösterilmiştir. 2 girişli modül ilk sonuçları 81 saat darbesi sonunda, 3 ve 4 girişli modüller ise 90 saat darbesi sonunda ilk sonuçları üretmişlerdir. 2 girişli için 81, 3 ve 4 girişli için 90 saat darbesinden önce modülün çıkışlarından herhangi bir çıkış alınamamaktadır. Bundan sonra her saat darbesinde modül çıkışlarından istenilen değerler alınabilmekte ve modül pipeline olarak çalışabilmektedir.



Şekil 6: 4 girişli softmax transfer fonksiyonu ünitesi Xilinx ISE Simülör sonuçları.

Tablo 1'de sentezleme işleminin ardından yapılan yerleştirme (place & route) sonrasında elde edilen FPGA çip istatistikleri verilmiştir. 2, 3 ve 4 girişli tasarımın minimum

darbe periyodu 3.753ns olduğundan bütün tasarımların maksimum frekansı 266.429MHz'dir. Bu sentez sonuçları Xilinx Virtex-6 çip ailesinin en küçük çiplerinden biri olan XC6VCX75T aracında yapılmıştır. Virtex-6 çip ailesinin daha büyük çipleri veya Virtex-7 ailesinin çiplerinden birisi kullanıldığında oldukça düşük çip istatistikleri elde edilecektir.

Tablo 1: FPGA çip istatistikleri.

Softmax Giriş-Çıkış Sayısı	Slice Reg. Sayısı / %	LUTs Sayısı / %	Slice FFs Sayısı / %	Bounded IOBs Sayısı / %	Max Saat Hızı (MHz)
2	8643 / 9	8508 / 18	2132 / 19	133 / 36	266.429
3	12351 / 13	12806 / 27	3278 / 19	198 / 55	266.429
4	16507 / 17	17851 / 38	4423 / 20	263 / 73	266.429

Tablo 2'de tasarımı yapılan modüllerin 100, 10000 ve 1000000 veriyi işleme süreleri verilmiştir. İki girişli modülün tasarımında daha az yapı kullanıldığından pipeline mesafesi kısalmakta ve sonuçları daha kısa süre içerisinde üretmektedir. 3 ve 4 girişli modüllerin verileri işleme sürelerinin aynı olmasının nedeni ise pipeline gecikmelerinin ve çalışma frekansının aynı olmasından kaynaklanmaktadır.

Tablo 2: Tasarlanan modüllerin verileri işleme süreleri.

Veri Sayısı	2 Girişli (µs)	3 Girişli (µs)	4 Girişli (µs)
100	0,4563	0,4653	0,4653
10000	4,644	4,743	4,743
1000000	4644,081	4743,09	4743,09

#### V. SONUÇLAR

FPGA çiplerinin kapasitelerinin ve hızlarının artması bu çiplerin pek çok alanda kullanılabilirliğini arttırmıştır. Bu uygulama alanlarından birisi de YSA'lardır. YSA'larda doğrusal ve doğrusal olmayan transfer fonksiyonları bulunmaktadır. Doğrusal olmayan transfer fonksiyonları genellikle üstel işlemleri içerdiğinden bunların geniş değer aralıkları içerisinde donanımsal olarak gerçekleşmeleri oldukça zor olmaktadır. Bu çalışmada doğrusal olmayan transfer fonksiyonlarından biri olan softmax transfer fonksiyonu FPGA tabanlı olarak tasarlanmıştır. Tasarım bir donanım tanımlama dili olan VHDL'de kodlanmıştır. Tasarımda, Virtex-6 FPGA çipi ve 32-bit IEEE 754-1985 kayan noktalı sayı standardı kullanılmıştır. Çalışmada 2, 3 ve 4 giriş-çıkışlı olmak üzere 3 tane softmax transfer fonksiyonu ünitesi modellenmiştir. Yazılan kod Xilinx'in ISE 13.1 aracı kullanılarak Virtex-6 FPGA çipi için sentezlenmiş ve test edilmiştir. Sonuçlara göre tasarlanan ünitelerin hepsinin çalışma frekansı 266.429MHz'dir. Ayrıca tasarımı yapılan 2 girişli softmax transfer fonksiyon ünitesi bir milyon veri takımını 4.644 ms'de, 3 ve 4 girişli softmax transfer fonksiyon üniteleri ise aynı sayıdaki veri takımını 4.743ms gibi çok kısa bir sürede

hesaplayabilmektedir. İleriki çalışmalarda sistemin boyutunu azaltmak amacıyla çeşitli optimizasyon çalışmaları yapılabilir ve ünitenin 64-bit versiyonu tasarlanabilir. Ayrıca diğer doğrusal olmayan transfer fonksiyonları FPGA’de modellenerek sentezlenebilir.

#### VI. KAYNAKÇA

- [1] Ö. Polat, T. Yıldırım, “FPGA implementation of a General Regression Neural Network: An embedded pattern classification system”, *Digital Signal Processing*, vol. 20, pp. 881–886, 2010.
- [2] J. Huang, J. Lee, Y. Ge, , "An array-based scalable architecture for DCT computations in video coding," *Neural Networks and Signal Processing, 2008 International Conference on* , vol., no., pp.451-455, 7-11 June 2008
- [3] Q. N. Le, J. W. Jeon, "Neural-Network-Based Low-Speed-Damping Controller for Stepper Motor With an FPGA", *IEEE Transactions on Industrial Electronics*, vol. 57, no. 9, 2010.
- [4] F. J. Lin, Y. C. Hung, “FPGA-Based Elman Neural Network Control System for Linear Ultrasonic Motor”, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 1, 2009.
- [5] C. J. Lin, H. M. Tsai, “FPGA implementation of a wavelet neural network with particle swarm optimization learning”, *Mathematical and Computer Modelling*, vol. 47, no. 9-10, pp. 982-996, 2008.
- [6] O. Banimelhem, R.B. Hani, "Neural network based optimization of CMOS transistor sizing for leakage power minimization," *Innovations in Information Technology (IIT), 2012 International Conference on* , vol., no., pp.167-172, 18-20 March 2012.
- [7] R. Rieger, S. Deng, S, "Double-Differential Recording and AGC Using Microcontrolled Variable Gain ASIC," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* , vol.PP, no.99, pp.1, 2012.
- [8] N. Kim, N. Kehtarnavaz, M. B.Yeary, S. Thornton, “DSP-Based Hierarchical Neural Network Modulation Signal Classification”, *IEEE Transactions on Neural Networks*, vol. 14, no. 5, 2003.
- [9] B. Yu, R. Chan, T. Mak, Y. Sun, C. Poon, C, "On-Chip Systolic Networks for Real-Time Tracking of Pairwise Correlations between Neurons in a Large-Scale Network," *Biomedical Engineering, IEEE Transactions on* , vol.PP, no.99, pp.1, 2012.
- [10] I. Sahin, I. Koyuncu, “Design and Implementation of Neural Networks Neurons with RadBas, LogSig, and TanSig Activation Functions on FPGA”, *Electronics and Electrical Engineering*”, no. 5(121), 2012.
- [11] I. Sahin, “A 32-bit floating-point module design for 3D graphic transformations”, *Scientific Research and Essays*, vol. 5 (20), pp. 3070-3081, 2010.
- [12] M. Papadonikolakis, C. Bouganis, "Novel Cascade FPGA Accelerator for Support Vector Machines Classification," *Neural Networks and Learning Systems, IEEE Transactions on* , vol.23, no.7, pp.1040-1052, July 2012.
- [13] I. Koyuncu, I. Sahin, “A CORDIC Based ex Calculator Unit Design for FPGA Chips”, *e-Journal of New World Sciences Academy*, vol. 6, no 4, 2011.