ÖNSÖZ

2004 yılından bu yana düzenlenen Akıllı Sistemlerde Yenilikler ve Uygulamaları Konferansı, **15-17 Ekim 2020** tarihinde Marmara Üniversitesi ev sahipliğinde çevrimiçi olarak düzenlenmiştir. Konferansın ana temasını oluşturan Akıllı Sistemler ve Yapay zeka kavramları günümüzde artık sadece Elektrik-Elektronik veya Bilgisayar Mühendisliği alanlarında değil; Tıp, Finans, Politika, Spor, Hukuk ve hizmet sektörleri gibi birbirinden çok farklı disiplinlerde kendine uygulama alanı bulabilmektedir. Ulusal bazda kendimize özgü yapay zeka stratejilerimizin planlanmasına ve hayata geçirilmesine ışık tutacak; akademik ve sektörel bazda yeni işbirliklerine kapı açacak bir konferans olması amacı ile Yıldız Teknik Üniversitesi'nin desteği ile düzenlediğimiz ASYU 2020 Konferansında **13 farklı tematik alanda, toplam 103 bildirinin sunumu** Türkçe veya İngilizce olarak gerçekleştirilmiştir.

Tüm bildiriler IEEE Xplore sayısal kütüphanesi kapsamında konferans kitapçığı şeklinde yayınlandıktan sonra; TRDizin'de taranan **International Journal of Advances in Engineering and Pure Sciences** isimli dergide bir Özel sayı çıkarılması planlanmış ve bu sayıda yer alabilecek en iyi bildiriler belirlenmiştir. Bu aşamada konferans yönetim programında bildiriler, hakemlerden aldıkları puanlara göre sıralanmış; ayrıca Düzenleme Kurulu tarafından yeniden gözden geçirilmiştir. Ardından ilgili yazarlara davet gönderilmiş ve davet giden bildiri yazarlarından %80 oranında olumlu cevap alınmıştır. Bunun üzerine yazarlara makalelerini genişletirken sadece literatür araştırması ve referanslar gibi unsurlarla değil; aynı zamanda kullanacakları yeni teknikler, algoritmalar, veri kümeleri ve bunlara ilişkin farklı deneysel ve benzetim sonuçları gibi önemli ölçüde içeriği artırmaları şeklinde bilgi verilmiştir. Yazarlardan genişletilmiş çalışmalarını Dergi yönetim platformu olan Dergipark'a yüklemeleri istenmiş ve tüm makaleler yeniden bir kör hakemlik değerlendirme sürecinden geçirilmiştir. IEEE Xplore konferans kitapçığında yer alan bildiriler ile genişletilmiş hali arasında %40'tan daha büyük bir benzerlik skoruna izin verilmemiştir. Sonuç olarak hakemlerden kabul alan 10 adet makale ile bu özel sayı oluşturulmuştur.

İşbirliklerinden ötürü tüm yazarlara ve değerlendirme sürecindeki özverili çalışmalarından dolayı tüm hakemlere teşekkür ederiz.

Saygılarımızla

Prof. Dr. Tülay YILDIRIM
ASYU Konferansları Genel Başkanı
Yıldız Teknik Üniversitesi

Prof. Dr. Hayriye KORKMAZ
Dergi Editörü/Konferans Eş Başkanı
Marmara Üniversitesi

Özel Sayı Alan Editörü
Dr. Öğr. Üyesi Zehra Aysun ALTIKARDEŞ
Marmara Üniversitesi

Özel Sayı Alan Editörü
Dr. Öğr. Üyesi Erkan DURSUN
Marmara Üniversitesi

Yayın Editörü
Arş. Gör. S. Enes HACIBEKTAŞOĞLU
Marmara Üniversitesi

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Simulation Studies for Motion Control of Multiple Biohybrid Microrobots in Human Synovial Fluid with Discontinuous Reference Signals

Süreksiz Referans Sinyalleri ile İnsan Sinovyal Sıvısında Birden Fazla Biyohibrit Mikrorobotun Hareket Kontrolü için Benzetim Çalışmaları

Ahmet Fatih TABAK[1] ORCID

[1] *Kadir Has University, Mechatronics Engineering, 34083 Istanbul, Turkey*

Abstract

It is envisioned that biomedical swarms are going to be used for therapeutic operations in the future. The utilization of a single robot in live tissue is not practical because of the limited volume. In contrast, a large group of microrobots can deliver a useful amount of potent chemicals to the targeted tissue. In this simulation study, a trio of magnetotactic bacteria as a task-force, Magnetospirillum *Gryphiswaldense* MSR-1, is maneuvered via adaptive micro-motion control through an external magnetic field. The magnetic field is induced by a single permanent magnet positioned by an open kinematic chain. The coupled dynamics of this small group in the human synovial tissue is simulated with actual magnetic and fluidic properties of the synovial liquid. The common center of mass is tracked by the equation of motion. The overall hydrodynamic interaction amongst all three bacteria is modeled within a synovial medium confined with flat surfaces. A bilateral control scheme is implemented on top of this coupled model. The position of the common center of mass is used as the reference point to the end-effector of the robotic arm. The orientation of the magnetic field is rotated to change the heading of the bacterial-group in an addressable manner. It has been numerically observed that controlling the common swimming direction of multiple bacteria is fairly possible. Results are presented via the rigid-body motion of the robotic task-force as well as the fluidic and magnetic force-components acting on the bacteria along with the bilateral control effort in all axes.

Keywords: Micro-Motion Control, Synovial Fluid, Magnetotactic Bacterium, Multiscale Robots, Bilateral Control, Adaptive Control

Öz

Gelecekte mikro robotik sürülerin medikal operasyonlar için kullanılması öngörülmektedir. Canlı dokuda tek bir mikro robotun kullanılması, sınırlı hacim nedeniyle pratik değildir. Ancak, kalabalık bir mikro robot grubu, hedeflenen dokuya yararlı miktarda faydalı kimyasallar iletebilir. Bu simülasyon çalışmasında, bir görev gücü olarak seçilen üç manyetotaktik bakterinin (Magnetospirillum *Gryphiswaldense* MSR-1) harici bir manyetik alan aracılığıyla adaptif mikro-hareket kontrol performansı araştırılmıştır. Manyetik alan, üç serbestlik dereceli açık bir kinematik zincir tarafından konumlandırılan tek bir doğal Neodimyum mıknatıs yardımı ile oluşturulur. Açık kinematik zincirin her ekseninde adanmış bir doğru akım (DC) motoru bulunmaktadır. İnsan sinovyal ekleminde hareket eden bu küçük bakteri grubunun katı cisim dinamikleri, sinovyal sıvının gerçek manyetik ve akışkan özellikleri üzerinden simüle edilir. Mikro robotların ortak kütle merkezi, hareket denklemi ile izlenir. Üç bakteri arasında hareket sırasında ortaya çıkan çapraz hidrodinamik etkileşim, sinovyal sıvı sınırları içinde modellenmiştir. Ortaya çıkan sistem dinamiklerinin üstüne çift yanlı bir adaptif kontrol yaklaşımı ile referans sinyali uygulanmaktadır. Ortak kütle merkezinin konumu, açık kinematik zincirin uç efektörüne referans noktası olarak geri beslenmektedir. Manyetik alanın yönü, bakteri grubunun yönünü adreslenebilir bir şekilde değiştirmek için kullanılmaktadır. Üç bakteriden oluşan bu görev gücünün ortak yüzme yönünü, sinovyal sıvı içerisinde ayrık kontrol sinyalleri ve adaptif kontrol çalışması ile idare etmenin kısmen mümkün olduğu, farklı referans fonksiyonları yardımı ile, sayısal olarak gözlenmiştir. Sonuçlar, robotik görev gücünün katı cisim hareketi ve tüm eksenlerde mevcut çift taraflı kontrol çabasına ek olarak bakterilere etki eden akışkan ve manyetik kuvvet bileşenleri aracılığıyla sunulur.

Anahtar Kelimeler: Mikro-Hareket Kontrolü, Sinovyal Sıvı, Magnetotaktik Bakteri, Çok Ölçekli Robotlar, Çift Yanlı Kontrol, Adaptif Kontrol

I. INTRODUCTION

The field of biomedical micro-robotics is perceived as a promising interdisciplinary field of medical science, robotic applications, and microtechnology drawing much attention for the past two decades [1–10]. The envisioned tasks for biomedical microrobots include kidney stone destruction, removal of blood-clots, targeting cancerous cells, and carrying potent chemicals to predefined locations [11,12]. The materials and approaches used include manufacturing artificial swimmers using organic materials [13] and inorganic materials [14,15], incorporating live cells into biohybrid systems [16,17], and directly employing self-propelling single-celled organism within the robotic system [18,19]. One very important benefit of the latter approach is biocompatibility for biomedical applications.

The biological well-being of the living organism is of the utmost importance. Therefore, the robotic device should be able to carry out the therapeutic task without any unforeseen complications in the living tissue. The strategy of relying on biocompatibility is the way to reduce the possibility of the immune system attacking the micro-robotic device or avoiding the possible toxic effects of the materials used in micro and nano-manufacturing processes [20–22]. The use of patient-based organic materials [23] and modified live cells [24], amongst other alternative solutions [25,26], are devised to address this practical but crucial problem. The use of natural cells in a robotic system causes a paradigm shift since power supply and motion control turn out to be detached issues as opposed to fully artificial systems. Furthermore, there are recent studies on micro-motion control in synovial liquid using biocompatible microrobots of different origins [27,28], most of which focus on incorporating magnetotactic bacteria in the complex medium.

In addition, groups of micro swimmers are being considered for the therapeutic applications [29–35]. Single microrobots are not capable of carrying substantial amounts of potent chemicals, nor on-boards sensory equipment can be incorporated. Thus, tracking and control of a single microrobot stands as a challenge whereas a large group can be tracked relatively more easily. Moreover, it will be possible to employ individual groups for different tasks simultaneously [36–38]. Finally, if the swarm is composed of biohybrid robots, energy supply problem could arguably be partially solved. The aforementioned reasons make swarm research more appealing for future studies.

There are several control studies in the literature demonstrating the performance of magnetotactic bacteria species Magnetospirillum *Gryphiswaldense* MSR-1 via electromagnetic (EM) coils and permanent magnets. The former approach demonstrating non-adaptive control effort on the swimming direction of real bacterium samples [39,40] while the latter investigating the adaptive control approach on the simulated bacterium and environment with continuous or discontinuous reference signals [28, 41–45]. In this study, a rudimentary control scheme for a biohybrid robotic system is numerically demonstrated with the help of a discontinuous set point yaw-angle reference using basic ramp and step functions. The reported studies in the literature are based on *in vitro* swimming conditions of *M. Gryphiswaldense* MSR-1 in an aqueous medium [39–45] and *in vivo* performance in the synovial liquid [28], however, acting as a single microrobot in the fluidic medium.

In this study, a swarm of three Magnetospirillum *Gryphiswaldense* MSR-1 is being controlled as a task force in the synovial liquid offering a novel analysis to the literature of biomedical micro-robotic research. The model and simulation strategy is based on previous studies [28,41] by expanding the analysis to swarms. Here, a detailed mathematical model representing the coupled system of an open kinematic chain, the task force of three magnetotactic bacteria, and one permanent magnet is cast to simulate the rigid-body and control dynamics of all the systems. Adaptive integral gains are incorporated in the proportional-integral-derivative (PID) control scheme. The control scheme is constructed to be bilateral; resulting in a two-way coupled system. Furthermore, two time-dependent reference functions are used to rotate the bacteria in clockwise and counter-clockwise directions, on separate occasions. The *Methodology* section discusses the details on the mathematical model of the system and the adaptive control followed by the *Results* section presenting the performance of the control algorithm and the prospect of controlling multiple bacteria *in vivo*.
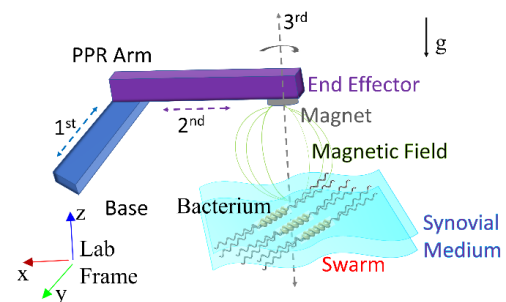
II. METHODOLOGY

Figure 1. PPR arm, the swarm comprised of three neutrally buoyant magnetotactic bacteria, the magnet at the end of the open kinematic chain, the magnetic field between the magnet and the swarm, the synovial medium of a certain thickness, gravitational attraction.

Figure 1 depicts the robotic system with the open kinematic chain, i.e. the prismatic-prismatic-revolute (PPR) robotic arm, the neutrally-buoyant group of bacteria, i.e., three hydrodynamically coupled Magnetospirillum *Gryphiswaldense* MSR-1 [39,40], and one $0.02 \times 0.02 \times 0.02$ m^3 permanent magnet, i.e., the N52-grade Neodymium [46] block, located at the end-effector of the robotic arm. The robotic arm is oriented in the workspace such that the magnet can traverse along the xy-plane and revolve along the z-axis in the lab frame. Each joint of the PPR arm is articulated by a dedicated brushless DC-motor [47]. The first two links of the PPR arm are selected to be aluminum prisms of $5 \times 5 \times 30 \times 10^{-6}$ m^3. The axes of the arm do not interact with gravitational attraction and are assumed to be rigid enough to withstand deformation along z-axes against own weight. The motion of the end-effector is assumed to be undisturbed by possible impurities in the construction of the open kinematic chain.

Each bacterium cell has two helical bundles consisting of two flagella each, articulated on the opposite sides

of its cell body [39,40]. The magnetic field penetrates the synovial liquid. As the magnet rotates, the resultant time-dependent magnetic field or the resultant electromagnetic field (EMF in Figure 2) exerts the torque necessary to change the yaw-angle of the swarm. The swarm propels itself fully submerged in the synovial liquid at body temperature.

In this study, each bacterium is simulated as a separate robot. They are far enough from each other so that they can swim in formation. Therefore, each bacterium has its equation of motion given as:

$$
\left[\begin{array}{l} \mathbf{F}_p(t) + \mathbf{F}_m(t) + \mathbf{F}_d(t) + \sum_{j,k} \mathbf{F}_{sw}(t) + \mathbf{F}_c(t) \\ \mathbf{T}_p(t) + \mathbf{T}_m(t) + \mathbf{T}_d(t) + \sum_{j,k} \mathbf{T}_{sw}(t) + \mathbf{T}_c(t) \end{array} \right]_i = \mathbf{0}
\tag{1}
$$

with subscripts 'p', 'm', 'd', 'sw', and 'c' denoting the total propulsive effect of the tails (Equation (2)), the stimulation of the Neodymium magnet (Equation (3), the viscous drag on the surface of the bacterium (Equation (4)), the hydrodynamic cross-coupling with the other two bacteria in the swarm (Equation (5)), and the contact with the solid boundary (Equation (6)), respectively. Also, the subscripts 'i', 'j', and 'k' denote the individual bacterium in the swarm.

The propulsive effect is modeled by integrating the local resistive force acting on the tails as:

$$
\left[\begin{array}{l} \mathbf{F}_p \\ \mathbf{T}_p \end{array} \right] =
$$

$$
\left(2\int_0^{L_1} \left[\begin{array}{cc} \mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}} & -\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}}\mathbf{S}_1 \\ \mathbf{S}_1\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}} & -\mathbf{S}_1\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}}\mathbf{S}_1 \end{array} \right] d\ell \right) \left[\begin{array}{c} \mathbf{0} \\ \mathbf{\Omega}_{tail_1} \end{array} \right]
$$
$$
+ \left(2\int_0^{L_2} \left[\begin{array}{cc} \mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}} & -\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}}\mathbf{S}_2 \\ \mathbf{S}_2\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}} & -\mathbf{S}_2\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}}\mathbf{S}_2 \end{array} \right] d\ell \right) \left[\begin{array}{c} \mathbf{0} \\ \mathbf{\Omega}_{tail_2} \end{array} \right]
\tag{2}
$$

Equation (2) calculates the 6-degrees-of-freedom propulsion force and torque vectors [48,49] with the help of 3×3 matrices; $\mathbf{R}$ giving the rotation matrix for local Frenet-Serret coordinates, $\mathbf{S}$ representing the local cross-products, and $\mathbf{C}$ containing the local fluid resistance coefficients for the tails.

The magnetic force and torque vectors on each bacterium cell are denoted by:

$$
\left[\begin{array}{l} \mathbf{F}_m \\ \mathbf{T}_m \end{array} \right] = \left[\begin{array}{l} (\mathbf{m}\cdot\nabla)(\mathbf{R}_{mag}\mathbf{B}) \\ \mathbf{m}\times(\mathbf{R}_{mag}\mathbf{B}) \end{array} \right]
\tag{3}
$$

with $\mathbf{m}$ signifying the magnetic moment possessed by the bacterium [40], $\mathbf{B}$ is the magnetic field felt by the bacterium, i.e., $\mathbf{B} = [B_x\ B_y\ B_z]^{\mathrm{T}}$, and $\mathbf{R}_{mag}$, is the rotation matrix between the bacterium and the end effector of the arm. Furthermore, the components of $\mathbf{B}$ are calculated based on the position of each bacterium, i.e., $[x\ y\ z]^{\mathrm{T}}$ relative to the position of the magnet in

the lab frame [41]. Thus, each bacterium is under the influence of different magnetic fields and field-gradients.

The viscous drag acting against the resultant translational and rotational rigid-body velocity, i.e., $[\mathbf{U}\ \mathbf{\Omega}]^{\mathrm{T}}_{sw}$, is given in Equation (6):

$$
\left[\begin{array}{l} \mathbf{F}_d \\ \mathbf{T}_d \end{array} \right] = - \left\{ 2\int_0^{L_1} \left[\begin{array}{cc} \mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}} & -\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}}\mathbf{S}_1 \\ \mathbf{S}_1\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}} & -\mathbf{S}_1\mathbf{R}_1\mathbf{C}_1\mathbf{R}_1^{\mathrm{T}}\mathbf{S}_1 \end{array} \right] d\ell \right.
$$
$$
+ 2\int_0^{L_2} \left[\begin{array}{cc} \mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}} & -\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}}\mathbf{S}_2 \\ \mathbf{S}_2\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}} & -\mathbf{S}_2\mathbf{R}_2\mathbf{C}_2\mathbf{R}_2^{\mathrm{T}}\mathbf{S}_2 \end{array} \right] d\ell
$$
$$
\left. + \int_0^{L_{body}} \left[\begin{array}{cc} \mathbf{R}_b\mathbf{C}_b\mathbf{R}_b^{\mathrm{T}} & -\mathbf{R}_b\mathbf{C}_b\mathbf{R}_b^{\mathrm{T}}\mathbf{S}_b \\ \mathbf{S}_b\mathbf{R}_b\mathbf{C}_b\mathbf{R}_b^{\mathrm{T}} & -\mathbf{S}_b\mathbf{R}_b\mathbf{C}_b\mathbf{R}_b^{\mathrm{T}}\mathbf{S}_b \end{array} \right] d\ell \right\} \left[\begin{array}{c} \mathbf{U} \\ \mathbf{\Omega} \end{array} \right]_{sw}
\tag{4}
$$

with $\mathbf{S}_b$ being the skew-symmetric matrix for the center of volume of the body of the bacterium.

The hydrodynamic cross-coupling effect, i.e., the hydrodynamic interaction (HDI in Figure 2), is calculated by [50]:

$$
\left[\begin{array}{l} \mathbf{F}_{sw}(t) \\ \mathbf{T}_{sw}(t) \end{array} \right] = \left[\begin{array}{c} \dfrac{100 e^{-100h_p\,t}}{1 - e^{-100h_p\,t}} \mathbf{d}_p(t) \\ \mathbf{p}_d(t)\times\mathbf{F}_s(t) \end{array} \right]
\tag{5}
$$

with $h_p(t)$ being the proximity between the opposite surfaces of two bacteria, the vector $\mathbf{d}_p(t)$ denoting the common normal between the closest points on the opposite surfaces, and $\mathbf{p}_p(t)$ giving the location of the closest point on the surface in the inertial frame of each bacterium [50].

The contact force (CF in Figure 2) acting on each bacterium cell should a collision with the solid boundary occurs is modeled based on the penalty approach [51] as:

$$
\left[\begin{array}{l} \mathbf{F}_c \\ \mathbf{T}_c \end{array} \right] =
$$
$$
\left[\begin{array}{c} \left(k \begin{array}{l} \delta \Leftarrow \delta <= 0 \\ 0 \Leftarrow \delta > 0 \end{array} + b \begin{array}{l} d\delta/dt \Leftarrow d\delta/dt > 0 \\ 0 \Leftarrow d\delta/dt <= 0 \end{array} \right) \mathbf{n}_c \\ \mathbf{S}_c\mathbf{F}_c \end{array} \right]
\tag{6}
$$

with necessary conditions imposed associated with the girid-body motion in lab frame.

In Equation (6), δ is the penetration depth, k and b represent the stiffness and damping at the point of

contact, respectively, where exists a surface normal, $\mathbf{n}_c$, employed to calculate the direction of contact force. Finally, $\mathbf{S}_c$ denotes the skew-symmetric matrix for the position of the point of contact in the inertial frame of the bacterium.

It must be noted that the matrices $\mathbf{C}_{\{1,2,b\}}$ in Equations (2) and (4) are diagonal [41–45] and contain the viscosity of the synovial fluid, μ, that is responsible for the lubrication of the synovial joint and exhibits non-Newtonian, i.e., shear-thinning, behavior [52,53]. The shear-thinning effect of the viscosity is given by:

$$\mu = \ 0.4814(2\pi f)^{0.385} - 0.07142 \ / 2\pi f \tag{7}$$

with f being the rotation frequency of the helical tails, 471.2 rad/s as reported [40].

The dynamics of the PPR arm [51] can be presented by Equation (8) given that the proposed open kinematic chain is not susceptible to gravity, Coriolis, and centrifugal forces due to its simple design and orientation in the lab frame:

$$\mathbf{D\ddot{q}} = \mathbf{K}_m\mathbf{I}_m \tag{8}$$

with $\mathbf{D}$ being the mass matrix of the PPR-arm [41]. The vector $\mathbf{q}$ denotes the generalized coordinates, $\mathbf{q} = [d_1(t) \ d_2(t) \ \theta_3(t)]^T$ along which the motion of the respective joint takes place. The Denavit-Hartenberg table [51] representing the link constants and joint variables is given in Table 1. Finally, $\mathbf{I}_m$ is the diagonal matrix of the motor currents leading to coupled electromechanics for the robotic arm as the motor current also appears in the electromechanical equations of the DC-motors.

Table I. Denavit-Hartenberg table of the PPR open kinematic chain arm depicted in Figure 1.

Link/ Joint	Twist angle (rad)	Link length (m)	Offset (m)	z-Rotation (rad)
1st	$-\pi/2$	0	$d_1(t)$	0
2nd	$-\pi/2$	0	$d_2(t)$	$\pi/2$
3rd	0	0	0.02	$\theta_3(t)$

Each one of the dedicated DC-motors is represented by a mathematical model of the form [51]:

$$L_m\dot{I}_m + R_mI_m = V_m - K_b\dot{q}_m \tag{9}$$

with I_m standing for the motor current, V_m denoting the applied control voltage, L_m and R_m representing the inductance and resistance of the DC-motor, respectively, and K_b being the back-emf constant of the armature. All motor specifications used in this study are of the Maxon EC 45 Flat brushless 48 V & 70 W DC-motor [47]. The brushless DC-motors are

selected based on the previous numerical simulation studies [28, 41–45].

Equation (8) and Equation (9) are solved together to find the acceleration of each joint and the corresponding DC-motor currents as a linear system of coupled equations. Time derivatives of link velocities and motor currents are determined following the respective control input, V_m, while satisfying the electromechanical constraints, e.g., upper limits of continuous current and velocity [47]. Position in the lab frame and the currents are predicted via simple integration over time.
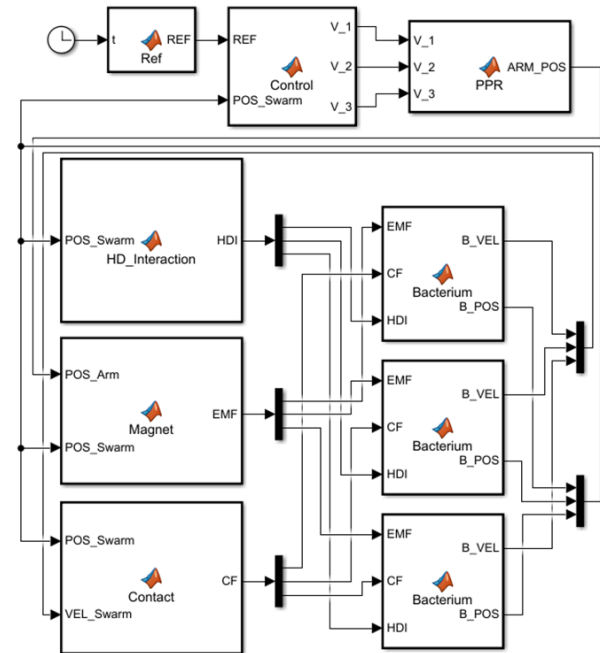

Figure 2. The control block (Control) generating the voltage values to the DC-motors on the PPR arm (PPR) based on the reference (REF) and swarm position (POS_Swarm). The arm position (ARM_POS) the magnet (Magnet), Magnet generates the field on each bacterium (Bacterium), each Bacterium interacts with each other (HD_Interaction) and the surrounding boundary (Contact). The position and velocity of each bacterium are denoted by B_POS and B_VEL, respectively.

The control loop and the system, which are constructed in MATLAB/SIMULINK environment, are depicted in Figure 2. The instantaneous position of each robot is averaged and used as a position reference to the PPR arm. Also, the average of the instantaneous z-angle of the bacterial swarm is used to calculate the yaw-angle error. The control law for the entire system to ensure the yaw-angle reference tracking with reasonable error is given as:

$$\tau_{\{x,y,z\}} = K_{p-\{x,y,z\}}e_{\{x,y,z\}}(t)$$
$$+ \int K_{i-\{x,y,z\}}e_{\{x,y,z\}}(t)\mathrm{d}t + K_{d-\{x,y,z\}}\dot{e}_{\{x,y,z\}}(t) \tag{10}$$

with

$$K_{i-\{x,y,z\}} = 1 \big/ \big((K_{ia-\{x,y,z\}} e_{\{x,y,z\}}(t))^2 + 1 \big) \qquad (11)$$

constituting the adaptive component.

Here, the coefficients $K_{p-\{x,y,z\}}$, and $K_{d-\{x,y,z\}}$ are the proportional and derivative gains for the PID employed in this study. Furthermore, the integral gain, $K_{i-\{x,y,z\}}$, is designed as an adaptive gain [28,41–45] dependent on a separate tuning parameter, $K_{ia-\{x,y,z\}}$. $K_{ia-\{x,y,z\}}$ is presented with a different power on the denominator to further modify its effect as opposed to previous studies [28,41-45], and it is possible to modify the term further to change the behavior of the adaptive contribution of the integral although not studied in this work.

The error functions, $e_{\{x,y,z\}}(t)$, for the joints is given as $e_x(t) = <y_{sw}(t)> - d_2(t)$, $e_y(t) = <x_{sw}(t)> - d_1(t)$, and $e_z(t) = \theta^{ref}(t) - <\theta_{z-sw}(t)>$. The bilateral control coerces the PPR-arm to follow the swarm while inducing the torque on the bacteria necessary to align with the yaw-angle reference thus coupling the dynamics of the multiscale robotic system. Once the control signal, $\tau_{\{x,y,z\}}$, is determined the corresponding pulse-width-modulation (PWM) signal generation follows [28]. The PWM signal is finally amplified to the nominal voltage of the DC-motors for operation, i.e. ±48 V [47] and fed back in Equation (9) to complete the closed-loop control depicted in Figure 2.

III. RESULTS AND DISCUSSION

The following results are obtained for two separate reference functions given by the combination of two ramp functions and one step function as $\theta^{ref}(t) = -(\pi/18)\cdot t\cdot(t >= 0)\cdot(t < 1) - (\pi/18)\cdot(t >= 1)\cdot(t < 2) + (\pi/18)\cdot t\cdot(t >= 2)\cdot(t < 3)$ rad for clockwise rotation and $\theta^{ref}(t) = (\pi/18)\cdot t\cdot(t >= 0)\cdot(t < 1) + (\pi/18)\cdot(t >= 1)\cdot(t < 2) - (\pi/18)\cdot t\cdot(t >= 2)\cdot(t < 3)$ rad for counter-clockwise rotation. The bacteria are initially located on the same xy-plane as depicted in Figure 1 and their relative location with respect to each other are given as $[0\ 0\ 0]^T$ at the center, $[4.625\times10^{-6}\ -4.625\times10^{-6}\ 0]^T$ m to the left and slightly ahead, and $[-4.625\times10^{-6}\ 4.625\times10^{-6}\ 0]^T$ m to the right and slightly behind. The bacteria are initially 1×10^{-2} m away from the bottom surface of the Neodymium magnet while swimming at a proximity of 0.5×10^{-6} m to the boundary of the liquid medium. The stationary surface that is taken to be flat in this study. The simulation interval is set to be $t = [0\ 3]$ s in real time. The tuning parameters for the PID controller are set to be $K_{p31} = 0.1$, $K_{p32} = 0.1$, $K_{i32} = 0.02$, $K_{p33} = 0.1$, $K_{i33} = 0.02$, $K_{p-x} = 75$, $K_{ia-x} = 5$, $K_{p-y} = 75$, $K_{ia-y} = 5$, $K_{d-x} = 0.1$, and $K_{d-y} = 0.1$, on both accounts. All results are obtained by variable step solver ode45, under MATLAB/SIMULINK environment.

Figure 3 is the result of the clockwise and counter-clockwise rotation control studies. The former exhibits spatial overshoots on the orientation of the swarm whereas the latter exhibits a phase delay, i.e., a spatial resistance to the rigid-body rotation. The fact that two different behaviors are observed is based on the direction of the rotation of the helical tails. As the helical tails rotate in the counter-clockwise direction, it is arguably much easier to rotate the swarm itself in the clockwise direction in the lab frame; a response arising due to the hydrodynamic interaction between the solid boundary and the rotating helical tail of each bacterium [54] that in turn affects the whole swarm.
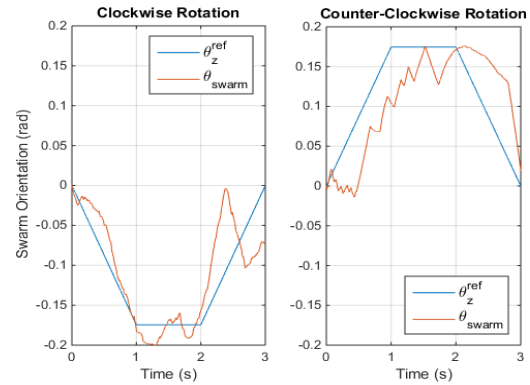

Figure 3. The orientation of the swarm under clockwise and counter-clockwise rigid-body rotation control in the lab-frame.

Figure 4 represents the PWM signals on the dedicated DC-motors at the respective joints of the PPR arm. Only the x-axis performances are observed to be akin to one another whereas the performances on the y-axis and the z-axis are almost the opposite as expected. It can be further observed that the PWM signals dictate the direction of rotation. It is also important to note that the PWM signals are not continuous but exhibit impulse function characteristics due to the incremental change in the reference signal over simulation time.
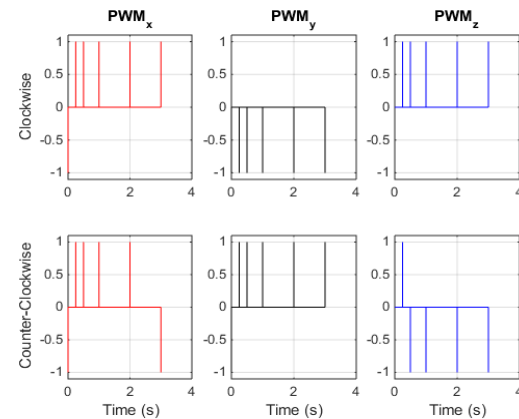

Figure 4. The dedicated PWM signals are generated on the x, y, and z axes of the PPR arm under clockwise and counter-clockwise rigid-body rotation control in the lab-frame.

Figure 5 depicts the currents on the dedicated DC-motors. The observed behavior mirrors the PWM signals, as expected. The currents along the x-axis and y-axis never exceeds 0.01 A; however, the extrema for the z-axis reaches ±0.02 A, although not shown in the figures to preserve the overall resolution. The sudden jerk [55] in the rotation to match the reference angle is the direct cause of this behavior.
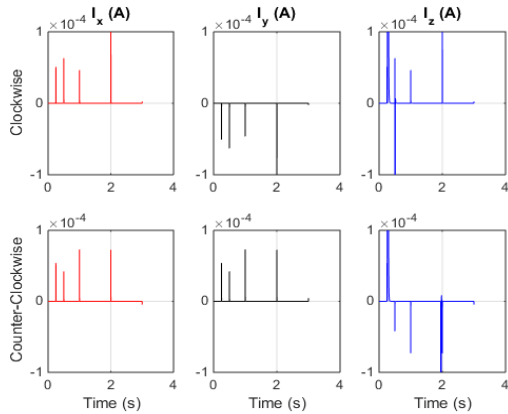

Figure 5. Dedicated DC-motor currents generated on the x, y, and z axes of the PPR arm under clockwise and counter-clockwise rigid-body rotation control in the lab-frame.

Figure 6 illustrates how the magnetic torque is adequate to rotate the swarm. The average of the propulsive torque, $<T_{p,z}>$, on the entire swarm is oscillating around zero meaning that it attains zero N.m several times per period. In the meantime, the average of the magnetic torque on the swarm, $<T_{m,z}>$, is seldom zero. It follows that, although the extrema of the magnetic torque are overwhelmed by the extrema of the propulsive torque, the magnetic torque is constant long enough to overcome the propulsive torque. It is important to note that, this analysis assumes that the magnetic torque can overcome Brownian noise [56]; an analysis previously carried out in detail for Magnetospirillum *Gryphiswaldense* MSR-1 and Neodymium magnet [28].
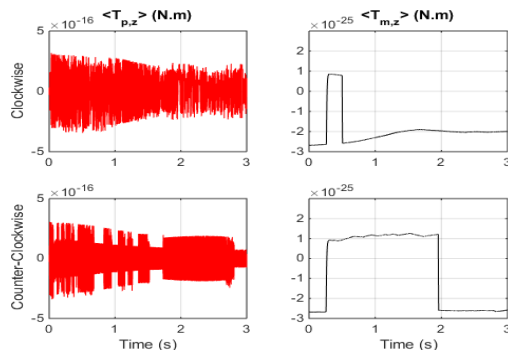

Figure 6. The average of the propulsive torque induced by and magnetic torque exerted on the swarm, both along the z-axis, under clockwise and counter-clockwise rigid-body rotation control in the lab-frame.

Finally, Figure 7 depicts the net repulsive force acting on the bacterium cell at the center over simulation time. It is known that two micro-objects moving and rotating in close formation will affect each other [50,57,58]. The force-components are calculated in the frame of the bacterium, i.e., the forward propulsive force is along the x-axis. The repulsive effect modeled here exhibits that the additional force along the direction of swimming has a net contribution to the x-traverse of the bacterium at the center. On the other hand, the lateral force along the swimming direction, i.e., y-axis, exhibits a swing in the direction as the swarm rotates.
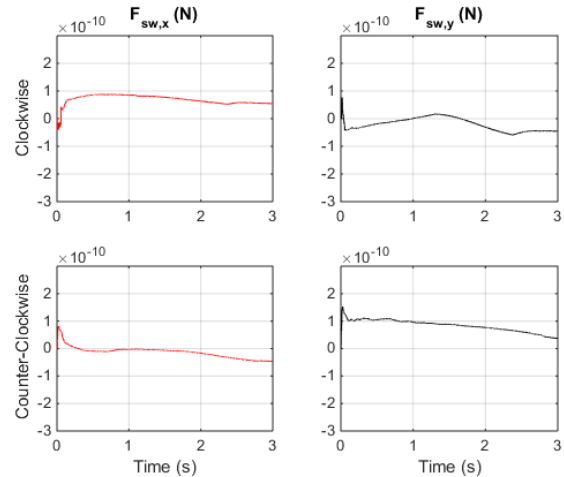

Figure 7. The total repulsion force exerted on the bacterium cell in the middle of the task-group, along the x-axis and y-axis of the frame of the bacterium, under clockwise and counter-clockwise rigid-body rotation control in the lab-frame.

IV. CONCLUSION

The swarm motion in complex biological fluids is of interest due to the envisioned application of swimming micro-robotic swarms, artificial or cybernetic. The study here focuses on the dynamics of three magnetotactic bacteria initially separated from each other at a certain distance and moving as a task force in the synovial fluid of a human at body temperature. The PID control scheme is set to be adaptive on the integral gain to reduce the steady-state error on the yaw-angle while the reference signal is set as discontinuous functions of time. The results show that (i) it is possible to control the swimming direction of a swarm of bacterial and (ii) the repulsive force has a measurable contribution to the force balance.

The study should be replicated for a relatively larger group of bacteria to see the effect of the repulsive force in all main axes. Furthermore, such a simulation study would also help on the effective volume of interest on which the magnetic field is focused. Finally, the study should be replicated under actual dimensions of a synovial joint to imitate the in vivo conditions as much as possible to further study the overall swimming dynamics of the swarm. The results

will be beneficial for the micro-robotic research community as well as the medical science, in the future.

Synovial joints are relatively safer in terms of flow conditions and immune system reactions given that they do not posses blood vessels. The microrobot, artificial or biohybrid, will arguably be subject to less biophysical and biochemical stimuli [59]. Thus, swarms of microrobots can be utilized in synoval cavity for minimally invasive biomedical applications while the patient, or the joint of interest, is stationary without load for the duration of the procedure. A very comparable promising field of application is ophthalmetry [60] where the eye offers a similar biophysical work environment, i.e., non-Newtonian flow with no motion, inspite of presence of blood vessels. In both cases, the flow field is none existent, except for the field induced by the swarm, and the location of the tissuse renders the procedure easier for relatively weaker magnetic fieds to penetrate in adequate strength to acuate and steer microrobots. Therefore, micro-robotic swarm can carry drugs or cells to repair tissue directly on targetted locations [13,19] increasing the effectiveness of the treatment and decreasing the time of recovery.

REFERENCES

[1] Agrahari, V., Agrahari, V., Chou, M.-L., Chew, C. H., Noll, J., & Burnouf, T. (2020). Intelligent Micro-/Nanorobots as Drug and Cell Carrier Devices for Biomedical Therapeutic Advancement: Promising Development Opportunities and Translational Challenges. *Biomaterials*, *260*, 120163.

[2] Cabanach, P., Pena- Francesch, A., Sheehan, D., Bozuyuk, U., Yasa, O., Borros, S., & Sitti, M. (2020). Microrobots: Zwitterionic 3D- Printed Non- Immunogenic Stealth Microrobots. *Advanced Materials*, *32*(42), 2070312.

[3] Ebrahimi, N., Bi, C., Cappelleri, D. J., Ciuti, G., Conn, A. T., Faivre, D., Habibi, N., Hošovský, A., Iacovacci, V., Khalil, I. S. M., Magdanz, V., Misra, S., Pawashe, C., Rashidifar, R., Soto- Rodriguez, P. E. D., Fekete, Z., & Jafari, A. (2020). Magnetic Actuation Methods in Bio/Soft Robotics. *Advanced Functional Materials*, 2005137.

[4] Field, R. D., Anandakumaran, P. N., & Sia, S. K. (2019). Soft medical microrobots: Design components and system integration. *Applied Physics Reviews*, *6*(4), 041305.

[5] Ghosh, A., Xu, W., Gupta, N., & Gracias, D. H. (2020). Active Matter Therapeutics. *Nano Today*, *31*, 100836.

[6] Gunduz, S., Albadawi, H., & Oklu, R. (2020). Robotic Devices for Minimally Invasive Endovascular Interventions: A New Dawn for Interventional Radiology. *Advanced Intelligent Systems*, 2000181.

[7] Hu, M., Ge, X., Chen, X., Mao, W., Qian, X., & Yuan, W.-E. (2020). Micro/Nanorobot: A Promising Targeted Drug Delivery System. *Pharmaceutics*, *12*(7), 665.

[8] Hunter, E. E., Brink, E. W., Steager, E. B., & Kumar, V. (2018). Toward Soft Micro Bio Robots for Cellular and Chemical Delivery. *IEEE Robotics and Automation Letters*, *3*(3), 1592–1599.

[9] Soto, F., Wang, J., Ahmed, R., & Demirci, U. (2020). Medical Micro/Nanorobots in Precision Medicine. *Advanced Science*, *7*(21), 2002203.

[10] Wang, J., Dong, R., Wu, H., Cai, Y., & Ren, B. (2019). A Review on Artificial Micro/Nanomotors for Cancer-Targeted Delivery, Diagnosis, and Therapy. *Nano-Micro Letters*, *12*(1), 1–19.

[11] Bogue, R. (2008). The Development of Medical Microrobots: A Review of Progress. *Industrial Robotics*, 35(4), 294-299.

[12] Felfoul, O., Martel, S. (2013). Assessment of Navigation Control Strategy for Magnetotactic Bacteria in Microchannel: Toward Targeting Solid Tumors. *Biomedical Microdevices*, *15*(6), 1015-1024.

[13] Yasa, C., Tabak, A. F., Yasa, O., Ceylan, H., Sitti, m. (2019). 3D-Printed Microrobotic Transporters with Recapitulated Stem Cell Niche for Programmable and Active Drug Delivery. *Advanced Functional Materials*, *29*(17), 1808992.

[14] Qiu, F., Nelson, B. J. (2015). Magnetic Helical Micro- and Nanorobots: Toward Biomedical Applications. *Engineering*, *1*(1), 21-26.

[15] Ghosh, A., Fischer, P. (2009). Controlled Propulsion of Artificial Magnetic Nanostructured Propellers. *Nano Letters*, *9*(6), 2243–2245.

[16] Dreyfus, R., Baudry, J., Roper, M. L., Fermigier, M., Stone, H. A., Bibette, J. (2005) Microscopic Artificial Swimmer, *Nature*, *437*, 862-865.

[17] Williams, B. J., Anand, S. V., Rajagopalan, J., Saif, M. T. A. (2014). A Self-Propelled Biohybrid Swimmer at Low Reynolds Number, *Nature Communications*, *5*, 3081.

[18] Xiong, X., Lidstrom, M. E., Parviz, B. A. (2007). Microorganisms for MEMS. *Journal of Microelectromechanical Systems*, *16*(2), 429-444.

[19] Alapan, Y., Yasa, O., Schauer, O., Giltinan, J., Tabak, A. F., Sourjik, V., Sitti M. (2018). Soft Erythrocyte-Based Bacterial Microswimmers for Cargo Delivery. *Science Robotics*, *3*(17), eaar4423.

[20] Yasa, I. C., Ceylan, H., Bozuyuk, U., Wild, A.-M., & Sitti, M. (2020). Elucidating the Interaction Dynamics Between Microswimmer Body and Immune System for Medical

Microrobots. *Science Robotics*, *5*(43), eaaz3867.

[21] Park, H., & Park, K. (1996). Biocompatibility Issues of Implantable Drug Delivery Systems. *Pharmaceutical Research*, *13*(12), 1770–1776.

[22] Horie, M., Fujita, K., Kato, H., Endoh, S., Nishio, K., Komaba, L. K., Nakamura, A., Miyauchi, A., Kinugasa, S., Hagihara, Y., Niki, E., Yoshida, Y., & Iwahashi, H. (2012). Association of the Physical and Chemical Properties and the Cytotoxicity of Metal Oxide Nanoparticles: Metal Ion Release, Adsorption Ability and Specific Surface Area. *Metallomics*, *4*(4), 350.

[23] Ceylan, H., Yasa, I. C., Tabak, A. F., Giltinan, J., Sitti, M. (2019). 3D-Printed Biodegradable Microswimmer for Theranostic Cargo Delivery and Release. *ACS Nano*, *13*(3), 3353–3362.

[24] Uenoyama, A., Miyata, M. (2005). Gliding Ghosts of Mycoplasma Mobile. *Proceedings of the National Academy of Sciences USA*, *102*(36), 12754-12758.

[25] Patiño, T., Feiner-Gracia, N., Arqué, X., Miguel-López, A., Jannasch, A., Stumpp, T., Schäffer, E., Albertazzi, L., & Sánchez, S. (2018). Influence of Enzyme Quantity and Distribution on the Self-Propulsion of Non-Janus Urease-Powered Micromotors. *Journal of the American Chemical Society*, *140*(25), 7896–7903.

[26] Sun, H. C. M., Liao, P., Wei, T., Zhang, L., & Sun, D. (2020). Magnetically Powered Biodegradable Microswimmers. *Micromachines*, *11*(4), 404.

[27] Go, G., Jeong, S.-G., Yoo, A., Han, J., Kang, B., Kim, S., Nguyen, K. T., Jin, Z., Kim, C.-S., Seo, Y. R., Kang, J. Y., Na, J. Y., Song, E. K., Jeong, Y., Seon, J. K., Park, J.-O., & Choi, E. (2020). Human Adipose–Derived Mesenchymal Stem Cell–Based Medical Microrobot System for Knee Cartilage Regeneration In Vivo. *Science Robotics*, *5*(38), eaay6626.

[28] Tabak, A. F. (2020). Bilateral Control Simulations for a Pair of Magnetically-Coupled Robotic Arm and Bacterium for In Vivo Applications. *Journal of Micro-Bio Robotics*, *16*(2), 199–214.

[29] Ahmed, D., Sukhov, A., Hauri, D., Rodrigue, D., Maranta, G., Harting, J., & Nelson, B. J. (2021). Bioinspired Acousto-Magnetic Microswarm Robots with Upstream Motility. *Nature Machine Intelligence*, 1.

[30] Dong, X., & Sitti, M. (2020). Controlling Two-Dimensional Collective Formation and Cooperative Behavior of Magnetic Microrobot Swarms. *The International Journal of Robotics Research*, *39*(5), 617–638.

[31] Keya, J. J., Kabir, A. M. R., Inoue, D., Sada, K., Hess, H., Kuzuya, A., & Kakugo, A. (2018).

[32] Mirzakhanloo, M., & Alam, M.-R. (2020). Stealthy Movements and Concealed Swarms of Swimming micro-robots. *Physics of Fluids*, *32*(7), 071901.

[33] Morozov, K. I., & Leshansky, A. M. (2020). Towards Focusing of a Swarm of Magnetic Micro/Nanomotors. *Physical Chemistry Chemical Physics*, *22*(28), 16407–16420.

[34] Servant, A., Qiu, F., Mazza, M., Kostarelos, K., & Nelson, B. J. (2015). Controlled In Vivo Swimming of a Swarm of Bacteria- Like Microrobotic Flagella. *Advanced Materials*, *27*(19), 2981–2988.

[35] Xu, T., Soto, F., Gao, W., Dong, R., Garcia-Gradilla, V., Magaña, E., Zhang, X., & Wang, J. (2015). Reversible Swarming and Separation of Self-Propelled Chemically Powered Nanomotors under Acoustic Fields. *Journal of the American Chemical Society*, *137*(6), 2163–2166.

[36] Yousefi, M., & Nejat Pishkenari, H. (2021). Independent Position Xontrol of Two Identical Magnetic Microrobots in a Plane Using Rotating Permanent Magnets. *Journal of Micro-Bio Robotics*, 1–9.

[37] Johnson, B. V., Esantsi, N., & Cappelleri, D. J. (2020). Design of the μMAZE Platform and Microrobots for Independent Control and Micromanipulation Tasks. *IEEE Robotics and Automation Letters*, *5*(4), 5677–5684.

[38] Khalil, I. S. M., Tabak, A. F., Hamed, Y., Tawakol, M., Klingner, A., Gohary, N. E., Mizaikoff, B., & Sitti, M. (2018). Independent Actuation of Two-Tailed Microrobots. *IEEE Robotics and Automation Letters*, *3*(3), 1703–1710.

[39] Khalil, I. S. M., Pichel, M. P., Abelmann, L., & Misra, S. (2013). Closed-loop control of magnetotactic bacteria. *The International Journal of Robotics Research*, *32*(6), 637–649.

[40] Khalil, I. S. M., Tabak, A. F., Hageman, T., Ewis, M., Pichel, M., Mitwally, M. E., El-Din, N. S., Abelmann, L., & Sitti, M. (2017). Near-surface effects on the controlled motion of magnetotactic bacteria. *2017 IEEE International Conference on Robotics and Automation (ICRA)*.

[41] Tabak, A. F. (2020). Motion Control for Biohybrid Multiscale Robots. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*.

[42] Tabak, A. F. (2020). Adaptive Motion Control of Modified E. Coli. *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*.

[43] Tabak, A. F. (2020). A Simulated Control Method for a Magnetically-Coupled Bacterium and Robotic Arm. *2020 International*

Conference on Manipulation, Automation and Robotics at Small Scales (MARSS).

[44] Tabak, A. F. (2020). Simulated Bilateral Motion Control of a Magneto-Tactic Bacterium via an Open Kinematic Chain. *2020 17th International Conference on Ubiquitous Robots (UR)*.

[45] Tabak, A. F. (2020). Independent Joint Control Simulations on Adaptive Maneuvering of a Magnetotactic Bacterium via a Single Permanent Magnet. *European Journal of Science and Technology*, 50–59.

[46] Dong, F., Huang, Z., Qiu, D., Hao, L., Wu, W., & Jin, Z. (2019). Design and Analysis of a Small-Scale Linear Propulsion System for Maglev Applications (1)—The Overall Design Process. *IEEE Transactions on Applied Superconductivity*, *29*(2), 1–5.

[47] *Maxon Motor Product Catalog*. Https://Www.Maxongroup.Com. Retrieved February 13, 2021, from https://www.maxongroup.com/maxon/view/product/397172

[48] Tabak, A. F. (2018). Hydrodynamic Impedance of Bacteria and Bacteria-Inspired Micro-Swimmers: A New Strategy to Predict Power Consumption of Swimming Micro-Robots for Real-Time Applications. *Advanced Theory and Simulations*, *1*(4), 1700013.

[49] Brennen, C., & Winet, H. (1977). Fluid Mechanics of Propulsion by Cilia and Flagella. *Annual Review of Fluid Mechanics*, *9*(1), 339–398.

[50] Ishikawa, T., Sekiya, G., Imai, Y., & Yamaguchi, T. (2007). Hydrodynamic Interactions between Two Swimming Bacteria. *Biophysical Journal*, *93*(6), 2217–2225.

[51] Spong, M. W., & Vidyasagar, M. (1989). *Robot Dynamics and Control* (1st ed.). Wiley.

[52] Mazzucco, D., McKinley, G., Scott, R. D., & Spector, M. (2002). Rheology of Joint Fluid in Total Knee Arthroplasty Patients. *Journal of Orthopaedic Research*, *20*(6), 1157–1163.

[53] Smith, A. M., Fleming, L., Wudebwe, U., Bowen, J., & Grover, L. M. (2014). Development of a Synovial Fluid Analogue with Bio-Relevant Rheology for Wear Testing of Orthopaedic Implants. *Journal of the Mechanical Behavior of Biomedical Materials*, *32*, 177–184.

[54] Lauga, E., DiLuzio, W. R., Whitesides, G. M., & Stone, H. A. (2006). Swimming in Circles: Motion of Bacteria near Solid Boundaries. *Biophysical Journal*, *90*(2), 400–412.

[55] Eager, D., Pendrill, A.-M., & Reistad, N. (2016). Beyond Velocity and Acceleration: Jerk, Snap and Higher Derivatives. *European Journal of Physics*, *37*(6), 065008.

[56] Ghosh, A., & Fischer, P. (2009). Controlled Propulsion of Artificial Magnetic Nanostructured Propellers. *Nano Letters*, *9*(6), 2243–2245.

[57] Goldfriend, T., Diamant, H., & Witten, T. A. (2015). Hydrodynamic Interactions between Two Forced Objects of Arbitrary Shape. I. Effect on Alignment. *Physics of Fluids*, *27*(12), 123303.

[58] Goldfriend, T., Diamant, H., & Witten, T. A. (2016). Hydrodynamic Interactions between two Forced Objects of Arbitrary Shape. II. Relative Translation. *Physical Review E*, *93*(4), 042609.

[59] Kuettner, K., & Cole, A. (2005). Cartilage Degeneration in Different Human Joints. *Osteoarthritis and Cartilage*, *13*(2), 93–103.

[60] Ullrich, F., Fusco, S., Chatzipirpiridis, G., Pané, S., & Nelson, B. J. (2014). Recent Progress in Magnetically Actuated Microrobotics for Ophthalmic Therapies. *European Ophthalmic Review*, *08*(02), 120.

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

3B Yazıcının Kontrolünde Optimizasyon Algoritmalarının Performansı

Performance of Optimization Algorithms in the Control of 3D Printer

Aytaç ALTAN[1] ⓘ **, Ahmet PARLAK[1]** ⓘ

[1]Zonguldak Bülent Ecevit Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, 67100, Zonguldak, Türkiye

Öz

Üç boyutlu (3B) biyo yazıcılar, rejeneratif tıp ve doku mühendisliği alanlarında özellikle kulak, burun ve yüz-çene protezlerinin biyo baskılarında yoğun şekilde kullanılmaktadırlar. Modelden modele farklılık gösteren baskı hataları yapay doku ve organların biyo-baskılarında sıklıkla görülmektedir. Modelin yüzeyinde meydana gelen hatalar yazdırılacak organın verimli kullanımına engel olmaktadır. Yapay doku ve organ biyo-baskısı sürecinde meydana gelen hataların en aza indirilebilmesi için 3B yazıcının kontrolcü performansının iyileştirilmesi gerekmektedir. Bu çalışmada, yapay doku ve organların biyo-baskıları için İyileştirilmiş Gri Kurt Optimizasyon (İGKO) tabanlı yeni bir uyarlanabilir PID kontrolcü geliştirilmiştir. Yerel minimumlardan kaçınmak için İGKO algoritması tercih edilmiştir. Önerilen algoritmanın yakınsama hızı PID kontrolcünün parametrelerinin hızlı ve doğru şekilde ayarlanabilmesine olanak sağlamaktadır. Geliştirilen İGKO tabanlı uyarlanabilir PID kontrolcünün performansı, performans metriklerinden biri olan zaman ağırlıklı karesel hatanın integrali (Integral of Time multiplied Squared Error-ITSE) yardımıyla ölçülmüştür. 3B yazıcı için önerilen kontrolcünün performansı, klasik PID ve Balina Optimizasyon Algoritması (BOA) tabanlı PID kontrolcülerin performansı ile karşılaştırılmıştır. Elde edilen deneysel sonuçlardan, önerilen İGKO tabanlı uyarlanabilir PID kontrolcünün 3B yazıcının geçici tepkisini önemli ölçüde iyileştirdiği ve yazdırılan burun ve kulak gibi organlardaki yüzey hatalarını en aza indirdiği görülmektedir.

Anahtar Kelimeler: 3B yazıcı, Yapay doku ve organ, Geliştirilmiş gri kurt optimizasyon, Uyarlamalı kontrol, Burun ve kulak protezi.

Abstract

Three-dimensional (3D) bio printers are used extensively in regenerative medicine and tissue engineering, especially in bioprinting of ear, nose and face-chin prostheses. Printing defects that differ from model to model are frequently seen in the bio-printing of artificial tissue and organ. Defects occurring on the surface of the model prevent the efficient use of the organ to be printed. The controller performance of the 3D printer needs to be improved so that defects in the artificial organ and tissue bio-printing process can be minimized. In this study, a novel adaptive PID controller based on Improved Grey Wolf Optimization (IGWO) has been developed for bio-printing of artificial tissues and organs. The IGWO algorithm has been preferred to avoid local minima. The convergence speed of the proposed algorithm allows the parameters of the PID controller to be adjusted quickly and accurately. The performance of the developed IGWO based adaptive PID controller has been measured with the help of Integral of Time multiplied Squared Error (ITSE), one of the performance metrics. The performance of the proposed controller for the 3D printer has been compared to the classical PID and Whale Optimization Algorithm (WOA) based PID controllers's performance. From the experimental obtained results, it can be seen that the proposed IGWO based adaptive PID controller significantly improves the 3D printer's transient response and minimizes surface defects in the printed organs such as the nose and ear.

Keywords: 3D printer, Artificial tissue and organ, Improved grey wolf optimization, Adaptive control, Nose and ear prosthesis.

I. GİRİŞ

3B yazıcılar mühendislik ve sağlık alanları başta olmak üzere endüstriyel sanatlar, otomotiv gibi birçok alanda yaygın olarak kullanılmaktadırlar. Üç boyutlu bir modelin yazdırılarak madde haline getirilmesini sağlayan 3B yazıcılar son yıllarda rejeneratif tıp ve doku mühendisliği alanlarında özellikle kulak, burun, kemik ve yüz-çene protezlerinin biyo baskılarında yoğun şekilde kullanılmaktadırlar [1-3]. Yapay dokuların biyofabrikasyonu için hücre-biyomateryal etkileşimlerini optimize ederek doku hasarı gibi zorlukların üstesinden gelen rejeneratif tıp alanında kullanılan 3B yazıcıların kontrolü genellikle klasik PID ile gerçekleştirilmektedir. 3B yazıcının ortopedik cerrahide kullanılan basit geometri ve sert yapıya sahip protez kemik modelinin yazdırılmasında gösterdiği performans ile rejeneratif tıp alanında kullanılan iç içe geçmiş karmaşık geometri ve yumuşak yapıya sahip kulak ve burun gibi modellerin yazdırılmasında gösterdiği performans arasında büyük farklılıklar görülmektedir [4, 5]. Hem sert hem de yumuşak dokuya sahip yüz-çene protezi gibi modellerin yazdırılmasında ise kontrolcü performansından kaynaklı yüzeysel baskı hataları ön plana çıkmaktadır. Bu doku ve organların biyo-baskılarında modelden modele değişen hataların en aza indirilmesi modelin etkin kullanımı açısından

oldukça önemlidir [6]. Yazdırılan modelin referans modele göre yüksek hata oranına sahip olması, hastanın yazdırılan protez modelden kaynaklı başka problemlerle karşılaşmasına neden olabilir. Bu hata oranının en aza indirilebilmesi amacıyla bu çalışmada İyileştirilmiş Gri Kurt Optimizasyon (İGKO) algoritması tabanlı yeni bir uyarlanabilir PID kontrolcü geliştirilmiştir.

Yüz, çene, kafatası, boyun ve ağız bölgesi ile ilgili hastalık, travma, doku eksikliği ve estetik sorunların tanı ve tedavisi maksillofasiyel cerrahi olarak adlandırılmaktadır. Plastik cerrahi, estetik cerrahi, rekonstrüktif mikrocerrahi, ortognatik cerrahi, diş hekimliği ve kulak burun boğaz hekimliği maksillofasiyel cerrahinin ilgi alanları arasında yer almaktadır. Hastanın yüz bölgesinde, özellikle kulak, burun ve çene bölgesindeki kusurlar, hastanın duygusal bütünlüğünü, sosyal etkileşimini, benlik algısını ve yaşam kalitesini olumsuz yönde etkilemektedir. Maksillofasiyal kusurları olan hastaların yaşam kalitesi, cerrahi rekonstrüksiyon veya protez sağlandığında bile genellikle risk altındadır [6-9]. Travmatik burun ve kulak kusurları, özellikle toplumda gerçekleştirilen aktivitelerde özgüven, moral ve motivasyon kayıplarına neden olabilmektedir [10]. Burun ve kulağın yüz bölgesindeki konumu göz önünde bulundurulduğunda, bu organların kaybı veya bu organlarda oluşacak kusurlar hastanın fiziksel görünümünü olumsuz yönde etkileyeceğinden, cerrahi operasyonun mümkün olmadığı durumlarda giyilebilir protez burun veya kulağın kullanımı büyük önem taşımaktadır [11]. Ayrıca, aynalanacak mevcut bir parça olmadığında, geleneksel taş kalıplar kullanılarak üretilen protez burun ve kulağın hastaya uygulanmasında problemler yaşandığı bildirilmektedir [12]. Literatürde yapılan çalışmalarda, burun ve kulak kaybı olan hastalara uygun protez organların basılması için hastanın yüz bölgesindeki yumuşak ve sert doku ayrı ayrı segmentlere ayrılmaktadır. Protezin hastanın cilt rengine uyumlu olması, hastanın protezi etkin kullanabilmesi açısından önemli olduğu için hastanın cilt rengi bir spektrofotometre ile ölçülmektedir. Protez organı takmadan önce kulak veya burun kaybı olan hastaların görünümü sırasıyla Şekil 1 (a) ve Şekil 2 (a) 'da gösterilmektedir. Hastanın cilt tonuna uygun olarak 3B yazıcıda basılan kulak ve burun, yüz bölgesinde segmentlere ayrılan yumuşak doku ve sert doku ile bütünleştirilmektedir. [1, 11]. 3B yazıcıda ile yazdırılan ve hastanın yüz bölgesine monte edilen protez kulak ve burnun görünümü sırasıyla Şekil 1 (b) ve Şekil 2 (b) 'de gösterilmektedir.

Bu çalışmada, yapay doku ve organların biyo baskıları için 3B yazıcıda kullanılmak üzere yeni bir İGKO tabanlı uyarlanabilir PID kontrolcü geliştirilmiştir. Geliştirilen uyarlanabilir kontrolcünün parametreleri, İGKO algoritması kullanılarak ayarlanmaktadır. Ayrıca, protez organların yazdırılmasında, üretimden kaynaklanan hataların en aza indirilebilmesi için özel

olarak tasarlanan ve üç farklı üretim malzemesini tek bir sistemde kullanabilen ekstruder sarkacının konum kontrolü de uyarlamalı olarak gerçekleştirilmektedir. Önerilen kontrolcü yaklaşımının performansı deneysel çalışmalarla doğrulanmaktadır.
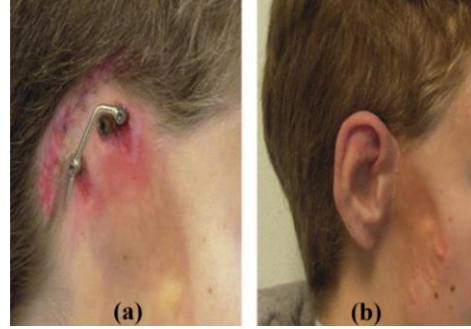
Şekil 1. Kulak kaybı olan (a) protezsiz (b) protezli hastanın görünümü [8]


Şekil 2. Burun kaybı olan (a) protezsiz (b) protezli hastanın görünümü [10]

Bu çalışmanın geri kalanı şu şekilde organize edilmiştir: yapay doku ve organların biyo-baskıları için 3B yazıcı tasarımı Bölüm II'de anlatılmaktadır. Geliştirilen modele dayanak oluşturan Gri Kurt Optimizasyon (GKO) algoritması Bölüm III'te sunulmaktadır. İGKO tabanlı olarak geliştirilen uyarlanabilir PID kontrolcü algoritmasının 3B yazıcıya uygulanması Bölüm IV'te gösterilmektedir. Bölüm V'te deneysel sonuçlar verilmekte ve elde edilen sonuçlar tartışılmaktadır. Çalışmanın sonuçları Bölüm VI'da vurgulanmaktadır.

II. DOKU VE ORGAN BİYO BASKISI İÇİN 3B YAZICI TASARIMI

Kartezyen, delta ve çekirdek xy gibi farklı tasarımlarda üretilen 3B yazıcılar genellikle xyz kartezyen koordinat sistemine bağlı ekstruder, nozul ve ısıtmalı yataktan oluşmaktadır [13]. Çalışmada kullanılan 3B yazıcının x, y ve z eksenlerinin hareketi aynı özelliğe sahip vidalı miller yardımı ile gerçekleştirilmektedir. Tasarlanan 3B yazıcının mekanik yapısı Şekil 3'te gösterilmektedir. Tasarlanan 3B yazıcının boyutları $600 \times 600 \times 700$ mm'dir. Yapay doku ve organların yazdırılması için kartezyen tip olarak tasarlanan 3B yazıcıda eksenler için dört adet (z ekseninde iki adet) ve filament sistemi için üç adet olmak üzere toplam yedi adet endüstriyel tip

servo motor kullanılmaktadır. Ayrıca, ekstruder sarkaç için bir adet step motor kullanılmaktadır.
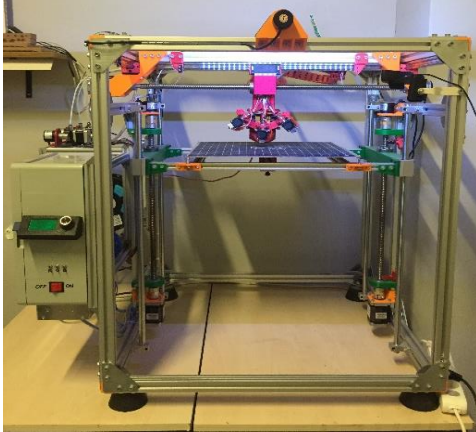
Şekil 3. Kulak ve burun gibi doku ve organların baskısı için tasarlanan 3B yazıcının mekanik yapısı

Tasarlanan 3B yazıcı 400×400 mm ısıtmalı yatağa ve 0.2 mm, 0.4 mm ve 0.6 mm çaplı nozullardan oluşan bir ekstruder sarkaş sistemine sahiptir. Bu çalışmada kullanılan 3B yazıcıyı diğer yazıcılardan ayıran en önemli özellik, ekstruder sarkaç sistemine sahip olmasıdır. Üç nozuldan oluşan ekstruder sarkaç sistemi Şekil 4'te gösterilmektedir. Yapay doku ve organların minimum hata ile basılabilmesi için 3B yazıcının eksen kontrolünün yanısıra ekstruder sarkaç sisteminin de etkin kontrolü oldukça önemlidir. Bu problemin üstesinden gelebilmek için bu çalışmada 3B yazıcılardaki klasik PID kontrolcü algoritması yerine İGKO tabanlı uyarlanabilir PID kontrolcü algoritması geliştirilmekte, 3B yazıcı üzerinde bulunan ve Şekil 5'te gösterilen kontrol kartı üzerine gömülmektedir. Geliştirilen kontrolcü algoritması aynı zamanda üç nozuldan oluşan ekstruder sarkaç sisteminin kontrolünü de gerçekleştirmektedir. Tasarlanan 3B yazıcı üç çeşit filamenti ekstruder sarkaç sistemi yardımıyla aynı anda kullanabilmektedir. Bu tür bir sistemin gereksinimi, modelin hastanın cilt rengindeki tonlamalara uygun olarak yazdırılabilmesine olanak sağlamasıdır. Ayrıca, model katmanlarında çeşitli tipte malzeme kullanımına imkân verilerek protez modelin yumuşak-sert doku uyumu gerçekleştirilebilmektedir.
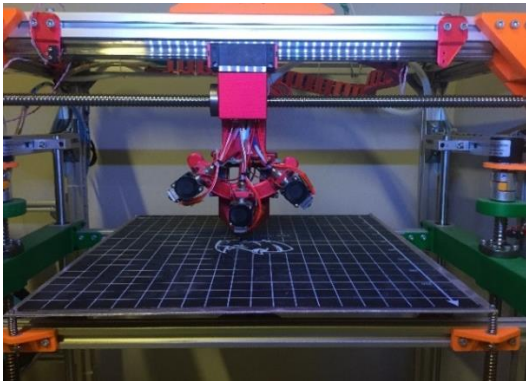
Şekil 4. Doku ve organ biyo-baskısı için tasarlanan 3B yazıcının ekstruder sarkaç sistemi
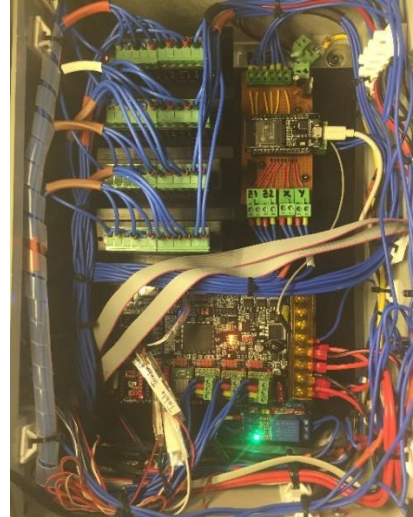
Şekil 5. Geliştirilen İGKO tabanlı uyarlanabilir PID kontrolcü algoritmasının gömülü olduğu kontrol kartı

III. GRİ KURT OPTİMİZASYONU

GKO, doğadaki gri kurtların avlanma davranışının ve sosyal liderliğinin simülasyonuna dayanan sürü zekâsı temelli evrimsel bir hesaplama yöntemi olan yeni ve etkili bir meta-sezgisel optimizasyon algoritmasıdır [14]. Gri kurtlardan esinlenen bu algoritmada gri kurdun liderliği ve avlanma mekanizmaları taklit edilir. Liderlik hiyerarşisini taklit etmek için alfa (α), beta (β), delta (δ) ve omega (ω) olmak üzere dört tür gri kurt tanımlanmaktadır. GKO algoritmasının doğası ω kurtlar tarafından takip edilirken α, β ve δ kurtları optimizasyon sürecine rehberlik etmektrdir. Gri kurt avının ana aşamaları:

- avı kovalamak, yaklaşmak ve takip etmek
- avın peşinde koşmak, taciz etmek ve etrafını sarmak
- sabit bekleme ve avına doğru saldırı.

Gri kurtların avlanma davranış aşamaları Şekil 6'da gösterilmektedir. Avın peşinden koşma, yaklaşma ve takip etme aşaması Şekil 6 (A)'da; avın peşinden koşma, taciz etme ve kuşatma aşamaları Şekil 6 (B-D)'de; durağan bekleme ve avına doğru saldırı aşamaları Şekil 6 (E)'de gösterilmektedir.

Şekil 6. Gri kurtların avlanma davranışı

Gri kurdun avlanma boyunca etrafı saran davranışı [14, 15]:

$$\vec{D} = |\vec{C} \cdot \overrightarrow{X_P}(t) - \vec{X}(t)| \tag{1}$$

$$\vec{X}(t+1) = \overrightarrow{X_P}(t) - \vec{A} \cdot \vec{D} \tag{2}$$

ile modellenmektedir. Burada, t mevcut yinelemeyi temsil etmektedir. $\vec{A}$ ve $\vec{C}$ katsayı vektörleridir. $\vec{X}(t)$ ve $\overrightarrow{X_P}(t)$, sırasıyla gri kurt ve avın konum vektörleridir. $\vec{A}$ ve $\vec{C}$ vektörleri,

$$\vec{A} = 2\vec{a} \cdot \vec{r_1} - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r_2} \tag{4}$$

yardımıyla hesaplanmaktadır. Burada, $\vec{r_1}$ ve $\vec{r_2}$, $[0, 1]$ aralığında rastgele vektörlerdir. $\vec{a}$ bileşenleri iterasyon boyunca doğrusal olarak 2'den 0'a azaltılmaktadır.

Gri kurtlar, avın konumunu tanıma ve onu kuşatma yeteneğine sahiptirler. Gri kurtların avlanma davranışını matematiksel olarak simüle etmek için β ve δ kurtların avın potansiyel konumu hakkında daha iyi bilgiye sahip olduğu varsayılarak α en iyi çözüm kabul edilmektedir. Bu süreç,

$$\overrightarrow{D_\alpha} = |\overrightarrow{C_1} \cdot \overrightarrow{X_\alpha} - \vec{X}| \tag{5}$$

$$\overrightarrow{D_\beta} = |\overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \vec{X}| \tag{6}$$

$$\overrightarrow{D_\delta} = |\overrightarrow{C_3} \cdot \overrightarrow{X_\delta} - \vec{X}| \tag{7}$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot (\overrightarrow{D_\alpha}) \tag{8}$$

$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot (\overrightarrow{D_\beta}) \tag{9}$$

$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot (\overrightarrow{D_\delta}) \tag{10}$$

$$\vec{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{11}$$

ile tanımlanmaktadır. Burada, t mevcut yinelemeyi ifade etmektedir. $\overrightarrow{A_1}$, $\overrightarrow{A_2}$ ve $\overrightarrow{A_3}$ rastgele vektörlerdir. $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$ ve $\overrightarrow{X_\delta}$ vektörleri sırasıyla α, β ve δ kurtlarının konumlarını belirtmektedirler.

Avın hareketi durduğunda gri kurtlar bir saldırı ile avı bitirirler. Avına yaklaşan matematiksel model için $\vec{a}$ değeri azaltılır. O sırada, $\vec{A}$'nın dalgalanma aralığı da $\vec{a}$ tarafından azaltılmaktadır. Bir arama ajanının bir sonraki konumu, mevcut konumu ile avın konumu arasındaki herhangi bir konum olabilir. Avlarını aramak için birbirlerinden ayrılan ve avına saldırmak için bir araya gelen gri kurtlar, çoğunlukla α, β ve δ kurtlarının konumuna göre arama yaparlar. Matematiksel model sapması için, arama ajanı avdan 1'den büyük veya -1'den küçük rastgele değerlerle

ayrılmalıdır. GKO algoritması prosedürü aşağıda tanımlanmıştır:

1: Gri kurt nüfusunu X_i $(i = 1, 2, \dots, n)$ **başlat**
2: a, A ve C **başlat**
3: Her arama ajanının uygunluğunu **hesapla**
    X_α: en iyi arama ajanı
    X_β: ikinci en iyi arama ajanı
    X_δ: üçüncü en iyi arama ajanı
4: $(t < maksimum\ yineleme\ sayısı)$ **iken**,
5: Her arama ajanı **için**
    5.1: Denklem (11) tarafından mevcut arama ajanının konumunu **güncelle**
6: a, A ve C **güncelle**
7: Her ajanın uygunluğunu **hesapla**
8: X_α, X_β ve X_δ **güncelle**
9: $t = t + 1$
10: X_α değerini **döndür**.

Bu çalışmada yerel minimumlardan kaçınmak için GKO algoritması iyileştirilmiş ve kontrolcü parametre katsayıları belirlenmiştir. İyileştirilen algoritma Bölüm IV'te tanıtılmaktadır.

IV. 3B YAZICI İÇİN GELİŞTİRİLMİŞ GRİ KURT OPTİMİZASYON TABANLI PID KONTROLCÜ TASARIMI

Yapay doku ve organların biyo-baskılarında kullanılan 3B yazıcıların kontrolü genellikle klasik PID kontrolcü algoritması ile gerçekleştirilmektedir. 3B yazıcı ile yazdırılan protezlerin imalat süreçlerinde meydana gelen hataların, özellikle yazdırılan protezlerin yüzeylerindeki hataların en aza indirilebilmesi için uyarlanabilir PID kontrolcü algoritması önerilmiştir [13]. Bu bölümde, yapay doku ve organların biyo-baskıları için tasarlanan 3B yazıcının geçici durum yanıtının iyileştirilebilmesi amacıyla önerilen İGKO tabanlı PID kontrolcü algoritması sunulmaktadır. Tasarlanan 3B yazıcı için geliştirilen İGKO tabanlı PID kontrolcünün blok yapısı Şekil 7'de gösterilmektedir. Şekil 7'de θ_r referans açısal konumu ifade etmektedir. θ_e açısal konum hatasını ve u_q kontrolcü çıkış sinyalini belirtmektedir. θ, kapalı döngü kontrol çıkışı olup gerçek açısal konumu göstermektedir. Geliştirilen İGKO tabanlı uyarlanabilir PID kontrolcünün performansı, en yaygın kullanılan performans ölçütlerinden biri olan zaman ağırlıklı karesel hatanın integrali (Integral of Time multiplied Squared Error-ITSE) yardımıyla ölçülmektedir. Denklem (12) ile ifade edilen bu performans ölçütü geliştirilen kontrolcünün parametrelerinin ayarlanmasında kullanılmaktadır.

$$ITSE = \int_0^\infty t[e(t)^2]dt \tag{12}$$

Önerilen kontrolcü yapısında kullanılan İGKO algoritması prosedürü aşağıda tanımlanmıştır [17]:

Girişler : n, d ve maksimum yineleme sayısı

Çıkış   : Global optimum

1: Gri kurt nüfusunu X_i ($i = 1, 2, …, n$) *başlat*

2: a, A ve C *başlat*

3: Maksimum yineleme sayısı *için*

    Her arama ajanının uygunluğunu *hesapla*

    X_α: en iyi arama ajanı

    X_β: ikinci en iyi arama ajanı

    X_δ: üçüncü en iyi arama ajanı

4: ($t < maksimum\ yineleme\ sayısı$) *için*,

5: Her arama ajanı *için*

    5.1: Denklem (8-10) ile konum *güncelle*

    5.2: Denklem (11) tarafından mevcut arama ajanının konumunu *güncelle*

    5.3: $R_i(t) = \|X_i(t) - X_i(t+1)\|$ *güncelle*

6: $d = 1$'den D *için*

    6.1: Boyut öğrenmeye dayalı avlanma için

$X_{i-DHL,d}(t+1) = X_{i,d}(t) + rand \times (X_{n,d}(t) - X_{r,d}(t))$

7: a, A ve C *güncelle*

8: X_α, X_β ve X_δ *güncelle*

9: En iyi boyut öğrenmeye daylı konum ve mevcut ara ajan konumunu *güncelle*

10: Her ajanın uygunluğunu *hesapla*

11: Gri kurt nüfusunu *güncelle*

12: Global optimum değerini *döndür*.

V. DENEYSEL SONUÇLAR

Yapay doku ve organların biyo-baskıları için tasarlanan 3B yazıcının performansı, sırasıyla klasik PID ve geliştirilen İGKO tabanlı uyarlanabilir PID kontrolcü algoritmaları ile test edilmiştir. 3B yazıcının MATLAB ortamında oluşturulan matematiksel modeli üzerinde yapılan sistem yanıt analizine göre klasik PID kontrolcünün kazanç parametreleri Ziegler-Nichols yöntemi kullanılarak belirlenmiştir. Çalışmada önerilen kontrol algoritmasının optimizasyon süreci için popülasyon büyüklüğü ve toplam adım sayısı sırasıyla 40 ve 50 olarak belirlenmiştir. Çalışmadaki tüm kodlar MATLAB 2020b ile derlenmiştir. Geliştirilen İGKO tabanlı uyarlanabilir PID kontrolcünün geçici durum yanıtı hem klasik PID kontrocünün hem de Balina Optimizasyon Algoritması (BOA) tabanlı PID kontrolcünün geçici durum yanıtı ile karşılaştırılmış ve sonuçlar Tablo 1'de sunulmuştur. Tablo 1'deki sonuçlar incelendiğinde, önerilen İGKO tabanlı uyarlanabilir PID kontrolcüsünün sistemin yanıtını önemli ölçüde iyileştirdiği doğrulanmıştır.

Tasarlanan 3B yazıcının performansı PID, BOA tabanlı PID ve İGKO tabanlı PID kontrolcü algoritmaları ile deneysel olarak test edilmiştir. Bu üç algoritma, sırayla tasarlanan 3B yazıcı üzerine gömülmüş ve protez kulak ve burun modelleri yazdırılmıştır. Klasik PID kontrolcüye sahip 3B yazıcı ile yazdırılan protez kulak ve burun modelleri sırasıyla

Şekil 8 ve Şekil 9'da verilmektedir. BOA tabanlı uyarlanabilir PID kontrolcüye ve önerilen İGKO tabanlı uyarlanabilir PID kontrolcüye sahip 3B yazıcı ile yazdırılan protez kulak ve burun modelleri sırasıyla Şekil 10 ve Şekil 11'de gösterilmektedir.
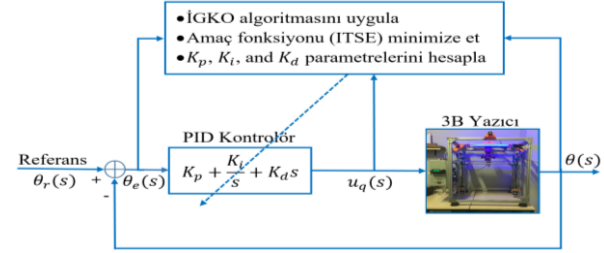
Şekil 7. Tasarlanan 3B yazıcı için geliştirilen İGKO tabanlı uyarlamalı PID kontrolcü blok yapısı

Tablo 1. 3B yazıcının geçici yanıt analizi sonuçları

| Kontrolcü	Yükselme zamanı (s)	Yerleşme zamanı (2%) (s)	Maksimum aşma (%)	Tepe zamanı (s)
PID [13]	0.78×10^{-3}	0.97×10^{-2}	41.67	0.36×10^{-2}
BOA-PID [16]	0.32×10^{-6}	0.08×10^{-5}	No overshot	0.14×10^{-5}
İGKO-PID	0.49×10^{-8}	0.19×10^{-7}	No overshot	0.27×10^{-7}

Tasarlanan 3B yazıcının performansı PID, BOA tabanlı PID ve İGKO tabanlı PID kontrolcü algoritmaları ile deneysel olarak test edilmiştir. Bu üç algoritma, sırayla tasarlanan 3B yazıcı üzerine gömülmüş ve protez kulak ve burun modelleri yazdırılmıştır. Klasik PID kontrolcüye sahip 3B yazıcı ile yazdırılan protez kulak ve burun modelleri sırasıyla Şekil 8 ve Şekil 9'da verilmektedir. BOA tabanlı uyarlanabilir PID kontrolcüye ve önerilen İGKO tabanlı uyarlanabilir PID kontrolcüye sahip 3B yazıcı ile yazdırılan protez kulak ve burun modelleri sırasıyla Şekil 10 ve Şekil 11'de gösterilmektedir.
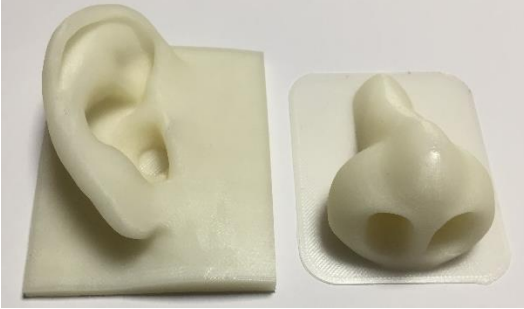

Şekil 8. Klasik PID kontrolcüye sahip 3B yazıcı ile basılan protez kulak modelleri [16]

Şekil 9. Klasik PID kontrolcüye sahip 3B yazıcı ile basılan protez burun modelleri [16]
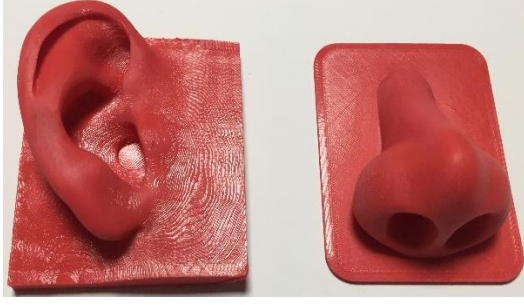
Şekil 10. BOA-PID kontrolcüye sahip 3B yazıcı ile basılan protez kulak ve burun modelleri [16]

Şekil 10. İGKO-PID kontrolcüye sahip 3B yazıcı ile basılan protez kulak ve burun modelleri

Çalışmada önerilen optimizasyon algoritmasının performansı çalışma [16]'da kullanılan algoritma ile karşılaştırıldığında, İGKO algortimasının yakınsama hızının BOA'nın yakınsama hızına göre daha iyi olduğu, yerel minimumlardan kaçınabildiği ve global optimum başarımının daha yüksek olduğu deneysel olarak doğrulanmıştır.

Önerilen kontrolcü algoritması ve karşılaştırılan diğer kontrolcü algoritmaları ile yazdırılan yapay organlardaki deformasyon miktarları Pearson korelasyonu ile belirlenmiştir. R, ölçülen ve istenen değerler arasındaki doğrusal ilişkinin düzeyinin bir ölçümünü sağlar. R değeri, hedefler ve çıktılar arasındaki ilişkinin bir göstergesi olarak tanımlanır. $R = 1$ ise, hedefler ve çıktılar arasında tam bir doğrusal ilişki vardır. R sıfıra yakınsa, hedefler ve çıktılar arasında doğrusal bir ilişki yoktur. R değeri [18];

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (13)$$

ile elde edilir. Burada, $\bar{x}$ ve $\bar{y}$ sırasıyla ortalama hedef değerler ve ölçülen değerlerdir. Klasik PID kontrolcüye sahip 3B yazıcı ile basılan protez kulak ve burun modelleri için deformasyon oranları sırasıyla ortalama % 19.76 ve % 17.62'dir. BOA tabanlı PID kontrolcüye sahip 3B yazıcı [16] ile basılan protez kulak ve burun modelleri için deformasyon oranları sırasıyla % 0.673 ve % 0.349'dur. Önerilen İGKO tabanlı uyarlanabilir PID kontrolcü algoritmasına

sahip 3B yazıcı ile basılan protez kulak ve burun modelleri için deformasyon oranları sırasıyla % 0.024 ve % 0.013'tür.

VI. SONUÇLAR

Bu çalışmada, kulak ve burun gibi yapay doku ve organların biyo-baskıları için parametreleri meta-sezgisel optimizasyon algoritmalarından İGKO ile kestirilen yeni bir uyarlanabilir PID kontrolcü geliştirilmiştir. Geliştirilen kontrolcü algoritması özgün olarak tasarlanan 3B yazıcı üzerinde test edilmiştir. İGKO algoritması, en iyi PID kontrolcü parametrelerinin ayarlanması için ITSE performans kriterlerine göre adım adım çalıştırılmıştır. Deneysel çalışmalardan elde edilen sonuçlar, önerilen yaklaşımın PID kontrolcünün parametrelerini hızlı ve etkili bir şekilde hesapladığını göstermiştir. Tasarlanan 3B yazıcıda kulak ve burun protez modelleri yazdırılarak yazıcının performansı deneysel olarak test edilmiştir. Geliştirilen İGKO tabanlı uyarlamalı PID algoritması ile elde edilen sonuçlar literatürdeki BOA tabanlı uyarlamalı PID ve klasik PID kontrolcü sonuçları ile karşılaştırılmıştır. Tasarlanan 3B yazıcı ile birlikte önerilen İGKO tabanlı uyarlanabilir kontrolcünün protez kulak ve burun üretim sürecindeki hataları belirgin biçimde azalttığı gözlemlenmiştir. Özellikle protez doku ve organların iç içe geçmiş kompleks geometrilere sahip baskılarında önemli bir sorun olan yüzey hataları önerilen algoritma ile en aza indirilmiştir. Gelecek çalışmalarda canlı yapıların, doku ve organların oluşturulabilmesi için canlı hücrelerin minimum hata ile konumlandırılmasını sağlayacak yapay zekâ temelli algoritmalar tasarlanan 3B yazıcının kontrolünde kullanılacaktır. Bu sayede, canlı hücrelerin hassas konumlandırılmasıyla hedef doku veya organların oluşturulması esnasında meydana gelen yüzey alanı ve üretim kusurları en aza indirilebilecektir.

TEŞEKKÜR

KAYNAKLAR

[1] Nuseir, A., Hatamleh, M. M. D., Alnazzawi, A., Al- Rabab'ah, M., Kamel, B., & Jaradat, E., (2019). Direct 3D printing of flexible nasal prosthesis: optimized digital workflow from scan to fit. *Journal of Prosthodontics*, 28(1), 10-14.

[2] Zhong, N., & Zhao, X., (2017). 3D printing for clinical application in otorhinolaryngology. *European Archives of Oto-Rhino-Laryngology*, 274(12), 4079-4089.

[3] Di Gesù, R., Acharya, A. P., Jacobs, I., & Gottardi, R., (2020). 3D printing for tissue

engineering in otolaryngology. *Connective Tissue Research*, 61(2), 117-136.

[4] Farook, T. H., Jamayet, N. B., Abdullah, J. Y., Rajion, Z. A., & Alam, M. K., (2020). A systematic review of the computerized tools and digital techniques applied to fabricate nasal, auricular, orbital and ocular prostheses for facial defect rehabilitation. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 121(3), 268-277.

[5] Bauermeister, A. J., Zuriarrain, A., & Newman, M. I., (2016). Three-dimensional printing in plastic and reconstructive surgery: a systematic review. *Annals of Plastic Surg*ery, 77(5), 569-576.

[6] Nemli, S.K., Aydin, C., Yilmaz, H., Bal, B.T., & Arici, Y.K., (2013). Quality of life of patients with implant-retained maxillofacial prostheses: a prospective and retrospective study. *The Journal of Prosthetic Dentistry*, 109(1), 44-52.

[7] Irish, J., Sandhu, N., Simpson, C., Wood, R., Gilbert, R., Gullane, P., Brown, D., Goldstein, D., Devins, G., & Barker, E., (2009). Quality of life in patients with maxillectomy prostheses. *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck*, 31(6), 813-821.

[8] Levine, E., Degutis, L., Pruzinsky, T., Shin, J., & Persing, J. A., (2005). Quality of life and facial trauma: psychological and body image effects. *Annals of Plastic Surgery*, 54(5), 502-510.

[9] Horlock, N., Vögelin, E., Bradbury, E. T., Grobbelaar, A. O., & Gault, D. T., (2005). Psychosocial outcome of patients after ear reconstruction: a retrospective study of 62 patients. *Annals of Plastic Surgery*, 54(5), 517-524.

[10] Hatamleh, M. M., Haylock, C., Watson, J., & Watts, D. C., (2010). Maxillofacial prosthetic rehabilitation in the UK: a survey of maxillofacial prosthetists' and technologists' attitudes and opinions. *International Journal of Oral and Maxillofacial Surgery*, 39(12), 1186-1192.

[11] Watson, J., & Hatamleh, M. M., (2014). Complete integration of technology for improved reproduction of auricular prostheses. *The Journal of Prosthetic Dentistry*, 111(5), 430-436.

[12] Hatamleh, M. M., & Watson, J., (2013). Construction of an implant- retained auricular prosthesis with the aid of contemporary digital technologies: a clinical report. *Journal of Prosthodontics: Implant, Esthetic and Reconstructive Dentistry*, 22(2), 132-136.

[13] Altan, A., & Hacıoğlu, R., (2018). The algorithm development and implementation for 3D printers based on adaptive PID controller. *Journal of Polytechnic*, 21(3), 559-564.

[14] Mirjalili, S., Mirjalili, S. M., & Lewis, A., (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, 46-61.

[15] Wang, X., Zhao, H., Han, T., Zhou, H., & Li, C., (2019). A grey wolf optimizer using Gaussian estimation of distribution and its application in the multi-UAV multi-target urban tracking problem. *Applied Soft Computing*, 78, 240-260.

[16] Altan, A., & Parlak, A. (2020). Adaptive Control of a 3D Printer using Whale Optimization Algorithm for Bio-Printing of Artificial Tissues and Organs. *IEEE Innovations in Intelligent Systems and Applications Conference*, October, 1-5.

[17] Nadimi-Shahraki, M. H., Taghian, S., & Mirjalili, S., (2021). An improved grey wolf optimizer for solving engineering problems. *Expert Systems with Applications*, 166, 113917.

[18] Tan, P. N., Steinbach, M., & Kumar, V., (2005). Introduction to Data Mining. *Pearson Addison Wesley.*

Speaker Accent Recognition Using MFCC Feature Extraction and Machine Learning Algorithms

MFCC Öznitelik Çıkarım Tekniği ve Makine Öğrenmesi Algoritmaları Kullanılarak Konuşmacı Aksanı Tanıma

Ahmet Aytuğ AYRANCI [1,2] (iD) , **Sergen ATAY²** (iD) , **Tülay YILDIRIM²** (iD)

[1] *İstanbul Kültür University, Electrical and Electronics Engineering, 34722, İstanbul, Türkiye*
[2] *Yıldız Technical University, Electronics and Communication Engineering, 34722, İstanbul, Türkiye*

Abstract
Speech and speaker recognition systems aim to analyze parametric information contained in the human voice and recognize it at the highest possible rate. One of the most important features in the voice for successful speaker recognition is the speaker's accent. Speaker accent recognition systems are based on the analysis of patterns such as the way that the speaker speaks and the word choice he uses while speaking. In this study, the data obtained by the MFCC feature extraction technique from voice signals of 367 speakers with total 7 different accents were used. The data of 330 speakers in the data set were taken from the "Speaker Accent Recognition" dataset in the UC Irvine Machine Learning (ML) open data repository. The data of the other 37 speakers were obtained by converting the voice recordings in the "Speaker Accent Archive" dataset created by George Mason University into data using the MFCC feature extraction technique. 9 ML classification algorithms were used for the designed speaker accent recognition system. Also, the k-fold cross-validation technique was used to test the data set independently. In this way, the performance of ML algorithms is shown when the dataset is divided into a k number of parts. Information about the classification algorithms used in the designed system and the hyperparameter optimizations made in these algorithms are also given. The performances of the classification algorithms are shown with using performance metrics.
Keywords: Mel-frequency cepstral coefficients, machine learning, speaker accent recognition, feature extraction

Öz
Konuşma ve konuşmacı tanıma sistemlerinde insan sesinin içerdiği parametrik bilginin sistem tarafından analiz edilip en yüksek başarı oranında tanınması hedeflenmektedir. Konuşmacı tanımanın başarılı bir şekilde yapılabilmesi için ses içerisindeki en önemli özelliklerden bir tanesi konuşmacının aksanıdır. Konuşmacı aksanı tanıma sistemleri konuşan kişinin konuşma şekli ve konuşurken kullandığı kelime seçimi gibi örüntülerin analiz edilerek tanınmasına dayanmaktadır. Konuşmacının ses sinyalinden gerekli öznitelik bilgilerini elde etmek için Mel-Frekans Kepstral Katsayıları (MFCC) öznitelik çıkarım tekniği kullanılmıştır. Bu çalışmada 12 katsayılı MFCC tekniği ile toplamda 7 farklı aksana sahip 367 konuşmacıya ait ses sinyallerinden elde edilen veriler kullanılmıştır. Kullanılan veri setindeki 330 konuşmacıya ait veriler UC Irvine Makine Öğrenmesi (ML) açık veri kaynağındaki "Speaker Accent Recognition" veri setinden alınmıştır. Diğer 37 konuşmacının verisi ise George Mason Üniversitesi tarafından oluşturulan "Speaker Accent Archive" veri setindeki ses kayıtlarının MFCC öznitelik çıkarım tekniği kullanılarak veriye dönüştürülmesi yoluyla elde edilmiştir. Tasarlanan konuşmacı aksanı tanıma sistemi için 9 farklı ML sınıflandırma algoritması kullanılmıştır. Bunun yanında veri setini bağımsız olarak test edebilmek amacıyla k-katlamalı çapraz doğrulama tekniği kullanılmıştır. Bu sayede veri setini farklı sayıda parçalara bölerek analiz edildiğinde sergilediği performans gösterilmiştir. Kullanılan sınıflandırma algoritmaları ve bu algoritmalarda yapılan hiper parametre optimizasyonları açıklanmıştır. Sınıflandırma yapılarının elde ettiği başarı sonuçları değerlendirme ölçütleri kullanılarak gösterilmiştir.
Anahtar Kelimeler: Mel-frekans kepstral katsayıları, makine öğrenmesi, konuşmacı aksanı tanıma, öznitelik çıkarımı

I. INTRODUCTION

Studies on speech and speaker recognition systems have gained popularity over the past 50 years. Many studies have been done on speech and speaker recognition [1-2]. In speaker recognition, it is aimed to analyze the parametric sound information of the human voice and to recognize this information by computers at the highest possible accuracy. The first process performed in speaker recognition is to analyze the sound and converting the parametric information of the sound into data. Obtained data is passed through the predetermined filters and then converted into a text form so that the system can easily understand. Research has shown that any two people's voices are different from each other [3]. When distinguishing human voices, it is quite difficult to distinguish it acoustically. The main distinctive features used to distinguish speakers are the structure of the speaker's vocal cords, way of speaking, accent, gender, and age [4-6]. These features are extracted from speech signals using feature extraction techniques.

With the developing technology, the number of systems based on speech and speaker recognition is increasing day by day. The performance of speaker recognition systems is of great importance in voice-controlled systems and devices. Speech and speaker recognition systems are actively used in areas such as online banking transactions, online shopping, applications requiring database access and personal security transactions. Since these systems are critical in terms of security, accuracy is of great importance. Filters and feature extraction techniques are used on the audio signal to increase the accuracy [7-8]. Also, noise and distorting effects in the environment where the sound is obtained play a big role in successful speaker recognition.

One of the biggest impacts on the accuracy of the speaker recognition system is the correct identification of the speaker's accent. An accent is a feature that distinguishes a certain group of people from other people in terms of the way they speak. Although there have been developments in the studies conducted in recent years, the differences in the speech styles of different local groups due to social and cultural reasons cause difficulties in identifying different accents. Therefore, it is seen that a limited number of accents are used to keep the accuracy high in the speaker recognition systems. The problem of speaker recognition appears to be intrinsically a pattern recognition problem. The way speakers speak and the pattern in the choices of words they use when speaking form, the basis of recognizing the speaker's accent.

The two main tasks within speaker recognition are speaker identification and speaker verification [9-10]. Feature extraction techniques and classification methods to be used on the sound data in speaker recognition systems are of great importance. In order to obtain important information from speech, it is necessary to determine the feature extraction technique to be used first. The most widely used feature extraction techniques in the literature are Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC), and Perceptual Linear Prediction (PLP) techniques [11-12]. The first process in speaker recognition is performing the feature extraction technique. The success performance of the method to be chosen here greatly affects the performance of other processes. The most popular feature extraction technique used in speech and speaker recognition systems is the MFCC technique. The MFCC technique works in a structure similar to the auditory system that people have. Since the MFCC processes the audio signal by estimating the frequency spectrum, the accuracy of the predicted frequency spectrum is of great importance [12]. The MFCC technique logarithmically receives components above 1000 Hz while keeping the frequency components

below 1000 Hz linear to maintain important properties of the sound signal. The MFCC technique receives components above 1000 Hz logarithmically while keeping the frequency components below 1000 Hz linear to maintain important properties of the sound signal. The windowing process is applied to divide the defined signals into frames. Then, frequency components are obtained by applying the Fast Fourier Conversion (FFT) process to the windowed signals. After these processes, Mel filter bank and Discrete Cosine Transform (DCT) operations are applied. Among the feature extraction techniques, the MFCC technique appears as the highest performance technique. Apart from MFCC, PLP and LPC techniques are also commonly used feature extraction techniques in the literature.

Classification techniques are used in order to perform speaker accent recognition of the data obtained from the audio signal by applying the feature extraction technique. Speaker recognition is performed using Machine Learning (ML) and Deep Learning (DL) techniques on the data obtained by feature extraction techniques. In our study on speaker accent recognition [13], we performed performance analysis using six ML algorithms on the dataset containing six different accent information. In [14], the authors conducted an accent recognition study using the K-Nearest Neighbor (K-NN) algorithm. In [15], the verification of the speaker was carried out by using the I-vector technique. In [16] where MFCC, PLP, and LPC feature extraction techniques are used, the authors have made performance analysis on the speaker recognition system using the Support Vector Machines (SVM) classification algorithm. In [17], the performance of the classifier algorithms was tested using the Speaker Accent Recognition dataset. In this study, the data set is extended by adding the Turkish accent.

In this study, the Speaker Accent Recognition dataset from UC Irvine (UCI) ML dataset repository was used. Also, voice recordings of 37 Turkish participants were taken from the Speech Accent Archive dataset created by Steven H. Weinberg from George Mason University were used. The voice recordings of 37 speakers in the Speech Accent Archive dataset were converted into data using a 12-coefficient MFCC feature extraction technique. With the added data, the dataset was expanded by obtaining more speakers and accent numbers. The accuracy of the system has been investigated using various ML classification algorithms on the dataset. Also, the k-fold cross-validation technique was used to show each classification algorithm's performance with unbiased accuracy and reliability. Using seven different speech accents, an extensive speaker accent recognition study was carried out.

II. FEATURE EXTRACTION TECHNIQUES

Feature Extraction is the process of obtaining important information about the speech from a stationary and short part of the audio signal. Parametric information in the audio signal is obtained using feature extraction techniques. The most commonly used feature extraction techniques in the literature are MFCC, PLP, and LPC techniques. While choosing the feature extraction technique to be used, the structure of the audio signal and the noise in the environment where the sound is recorded should be considered. The performance of the feature extraction technique in speaker recognition also greatly affects the performance of other processes.

The choice of feature extraction technique has a big impact on the accuracy of speaker recognition. The reason for that, speaker recognition is based on the data obtained by the feature extraction technique. In this study, the MFCC technique is used. MFCC will be further discussed in detail in Section 2.1. PLP technique, one of the other widely used feature extraction techniques, aims to achieve high performance by using three important structures to remove unnecessary information in the audio signal. PLP feature extraction technique combines the critical bands, intensity to loudness compression, and equal loudness pre-emphasis structures in the extraction of parametric information from audio signal. PLP has been developed by taking the human auditory system as a model. The most important disadvantages of the PLP technique compared to MFCC and LPC feature extraction techniques are the lower computational speed and noise resistance. LPC feature extraction technique is designed with inspiration from the Human vocal tract. LPC evaluates the audio signal by approximating the formants, getting rid of its effects from the audio signal, and estimates the concentration and frequency of the left behind residue [18]. LPC performs better in situations where the speakers speak shortly and say the same words. The space between people's vocal cords and larynx produces a buzz. Frequency and intensity content varies according to pitch. The vocal paths are separated according to their resonant frequencies, and this is called formants. Using this technique, the positions of the formants in an audio signal are predictable by calculating the linear predictive coefficients above a sliding window and finding the crests in the spectrum of the subsequent linear prediction filter [18]. After the formants are obtained, the conversion is achieved by performing the reverse filtering process. LPC has a high performance similar to the MFCC technique in terms of computational speed, accuracy, and obtaining acoustic information of the sound.

2.1. Mel-Frequency Cepstral Coefficients

MFCC is a feature extraction technique designed based on the auditory systems of humans. MFCC converts the signal from the time domain to the frequency domain. Speech signals contain tones of varying frequencies MFCC computes these frequencies on the Mel scale. The Mel scale is approximately linear up to 1 kHz and logarithmic above the 1kHz threshold. Since the sensitivity of the human ear decreases after the 1 kHz threshold, this scaling is of great importance in sound signal extraction [11]. MFCC technique is a replication of the human auditory system intending to artificially implement the human's hearing principle to the computers. The relation between frequency of speech and Mel scale

$$Mel(f) = 2595 * log_{10}\left(1 + \frac{f}{700}\right) \qquad (1)$$

can be given. It is possible to see the processes taking place in the MFCC feature extraction technique on the block diagram in Figure 1.
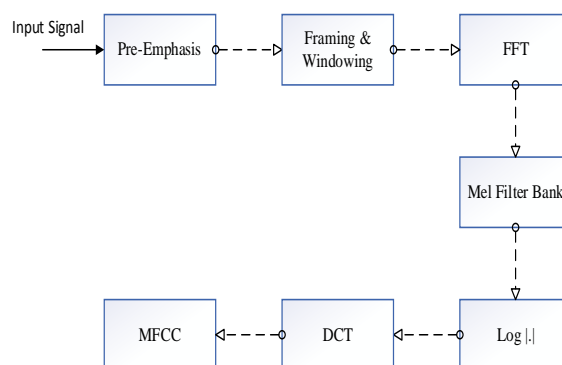

Figure 1. Block diagram of MFCC technique

Figure 1 shows the processes in the MFCC technique, respectively. In the pre-emphasis stage, the signal at high frequencies is made more pronounced. In this way, the amplitude of the high-frequency audio signal approximates the amplitude of the low-frequency audio signal and makes it suitable for comparison. This process is done by increasing the energy of the high-frequency signal. In framing, it is used to help calculate the FFT computation easily. At this stage, discontinuity is prevented. In the windowing process, each individual frame is minimized in order to prevent signal discontinuities at the beginning and end of each frame. Hamming, Hanning, Blackman, and Gauss functions are examples of commonly used windowing functions. The next step after windowing is the FFT process. Each frame is converted from the time set into a set of frequencies. This process is implemented to calculate the power spectrum. Then it is defined which frequencies are presented in which frame. One of the most important steps in the MFCC technique is the Mel filter bank. Since the human hearing ability is different in each frequency band, scaling is performed in the Mel filter bank. Since the MFCC technique is similar to the human auditory system, its sensitivity decreases as higher frequencies are reached, so it is aimed to reduce losses by taking values through a

logarithmic function in the scaling process. In the last stage, the discrete cosine transformation is made. The purpose of discrete cosine transformation is to convert the Mel spectrum back to the time domain and decorrelate the filter bank energies. The result obtained is called the MFCC. One of the most important usage advantages of MFCC is its high accuracy rate in high-frequency signals. Also, MFCC has the highest performance in computational speed and capturing the sound characteristic in speech.

III. MATERIAL AND METODOLOGY

3.1. Dataset
Success performance in speaker recognition problems depends on the dataset, feature extraction techniques, and classification algorithms used. The noise in the audio file used, the variety of accents in the data set, and the selection of the feature extraction technique are of great importance in speaker accent recognition.

In this study, the data from the Speaker Accent Recognition data set taken from the UCI ML data set source was used. Also, Turkish voice recordings in the Speech accent archive dataset created by Steven H. Weinberg were converted into data with 12-coefficient MFCC and used. Voice recordings of 37 Turkish participants were taken from the Speech accent archive dataset and converted into data. Turkish voice recordings were converted into 12-coefficient MFCC data using the Praat program and added to the dataset. The number of speakers in the dataset is 367, and there are 7 accents in total.

Table 1. Dataset information

Speaker Accent	Participant
Spanish Accent	30
French Accent	30
German Accent	30
Italian Accent	30
English Accent	45
US Accent	165
Turkish Accent	37

3.2. K-Nearest Neighbors
K-Nearest Neighbor (K-NN) is one of the simplest ML algorithms based on the Supervised Learning technique. K-NN is a widely used algorithm in classification and regression problems. K-NN algorithm was developed in 1967 by T. M. Cover and P. E. Hart. Unlike other supervised learning algorithms, the K-NN classifier does not have a

training phase. K-NN can be classified as a lazy learning algorithm because it does not have a specialized training phase. K-NN has a system based on memorizing rather than learning. It can be easily integrated into the classification problems. K-NN algorithm has high performance against noisy data. Because of this, it is widely used in speech and speaker recognition problems [13-14].

First, the k parameter is determined in the K-NN algorithm. This parameter is the number of neighbors closest to a given point. The distance of the new data from the existing data is calculated according to the selected k parameter. Choosing the optimal k value is important for the algorithm. In general, a large k value is more preferrable as it reduces the overall noise but there is no guarantee that algorithm performs better. The main distance functions used in the K-NN algorithm can be given as Euclidean and Manhattan distance functions. Formulations of Euclidian and Manhattan distance functions

$$\sqrt{\sum_{n=1}^{k}(x_n - y_n)^2} \tag{2}$$

$$\sum_{n=1}^{k}|x_n - y_n| \tag{3}$$

can be given. The formula given in Equation (2) shows the Euclidean distance formula. Here the x indicates the incoming value, while the y value indicates the center point. Euclidean and Manhattan distance functions can be used when the variables are continuous. In the instance of categorical variables, the Hamming distance should be used. The working method of the K-NN algorithm is given below.

- The k parameter to be used in the classification is determined. The chosen k parameter has a big impact on the accuracy of the algorithm.
- Distance calculations are made with the help of selected distance functions. The distance function selection is made according to the type of variables in the dataset used.
- K closest neighbors are selected from the related distances. The selected class is considered to be the class of the observation value expected to be estimated.

K-NN algorithm has a flexible and simple structure. It is one of the most widely used algorithms in studies on classification.

3.3. Multi-Layer Perceptron
Multi-Layer Perceptron (MLP) algorithm is used in classification and regression problems. MLP is a multi-layered feedforward neural network. MLP consists of an input layer, hidden layer (can be more than one) and an output layer.
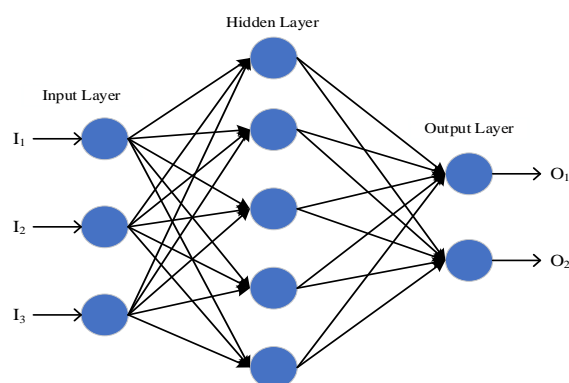
Figure 2. MLP structure

Figure 2 shows the structure of an MLP algorithm with 3 input and 2 output values. In the input layer, which is the first layer of the algorithm, data from the outside world is transmitted to the hidden layer. No operations or calculations are made in the input layer, and information is only transmitted to the hidden layer. Data from the input layer is processed in the hidden layer. There can be more than one hidden layer in the MLP structure. Information processed in the hidden layer is transmitted to the output layer. In the output layer, the data coming from the hidden layer is produced as an output.

In the MLP algorithm, there are hyper parameters that must be determined to increase the success performance. These parameters are the number of hidden layers in the algorithm, the number of iterations to be applied during the training phase, and the selection of the activation function. Determining these parameters has a big impact on accuracy. The most used activation function in the MLP algorithm is the Sigmoid function. In the MLP algorithm, all input values are multiplied and summed with the determined weight values before passing to the hidden layer. Then the obtained value is sent to the selected activation function and the output value is obtained. Accuracy of the MLP algorithm can be increased significantly with hyperparameter optimizations.

3.4. Radial Basis Function Network
Radial Basis Function (RBF) networks are an Artificial Neural Network (ANN) model developed by Moody and Darken in 1988. RBF networks are inspired by biological nerve cells. RBF networks form a special class of ANN, which consist of three layers. These layers are the input layer, hidden layer, and output layer. The feature that distinguishes RBF networks from standard ANN models is that RBF uses radial-based activation functions in the transition from the input layer to the hidden layer [19]. The hidden layer contains a number of nodes, which apply a nonlinear transformation to the input variables. The output of the RBF network is a linear combination of radial basis functions of the inputs and neuron variables.

RBF networks have high accuracy in classification problems because of the simplicity and fast training of the network architecture. In the input layer, weightless connections pass inputs to the hidden layer without any processing. The hidden layer contains a number of neurons, and each neuron consists of RBF. The hidden layer applies a nonlinear transformation to the input variables, using a radial basis function, such as the Gaussian function. The variable coming from a neuron in the hidden layer is multiplied by a weight associated with the neuron. The weights are applied to the RBF function outputs as it goes to the output layer. Output in RBF networks,

$$y_n = \sum_{k=1}^{N} w_{nk} \varphi_k(x, c_k) = \sum_{k=1}^{N} w_{nk} \varphi_k \|x - c_k\|_2 \qquad (4)$$

can be expressed. In Equation (4), y indicates the output of the network, x is the input vector of the network, and φ_k indicates the radial-based activation function. N represents the number of cells in the hidden layer and w_{nk} represents the weights determined in the output layer. Optimization of two parameters is of great importance in RBF networks. These parameters are the number of neurons in the hidden layer and their output weight. Since the number of neurons that should be used in the network structure also determines the number of RBFs to be used, it can be said that it is the most important parameter.

3.5. Decision Tree
The Decision Tree (DT) algorithm is an algorithm in the supervised learning structure used in classification and regression problems. The DT algorithm has high accuracy in complex data sets or in problems where multiple classifications are required. The structure of the DT algorithm is similar to the thought structures of humans. In this way, it can be mentioned that it is a structure that is easy to understand or make inferences. Algorithm's structure starts from the root node, goes to the inner nodes and from there to the leaf nodes, which is the final decision. The root node creates the first decision cell of the DT structure. According to the principles determined here, it is divided into inner nodes and from there into leaf nodes. The DT algorithm can be used to classify both numeric and categorical data. The creation of tree structure in numeric data sets in line with entropy values makes it difficult to understand the general structure.

Various parameters are considered when creating trees in the algorithm. While creating the DT structure, it is aimed to minimize the generalization error. The most common variants of DT algorithm used in data mining are ID3 and C4.5 algorithms. The biggest disadvantage of the DT algorithm is the problem of the over-fitting. In order to prevent this situation, restrictions or pruning can be made on parameters. Pruning the DT provides a smaller tree by removing

the lower branches containing low statistics from the structure [20].

3.6. Logistic Model Tree
The Logistic Model Tree (LMT) algorithm is a classification algorithm that combines the properties of the two widely used classification algorithms. These algorithms are namely DT algorithm with the Logistic Regression (LR) algorithm. Unlike standard DT algorithms, LMT algorithm uses Logistic Regression functions on the leaf nodes. With the Logistic Regression functions used, the Logit Boost algorithm produces an LR model on each node of the tree. The nodes are then split using the c4.5 criteria [20]. LMT algorithm makes it possible to classify binary and multiple variables in the algorithm structure. Parameter optimizations and selected Logistic Regression functions greatly impacts the accuracy of the algorithm.

LMT algorithm contains the basic features of both DT and LR algorithms. However, unlike other Decision Tree algorithms, boosting is applied in the LMT algorithm to fit the LR structure to the leaves as the tree structure grows. The biggest shortcoming of the LMT is the time required to construct the tree and LR model due to its complex structure.

3.7. Random Forest
The Random Forest (RF) algorithm is a classification algorithm developed by Leo Breiman in 2001. RF can be used in both classification and regression problems. The RF algorithm is based on the combination of bagging method and random feature selection. RF is composed of multiple DT, each with the same nodes, but using different data that leads to different leaves. RF merges the decisions of multiple DT in order to find the most probable answer.

Increasing the number of trees produced in the RF structure increases the probability of obtaining a definite result. RF is considered as an ensemble learning, such that it creates more accurate results by using multiple models to come to its conclusion. This increases the accuracy of the model since RF looking at the results of many different DT structures in order to find the most probable result. [22]. The RF algorithm provides a solution to the overfitting problem, which is the biggest problem of the DT algorithm. In addition, it can provide definitive solutions to both classification and regression problems. At the same time, the RF algorithm can be used in case of missing data in the data set or data with scattered distribution. The RF structure also has high performance against noisy data, which is one of the biggest problems in the speech and speaker recognition system. Another important feature of the algorithm is that it shows the importance of attributes. The RF algorithm uses the GINI index to measure the achievements of trees and branches. The GINI index measures the success rates of the sampling given on each node. The lower the GINI index, the more homogeneous and successful that branch can be said. The algorithm can say that if the GINI index of each sub-branch is lower than the upper branch when moving towards the branches, that branch is more successful [23].

3.8. K-Star
K-Star (K*) is an entropy-based algorithm commonly used in classification problems. It has a structure based on lazy learning, such as the K-NN algorithm. The K* algorithm uses an entropic offset function to determine the similarity between attributes. K* processes the probabilities of all attributes in a category using the sum of their probabilities. It chooses the highest probability among these possibilities. The algorithm requires the optimization of a single parameter as a structure. This parameter is a global blend parameter. Global blend parameter can take values between 0 and 100. K* algorithm's function can be obtained by,

$$K^*(b|a) = -log_2 P^*(b|a) \qquad (5)$$

using the Equation (5). The P^* expression in this equation indicates the probability function. The K* algorithm uses the Kolmogorov distance to find shortest distance between two variables. The Kolmogorov distance or test can be used to compare two one-dimensional distributions. The Kolmogorov distance uses Cumulative Distance Functions (CDF) of the data set. Using the CDF, the distribution and maximum distances between the data are found.

3.9. Logistic Regression
LR is an ML algorithm used for classification problems. LR basically is an extension of Linear Regression model for classification problems. LR classified under Supervised Learning technique. While the Linear Regression algorithm is used for solving regression problems, and the Logistic Regression algorithm is used for solving classification problems. The most widely used LR algorithm models are binary, ordinal, and multinomial. LR can be used where the probabilities between two classes is required. In Linear Regression, the data are based on estimating the outputs by finding the best regression line. In LR, instead of fitting a regression line, we fit an "S" shaped logistic or sigmoid function, which predicts between 0 and 1. This prediction cannot exceed these limits. The sigmoid function is a mathematical function used to map the predicted values to probabilities. LR is using the concept of Maximum Likelihood estimation to estimate the accuracy. According to this estimation, the observed data should be most probable.

3.10. Naive Bayes

Naive Bayes (NB) is a probabilistic ML algorithm based on the Bayes Theorem. Bayes Theorem is a simple mathematical formula used for calculating conditional probabilities. NB Algorithm is one of the simplest Classification algorithms which helps in building fast ML models that can make quick predictions. The NB algorithm classifies the data submitted to the system with probability calculations. NB algorithm has a flexible use because it requires a small amount of data for training. In studies on speaker and speaker accent recognition, it is seen that the NB classification algorithm has a low performance [13]. Considering the results obtained in this study on speaker accent recognition, it is seen that the algorithm with the lowest accuracy is NB.

3.11. K-Fold Cross Validation

The goal in the K-Fold cross-validation structure is to be able to create unbiased observation sets. Dataset is divided into k number of parts and the training set is tested for all observation sets. While the dataset is divided into k parts, all parts are expected to be of equal size and quality. In K-fold cross-validation structure, data is divided into k equal-sized folds, then the classifier was trained using the k-1 folds and tested on the remaining partition. This process is repeated k times [24]. Using the K-fold cross-validation technique, both the performance of the classification structure and the performance change according to the k parameter can be seen [25]. K-fold cross-validation shows the prediction accuracy and usability of the classifier methods. K-fold cross-validation shows unbiased prediction accuracy and generalization performance of a classifier structure. In this study, the 10-fold cross-validation method was used in order to see the performance of algorithms with unbiased.

IV. RESULTS AND DISCUSSION

Performance tests will be conducted by using 9 ML classification algorithms on the dataset. The performance of the algorithms explained through the evaluation criteria. In this section, their performance metrics will be explained first. Then the obtained algorithm results will be given.

4.1. Performance setrics

In this section, the evaluation criteria used to compare the performance of the classification algorithms will be explained. Information will be given about the criteria that show the accuracy performance of the algorithms and the criteria that show the error performance. The main parameters used to calculate evaluation criteria are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

4.1.1. Accuracy

It is the basic evaluation criterion used to see the performance of the algorithms used in classification and regression problems. Accuracy is the most basic evaluation criterion that specifies the performance of classification algorithms. The accuracy rate measures the percentage of the number of data that the algorithm correctly estimates in the total data. To calculate the accuracy rate,

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (6)$$

expression can be used. It is used to determine the algorithm's performance across classes and on the entire data set.

4.1.2. Precision

Precision shows the success rate within positive estimates. The accuracy value specifically indicates the importance of TN values. Precision value is of great importance when determining the performance of the algorithm. Precision value,

$$Precision = \frac{TP}{TP+TN} \qquad (7)$$

can be calculated using Equation (7).

4.1.3. Recall

The recall parameter measures how many of the parameters that need to be positively estimated are correctly predicted. Recall criteria are used to examine the effect of positively predicted results. The formula

$$Recall = \frac{TP}{TP+FN} \qquad (8)$$

is used for the sensitivity account.

4.1.4. Kappa statistics

Kappa statistics or coefficient is a criterion used in multiclass classification problems. Kappa Statistics are used because the accuracy, precision, and sensitivity criteria do not fully show the performance in multi algorithm cases. When examining Kappa Statistics, the confusion matrix should also be considered. Because of the observed accuracy and expected accuracy parameters should be used when calculating Kappa statistics confusion matrix should be given. The Kappa Statistics are always 1 or less than 1. In cases where kappa statistics are less than 0, it can be said that the classification structure has failed. Kappa Statistics, which has a more complex structure, can be shown as the most accurate evaluation criteria among classification structures. In order to calculate Kappa Statistics, two different probability calculations are required. These possibilities are observed probability $pr(o)$ and expected probability $pr(e)$. Kappa Statistics using these probability values,

$$\kappa = \frac{pr(o)-pr(e)}{1-pr(e)} \qquad (9)$$

expressed in the form [26].

4.1.5. F-Score

F-score is a performance metric that compares precision and sensitivity together. Together with the Kappa Statistics, it is the most useful criteria for performance evaluation when there is an imbalance between classes in the dataset. To calculate F-score;

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (10)$$

Equation (10) is used.

4.1.6. Mean Absolute Error

Mean Absolute Error (MAE) is a statistical error measure used to compare classification structures. MAE basically shows the absolute difference between two continuous variables. MAE performance metric displays the absolute distance the values in the dataset and the regression line. Because it is an easy-to-interpret criterion, it is often used to indicate performance in classification and regression problems. MAE criteria are measured regardless of the direction of errors in the forecast. MAE is a linear criterion where all individual errors are weighted equally on average. The formula,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \qquad (11)$$

is used to calculate the MAE criteria.

4.1.7. Root Mean Square Error

Root Mean Square Error (RMSE) is a widely used evaluation criteria in classification algorithms. It defines the standard deviation of difference between estimated values and true values. RMSE is also a second degree (quadratic) criteria to indicate the standard deviation of estimate errors. RMSE criteria points out the data density around regression line. As RMSE value gets close to 0, it shows that classification structure performs better. To calculate the RMSE criteria;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}} = \sqrt{MSE} \qquad (12)$$

formulation can be used. The MSE parameter represents Mean Square Error. RMSE value can be calculated via taking square root of MSE. MSE criteria is similar to RMSE criteria and defines the distance of a series variables to regression line.

4.2. Results

Advised classification structures in created data sets are tested by Weikato Environment for Knowledge Analysis (WEKA) program. Data set split into two parts as %70 education and %30 test and tested in 9 different classification algorithms. In used classification algorithms, various hyperparameter optimizations applied to achieve better success ratio. For example, learning speed set to 0.35 and

momentum set to 0.2 for MLP structure. In K* algorithm, when global blend parameter set to 30, highest success ratio achieved. When hyper parameter optimizations like this used in classification algorithms, accuracy in designed structure's success increased.

Table 2. Evaluation criteria of algorithms

Algorithms	Accuracy	Precision	Recall
MLP	**89.1**	**89.6**	**89.1**
RBF	78.18	77	78.2
K-NN	**88.2**	**89.5**	**88.2**
K*	84.55	85.6	84.5
DT	70	70.4	70
LMT	72.8	73.6	72.7
RF	81.8	83.3	81.6
LR	69.1	69	69.2
NB	57.3	73	57.3

Table 3. Kappa Statistics and F-Score of algorithms

Algorithms	Kappa Statistics	F-Score
MLP	**84**	**88.5**
RBF	67.45	76.7
K-NN	**83.35**	**88.6**
K*	77.9	84.6
DT	56.2	69.5
LMT	58.3	71.3
RF	72.7	81.4
LR	55.4	68.5
NB	47.9	58.5

The performance metrics of the algorithms are shown in table 2 and Table 3. When the tables are examined, it is seen that the algorithms that provide the highest performance are MLP and K-NN algorithms. The classification structures are expected to achieve the highest possible values, as the above evaluation criteria show the performance of algorithms.

Table 4. Error metrics of algorithms

Algorithms	MAE	RMSE
MLP	**5.3**	**17**
RBF	7.3	24
K-NN	**3.95**	**18.2**
K*	6.5	19.1
DT	10.2	28.2
LMT	10.15	22
RF	11.5	21.2
LR	9.2	24.3
NB	13.2	30.1

Table 4 shows the evaluation criteria of classification algorithms based on error performance. When the two criteria are evaluated together, it is seen that the algorithms that provide the highest error performance

are MLP and K-NN. Algorithms with the lowest error performance appear to be DT and NB algorithms.

In Figure 3, accuracy, precision and recall performance metrics of algorithms are shown. As can be seen in the figure, MLP and K-NN algorithms have shown the highest performance. Hyperparameter optimizations made in MLP, K-NN, and K* algorithms affected the algorithms' high performance.
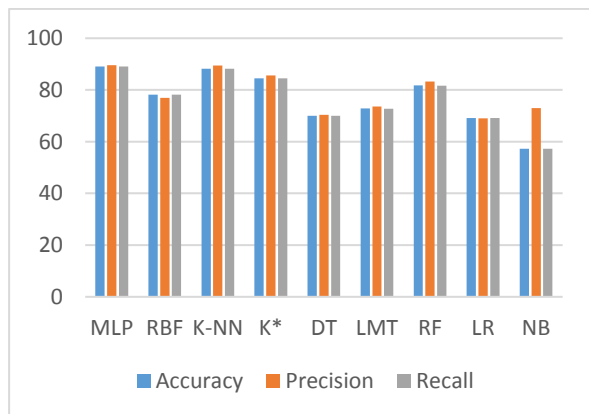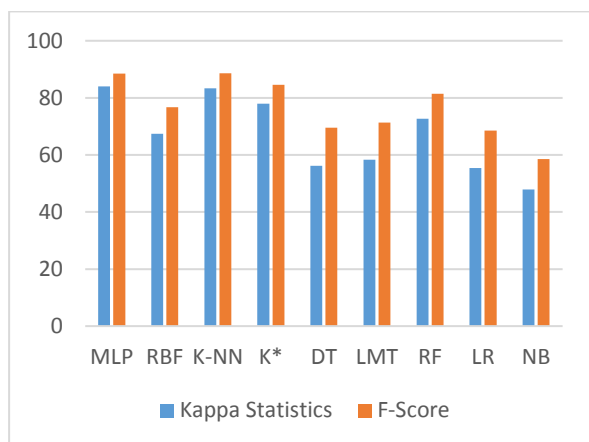

Figure 3. Accuracy, precision, and recall values

Figure 4. Kappa statistics and F-score parameters

In Figure 4, Kappa Statistics and F-Score results are shown through graphs.
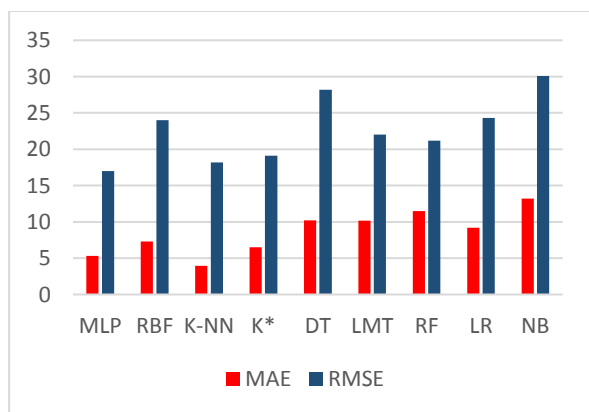
Figure 5. MAE and RMSE values of algorithms

In Figure 5, the error performance results obtained for the algorithms are shown in the graphs. Evaluation criteria are of great importance when comparing the performances of the classification structures. In addition to the evaluation criteria, a confusion matrix is used to show the performance of the classification structures and the accuracy of each class in the dataset. The confusion matrices of the classification algorithms are given below. The confusion matrices of the algorithms used are given in the tables below. The letters shown in the table (a = ES, b = FR, c = GE, d = IT, e = UK, f = US, g = TR) indicates the accents of speakers.

Table 5. Confusion matrices of algorithms.

MLP	a	b	c	d	e	f	g
a	7	0	0	0	0	2	0
b	1	9	0	0	0	0	0
c	0	0	6	0	0	1	0
d	0	0	1	3	2	1	0
e	0	0	0	0	12	2	0
f	1	0	0	0	1	54	0
g	0	0	0	0	0	0	7

a)   MLP

RBF	a	b	c	d	e	f	g
a	7	0	0	0	0	2	0
b	0	2	0	1	2	5	0
c	0	0	4	0	0	3	0
d	0	0	0	6	1	0	0
e	1	0	0	0	10	3	0
f	0	2	1	0	3	50	0
g	0	0	0	0	0	0	7

b)   RBF

KNN	a	b	c	d	e	f	g
a	8	0	0	0	0	1	0
b	0	8	1	0	1	0	0
c	0	0	6	1	0	0	0
d	0	0	1	6	0	0	0
e	0	0	0	0	12	2	0
f	0	1	2	0	3	50	0
g	0	0	0	0	0	0	7

c)   K-NN

K*	a	b	c	d	e	f	g
a	8	0	0	0	0	1	0
b	0	7	1	1	0	1	0
c	0	0	6	0	0	1	0
d	0	0	1	6	0	0	0
e	0	0	1	1	9	3	0
f	1	0	1	0	4	50	0
g	0	0	0	0	0	0	7

d)   K*

DT	a	b	c	d	e	f	g
a	5	0	0	0	1	3	0
b	0	4	0	2	1	3	0
c	0	0	6	0	0	1	0
d	0	1	2	3	0	1	0
e	1	0	1	1	6	5	0
f	1	1	3	2	1	47	1
g	0	0	0	0	1	0	6

e)   DT

LMT	a	b	c	d	e	f	g
a	6	0	0	0	0	3	0
b	0	1	3	1	1	4	0
c	0	0	4	0	0	3	0
d	0	0	2	4	0	1	0
e	0	0	0	1	9	4	0
f	0	2	3	0	0	51	0
g	0	0	0	0	0	2	5

f)   LMT

RF	a	b	c	d	e	f	g
a	6	0	0	0	0	3	0
b	0	6	0	3	0	1	0
c	0	0	4	0	0	3	0
d	0	0	2	5	0	0	0
e	0	0	0	0	9	5	0
f	0	0	0	1	2	53	0
g	0	0	0	0	0	0	7

g)   RF

LR	a	b	c	d	e	f	g
a	6	0	0	0	0	3	0
b	0	2	2	3	1	2	0
c	0	0	4	0	0	3	0
d	0	0	0	4	2	1	0
e	0	0	0	2	8	4	0
f	4	3	4	0	0	45	0
g	0	0	0	0	0	0	7

h)   LR

NB	a	b	c	d	e	f	g
a	7	2	0	0	0	0	0
b	1	2	0	7	0	0	0
c	0	0	7	0	0	0	0
d	0	0	0	6	1	0	0
e	0	1	1	1	10	1	0
f	7	9	3	6	6	24	1
g	0	0	0	0	0	0	7

i)   NB

The confusion matrices of the classification algorithms are given in Table 5. The letters in the matrices indicate the Spanish accent, the French accent, the German accent, the Italian accent, the British accent, the American accent, and the Turkish accent, respectively. Successful predictions made by algorithms for each accent can be seen over tables. In Table 2, the most successful results are obtained in MLP and K-NN algorithms. When the confusion matrix tables of these algorithms are examined (Tables 5. a, and 5. c), it is seen that these algorithms have shown more accurate results than other algorithms, and their margin of error is low. By examining the confusion matrix, it is possible to see which features of the dataset causes errors in the classification algorithm and the distribution of these errors. The accuracy of each class in the dataset also can be seen for all algorithms via the confusion matrix.

To see the performance of algorithms unbiased, k-fold cross-validation technique was used in the study. Here, the k number was chosen as 10 to obtain the most unbiased results. Table 6 shows the results obtained for 10-fold cross-validation in each algorithm used.

Table 6. Performance metrics for 10-fold cross-validation

Algorithms	Accuracy	Kappa Statistics	F-Score
MLP	**82.8**	**76.8**	**82.4**
RBF	81.15	74.5	81.1
K-NN	**82.78**	**77.1**	**83**
K*	78.15	70.8	78.8
DT	63.7	51.5	63.7
LMT	72.95	63.5	73.1
RF	82	74.9	81.5
LR	74.6	65.6	74.5
NB	60.4	52.1	60.6

When 10-fold cross-validation is used, the evaluation criterion values obtained in the algorithms are given in Table 6. The algorithms with the highest performance also appear to be MLP and K-NN algorithms.

V. CONCLUSION

A study of speaker accent recognition was conducted using the MFCC feature extraction technique and ML classification algorithms. In the dataset, a total of 7 accent data provided for 367 speakers. ML classification algorithms are used in this system to test the dataset's performance. To achieve the optimum performance in the system, hyperparameter optimizations are applied in classification algorithms. The performances of the algorithms are shown in tables and graphs using evaluation criteria.

In this study, the data set was expanded compared to [13] and different ML algorithms were added. It is seen that higher performance is obtained in the study with the extended data set. MLP and K-NN were the highest performing algorithms among the ML classification algorithms that used. The algorithms that perform the worst accuracy are DT and NB algorithms. This was valid both when the data set was used for 70% training and 30% testing and when the 10-fold cross validation technique was used. It has been observed that algorithms with the high noise resistance attribute performs better performance.

REFERENCES

[1] Van Leeuwen D. A. Martin A. F., Przybocki M. A., and Bouten J. S., (2006). NIST and TNO-NFI evaluations of automatic speaker recognition. Comput. Speech Lang., vol. 20, pp. 128–158.

[2] Furui, S. (1970). 50 Years of Progress in Speech and Speaker Recognition Research.

[3] Kinnunen T. and Li H. (2010). An overview of text-independent speaker recognition: From features to supervectors, Speech communication, vol. 52, no. 1, pp. 12–40.

[4] Nakagawa S., Wang L. and Ohtsuka S. (2012). Speaker identification and verification by combining MFCC and phase information. *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 4, pp. 1085-1095.

[5] Faria, A. (2005). Accent classification for speech recognition *proceedings of the Second Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI '05).*

[6] Turner C. and Joseph A. (2015). A wavelet packet and mel-frequency cepstral coefficients-based feature extraction method for speaker identification Procedia Computer Science, 61, pp. 416-421.

[7] De-la-Calle-Silos F. and Stern R. M. (2017). Synchrony-Based Feature Extraction for Robust Automatic Speech Recognition. in IEEE Signal Processing Letters, vol. 24, no. 8, pp. 1158-1162.

[8] Ranjan R. and Thakur A. (2019). Analysis of feature extraction techniques for speech recognition system, *International Journal of Innovative Technology and Exploring Engineering,* vol. 8, no. 7C2, pp. 197-200.

[9] Reynolds D. A. and Rose R. C. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Trans. Speech and Audio Processing,* vol. 3, no. 1, pp. 72-83.

[10] Campbell J. P. (1997). Speaker recognition: A tutorial. Proc. IEEE, vol. 85, no. 9, pp. 1437–1462.

[11] Dave N. (2013). Feature extraction methods LPC PLP and MFCC in speech recognition *Int. J. for Advance Research in Eng. and Technology,* vol. I, no. 6, pp. 1-4.

[12] Harris F. (1998). On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, vol. 66, no. 1, pp. 51-84.

[13] Ayrancı A. A., Atay S. and Yıldırım T. (2020). Speaker Accent Recognition Using Machine Learning Algorithms. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)* Istanbul, Turkey, pp. 1-6.

[14] Widyowaty D. S. and Sunyoto A. (2020). Accent Recognition by Native Language Using Mel-Frequency Cepstral Coefficient and K-Nearest Neighbor. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)* Yogyakarta, Indonesia, pp. 314-318.

[15] Alam M. J.,Kinnunen T., Kenny P., Ouellet P. and O'Shaughnessy D. (2011). Multi-taper MFCC features for speaker verification using I-vectors. *2011 IEEE Workshop on Automatic Speech Recognition & Understanding,* Waikoloa, HI, pp. 547-552.

[16] Aslan Z. and Akın M. (2019). Performing accurate speaker recognition by use of SVM and cepstral features. The International Journal of Energy and Engineering Sciences, vol. 3, no. 2, pp. 16-25.

[17] Okkan U., Dalkılıç H. (2012). Radyal Tabanli Yapay Sinir Aglari ile Kemer Barajı Aylık Akımlarının Modellenmesi. İMO Teknik Dergi, 379, 5957-5966.

[18] S. A. Alim and N. K. A. Rashid. (2018). Some commonly used speech feature extraction algorithms. in From Natural to Artificial Intelligence-Algorithms and Applications, IntechOpen.

[19] Onan A. (2015). Şirket İflaslarının Tahminlenmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi. Bilişim Teknolojileri Dergisi, vol. 8, no. 1, pp. 0.

[20] Landwehr, N., M. Hall, and E. Frank, Logistic model trees. Machine learning, (2005). 59(1-2): p. 161-205.

[21] Akar Ö. and Güngör O. (2012). Rastgele Orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması, Jeodezi ve Jeoinformasyon Dergisi, no. 106, pp. 139-146.

[22] Erdem F., Derinpınar M., Nasırzadehdızajı R., Oy S., Şeker D. and Bayram B. (2018). Rastgele Orman Yöntemi Kullanılarak Kıyı Çizgisi Çıkarımı İstanbul Örneği. Geomatik, vol. 3, no. 2, pp. 100-107.

[23] Kartal C. (2020). Modeling Bitcoin with K-Star Algorithm. bmij, vol. 8, no. 1, pp. 213-231.

[24] Mutlu, A. Y. and Yucel, O. (2018). An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. Energy, vol. 165, pp. 895–901.

[25] Elmaz F. , Yücel Ö. and Mutlu A. (2020). Machine learning based approach for predicting of higher heating values of solid fuels using proximity and ultimate analysis", *International Journal of Advances in Engineering and Pure Sciences*, vol. 32, no. 2, pp. 145-151, Jun. 2020.

[26] Landis J. R. and Koch G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1):159.

[27] Kılıç, S. (2015). Kappa Testi. Journal of Mood Disorders, 5(3), 142-144.

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Aspect Based Opinion Mining on Hotel Reviews

Otel Değerlendirmeleri Üzerinde Hedef Tabanlı Fikir Madenciliği

Semih DURMAZ[1] , **Yunus Emre DEMİR**[1] , **Ahmet ELBİR**[1] , **İbrahim Onur SIĞIRCI**[1] ,
Banu DİRİ[1]

[1] *Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul, Türkiye*

Abstract

Users often use online reviews to assess the quality of hotels according to their various attributes. In this study, a sentiment analysis of online reviews has been conducted using eleven attributes the most frequently reviewed pertaining to hotels. Using this analysis, users' overall assessments of hotels have been determined and summarized from reviews left for a group of various hotels. To identify words with similar meanings to the eleven predetermined hotel attributes, the Word2Vec method has been employed. Additionally, the FastText method has been used to detect words containing spelling errors. The sentiment analysis of the comments has been made by using three different methods belonging to two different approaches. These methods are VADER method as dictionary-based approach, BERT and RoBERTa as machine learning approaches. Using these methods, the reviews have been evaluated in three categories as positive, negative, and neutral, and the quality score has been calculated. In addition, a software with a user-friendly graphical interface has been implemented in an effort to easily use all the methods used in this study.

Keywords: opinion mining, sentiment analysis, aspect based, social media, hotel reviews.

Öz

Kullanıcılar, çevrimiçi yorumları kullanarak otelleri çeşitli özelliklerine göre değerlendirmektedirler. Bu çalışmada; oteller ile ilgili yorumlar içerisinde hakkında en çok değerlendirme yapılan on bir özellik belirlenmiş ve bu özellikleri içeren yorumların duygu analizleri yapılmıştır. Bu sayede otelin bir niteliği hakkında yapılan yorumlardan kullanıcıların genel görüşü tespit edilmiş ve özetlenmiştir. Çalışmada belirlenen on bir özelliği temsil edecek benzer anlamlı kelimelerin tespiti için Word2Vec ve yazım hataları içeren kelimelerin tespiti için FastText yöntemi kullanılmıştır. Yorumların duygu analizi, iki ayrı yaklaşıma ait üç farklı yöntem kullanılarak yapılmıştır. Birincisi, sözlük tabanlı yaklaşımlardan VADER, ikincisi makine öğrenmesi yaklaşımlarından BERT ve RoBERTa'dır. Bu yöntemler ile yorumlar; olumlu, olumsuz ve nötr olmak üzere üç kategoride karşılaştırmalı olarak değerlendirilerek nitelik skoru hesaplanmıştır. Buna ek olarak, bu çalışma kapsamında kullanılan tüm yöntemleri kolay bir şekilde uygulamak için açık kaynaklı ve kullanıcı dostu bir grafik ara yüze sahip yazılım gerçeklenmiştir.

Anahtar Kelimeler: fikir madenciliği, duygu analizi, hedef tabanlı, sosyal medya, otel yorumları.

I. INTRODUCTION

With the advent of technology and the increasing importance of the internet in human life, people's habits have undergone substantial change. Processes that previously required significant effort have been facilitated by the Internet and technology. Especially, reservations and shopping can be done quickly through the internet. The Internet also triggers people's desire to share experiences. This situation has vastly increased the number of comments on the internet.

In the past, visitors to places would write their opinions in guestbooks. These guestbooks had to be read in order to learn more about past visitors' experiences and include information on cleanliness, food quality, and other details. However, technology has enabled people to carry out such activities on a different platform. To this end, most establishments, in particular hotels and restaurants, have now transferred these operations onto the internet. In addition to online booking systems, online review systems have been put into place by establishments to ensure customers and visitors can continue to leave reviews. By using review systems, people can easily express their good or bad opinions regarding any establishment. While these reviews have an important place in terms of guiding future customers, they are also of great importance to a company to assess itself from the customer's perspective.

The sheer volume of comments shared on the internet makes it difficult to read and evaluate all of them. As a result, sentijent analysis studies are used to determine the sentiments contained in massive comment datasets. Sentiment analysis is defined as the classification and interpretation of various sentiments contained in texts.

Thanks to sentiment analysis studies, customer evaluations and feelings for establishments or their services, for which opinions and feedback are provided online, can be summarized conveniently [1]. Movie comments [2],[3], twitter comments [4], food comments [5], and hotel comments are some of the principal application areas in which sentiment analysis is performed.

In the literature, there are two basic approaches to sentiment analysis, machine learning and dictionary-based approaches. In this study, both approaches have been implemented. Dictionary-based approaches use a variety of predetermined words while evaluating a particular piece of content. The strongest aspect of this technique is that training data is not required, while the weakest property is that the number of words in the sentiment dictionary is not sufficient [6]. These words are obtained by [7] statistical and semantic techniques. In the dictionary-based approach, sentiment scores are presented by evaluating words and short contexts with various counting methods [8].

One of the most common challenges in sentiment analysis studies of comments is that any comment might include more than one sentiment. For example, customers who like the food in the hotel but do not like the cleanliness of the hotel can express these two evaluations in one sentence. In this study, sentiment analysis has been conducted relating to certain features determined to apply to hotel terminology using an English dataset. This dataset has been collected from various hotel booking sites and includes user comments about hotels.

The following part of this article is organized as follows. In Part II, comprehensive information about the Word2Vec, FastText, VADER, BERT, and RoBERTa techniques used in the study is provided. In Part III, the dataset used, and the flow of the proposed method are explained in detail and the performance results are demonstrated. An evaluation of the results and proposed method, as well as information about future studies are provided in section IV.

II. METHODS
In this section, the Word2Vec, FastText, VADER, BERT, and RoBERTa methods, which constitute the milestones of the study, are explained, respectively.

2.1. Word2Vec
Word2Vec [9] consists of a trained two-layer neural network that represents words in vector space according to their linguistic context. With the help of Word2Vec, the distance between words can be calculated vectorially [10]. In this way, words and analogies that are closest to a specified word in context can be found. There are two different Word2Vec approaches, skip-gram and CBOW. In the skip-gram approach; the input is the target word, and the outputs are the neighboring words of the target word in the sentence. In the CBOW approach, the input is the target word for adjacent words in the sentence, while the output is the target word [11]. With the CBOW approach, when a word is given its neighboring words, it is provided to predict itself. The CBOW approach is used in this study because it requires less computational complexity.

2.2. FastText
FastText was developed by Facebook in 2016 as an extension of the Word2Vec method [12]. Instead of giving words to an artificial neural network, it gives them in chunks with the letters *n*. In this approach, also called the *n*-gram model, the number of *n* indicates how many times the word will be divided. The fragmentation of the words increases the number of data, which results in the duration of the training. Thanks to the *n*-gram approach; vector representations can also be obtained for words that are caused by spelling errors and that do not actually exist [13]. In this study, the semantically adjacent words of a given word have been determined by using trained models.

2.3. VADER
VADER (Valence Aware Dictionary and sEntiment Reasoner) [14] is a dictionary and rule-based sentiment analysis tool prepared in accordance with the sentiments expressed in social media. By using VADER, we can learn whether a sentence is positive or negative. When analyzing sentiments, the use of words, punctuation marks and emoji are also considered to make the results more precise. Since VADER is a dictionary-based solution, it does not need training data and provides fast results. In addition to sentiment analysis, information about the degree of positivity of the sentence is also obtained by VADER. With this feature, a degree in the range of [-1, +1] is presented [15]. The negativity of the sentence increases as this degree approaches -1, while the positivity of the sentence increases as it approaches +1.

2.4. BERT
BERT (Bidirectional Encoder Representations from Transformers) algorithm is developed by Google to be used for many different NLP tasks, such as Classification, Question Answering, Sentiment Analysis etc [16]. BERT was trained on Wikipedia and Bookcorpus, more than 3 billion words [17]. It obtained the best accuracy ratio for some of the NLP tasks. In this study, BERT will be used to decide whether the review better has a positive, neutral or negative meaning. The BERT model which is used during this study was fine-tuned for sentiment analysis on product reviews in six languages. It was trained with 150k comments in English. BERT contains lots of pretrained models trained by different people on different datasets. The model of *bert-base-multilingual-uncased-sentiment* is used during this study [18]. This model is fine-tuned for sentiment

analysis on reviews. It gives a result as the sentiment of the review as a number of stars (between 1 and 5).

2.5. RoBERTa

RoBERTa (Robustly optimized BERT approach) is a language model developed by the Facebook AI team [19]. It was built on BERT's language masking strategy. RoBERTa allows it to improve the masked language modeling objective that helps to achieve better performance by modifying the basic hyper-parameters in the BERT model [17, 20]. RoBERTa is a better version of BERT by using 10 times more data and computing power. RoBERTa, just like BERT, contains lots of pretrained models trained by different people on different datasets. The model of *twitter-roberta-base-sentiment* is used during this study [21]. This model trained on 58M tweets and fine-tuned for sentiment analysis with the TweetEval benchmark. It gives a result as the sentiment of the review as a label where Label0 is negative, Label1 is neutral and Label2 is positive.

III. PROPOSED METHOD AND RESULTS

In this section, the flow of the proposed method is expressed by introducing the dataset used and implementation of sentiment analysis. The flowcharts of proposed methods are shown in Figure 1. Phase 1 shows the determination of the attribute set. Phase 2 illustrates the step of making sentiment analysis
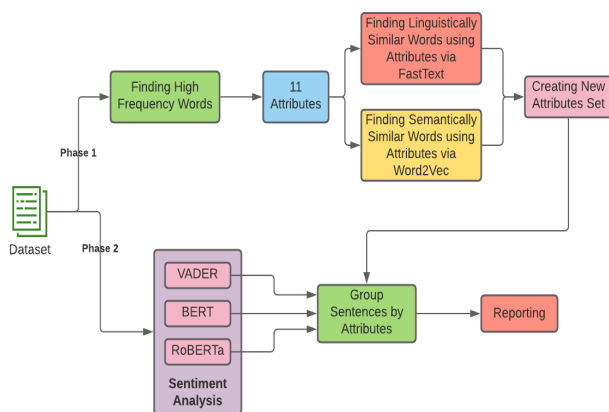
Figure 1. Flowchart of proposed method.

3.1. Dataset

In this study, approximately 27329 reviews written online for the 10 most expensive hotels in London have been used as the dataset. The dataset has been obtained from *kaggle.com* [21]. Since the dataset is suitable for the purpose of the study, it has been considered sufficient. Firstly, some of pre-processing operations has been implemented since it had been not cleaned. 431 of the comments contain blank lines, and 3350 of them are written in a language other than English. These erroneous comments, non-English letters, and word groups such as "på, ich, wir, des, ò,

ci, Bij, piu, Che dire, Ci, è, un, ed, ó, á, ä, å, di, ç, ğ, ş, ö, ü" have been removed. Comments exceeding 2000 letters in length have been removed from the dataset. As a result of all these data preprocessing operations, 22075 comments have been selected to work on. Moreover, the dataset includes the evaluation score, or "Star" rating given by the reviewer as an integer out of 5, as well as which hotel the reviews are for. Thus, the accuracy of the methods has been calculated.

3.2. Sentiment Analysis

The VADER sentiment analysis tool provides positive, negative, and neutral scores of a sentence given as input such that the total of them is 1.0. Also, it gives sentimental level information in the range of $[-1, +1]$. The outputs on a piece of sample text can be seen in Figure 2. According to this output, while the sentence is a neutral sentence at a rate of 63.3%, it is a positive sentence at a rate of 36.7%. In addition, the degree of positivity is very high given its proximity to +1 as 0.9583.

Figure 2. VADER sample output

The BERT analyzes the sentiment and provides the number of stars between 1 and 5 where 5-star indicates the highest positive sentiment and 1-star indicates just the opposite. It also gives sentimental level information in the range of $[-0, 1]$ as "score". The higher score means higher stability in the given number of stars. According to the output shown in Figure 3, the label is 5 stars, that means the sentence is a positive sentence and score is 0.851 that shows the review deserved 5 stars with a stability rate of 85.1%.

Figure 3. BERT sample output

The RoBERTa analyzes the sentiment and provides a label about the sentiment of the sentence. There are three possible labels as an output of RoBERTa, these are: *label0*, *label1* and *label2*. Label2 indicates the sentence has a positive sentiment, label1 indicates that it has a neutral sentiment, and label0 indicates that it has a negative sentiment. RoBERTa also gives stability information in the range of $[-1, +1]$ as "score". The higher score means higher stability in the given number of stars. According to the output shown in Figure 4, the label is Label2 which means the sentence is a positive sentence and score is 0.989 that shows the review is positive with a stability rate of 85.1%.

Figure 4. RoBERTa sample output

3.3. Hotel Attributes

In this study, the words in the reviews have been ordered according to their frequencies of use. While only the words related to the hotel have been selected, those which are not relevant have been removed from the list. For instance, although the names of the cities are mentioned very often, they have been excluded from the list because they are not related to the hotel. After the ranking mentioned above, the top 11 words have been determined as hotel attributes in this study. The frequencies of these selected words are shown in Table 1. The reason why the number of features is eleven is that other frequently repeated words in the list have close meanings with these eleven words and they have included in the same cluster. Additionally, statistical, and unsupervised learning methods can be used to detect attributes in such studies, so the number of features may vary according to different selection methods.

Table 1. Frequency of selected hotel features

Attribute	Frequency
Room	39174
Staff	18214
Service	13370
Breakfast	12217
Location	8806
Restaurant	7014
Bed	6015
Bathroom	5377
Food	5368
View	4586
Hotel	4392

Since these eleven detected keywords can be expressed with synonyms in different comments, we sought to identify words that could be synonymous with them. Similar 10 words have been determined using the Word2Vec approach in line with this goal. The FastText model also has been used to detect linguistically similar 10 words and find similar words since it is sensitive to spelling errors. Similar words obtained with the help of these models are shown in Table 2.

Table 2 shows that the Word2Vec method focuses on synonyms, while the FastText method focuses on spelling errors. As an example, when the word "breakfast" is examined, words resembling the word breakfast have been found in the Word2Vec approach; in the FastText method, typos such as "breakfats, breakfat, brekfast, breakfeast, breafast" have been found as the closest words. By examining the words obtained by both methods, a new list has been prepared by selecting the words related to a certain attribute.

In Figure 5, blue rows show the number of the reviews that contain the related attribute only. On the other hand, orange rows show the number of the reviews that contain the related attribute and related words that are found with the help of Word2Vec and FastText. As it can be seen in Figure 5, with the addition of related words, 15.45% more reviews became available for analyzing.

3.4. Reporting

The evaluations made in this study have been reported for each of the hotels according to their attributes. Figure 6 shows an example of this reporting by using VADER. In the Figure 6, selected eleven attributes of any hotel and their positive and negative review rates are presented on a pie chart. It is easy to observe which features of the hotel are good and which of them are bad. For example, 4% of the comments on the "Staff" attribute are negative, 96% of them are positive. In addition, by means of the software implemented in this study, all reviews of the relevant attribute can be viewed comprehensively by clicking on any selected graphic.

III. CONCLUSION

In the study, sentiment analysis has been conducted for the 10 most expensive hotels in London related to various attributes determined by using online comments. The attributes have been reduced to eleven by selecting them as keywords from among the most frequent words in the dataset. Then Word2Vec, which gives synonyms of the eleven keywords, and FastText, which ignores typos and finds similarities, have been applied to the data. The comments have been evaluated according to their qualities by using the words selected among the words determined by these approaches. By increasing the eleven keywords with Word2Vec and FastText methods, a total of 15.45% more comments have been evaluated. Thus, instead of analyzing an average of 7162 comments per feature, 8268 comments have been analyzed. The VADER, BERT and RoBERTa methods has been used to analyze the sentiments of the comments. When making a comparison between these three methods, a three-categorized structure with a result close to user scores has been evaluated: positive, negative, and neutral. When the results found are compared with the user scores, accuracy score was used to calculate the success ratio, VADER's success is 91%, BERT's success is 89.2% and RoBERTa's success is 92.6% as shown in Table 3. To understand why the RoBERTa is more successful, it's important to search how it has been trained. RoBERTa model which is used for this study is trained on 58M tweets while the BERT model used is trained on 500K reviews. As a result, the fact that RoBERTa has been trained with more datasets in both pre-training and fine-tuned stages is considered the most important factor increasing its success. By conducting a sentiment analysis with all methods used in this study on the comments, customers' sentiments

pertaining to the specific nature of a hotel has been determined. In this way, reports have been made about the sentiment analysis of all comments regarding the hotels, as well as regarding the feelings of the customers as they relate to the hotels' attributes in particular.

In future studies, we plan to use up-to-date deep learning approaches relating to context for sentiment analysis and detection of close words. In this way, we aim to increase the number of comments that can be analyzed. We expect that the performance of sentiment analysis techniques will improve further as a result.

Table 2. Similar words table for 11 selected words

| Category	Word2Vec	FastText	Selected words
Hotel	property, accommodation, place, establishment, accomodation, hotels, city, comparison, london, stay	hotels, otel, hotelrooms, whatahotel, motel, hotelier, hotelroom, rhodeshotel, property, hote	hotels, otel, hotelrooms, motel, hotelier, hotelroom, property, accommodation, place, establishment, accomodation, building
Staff	personnel, employee, team, everyone, informative, professional, approachable, chatty, incredibly, genuinely	staffed, staffer, staf, naff, barstaff, waitstaff, stafford, quaff, doorstaff, raff	staffed, staffer, staf, waitstaff, personnel, employee, everyone, team
Location	position, attraction, locate, situate, shopping, proximity, buss, neighborhood, center, subway	allocation, localization, position, occation, education, cation, staycation, located, locate, disposition	position, locate, located, center
Room	bedroom, double, bed, executive, functional, amenity, sufficiently, adequately, bathrooms, deluxe	rooom, roomy, inroom, zoom, roomier, broom, roooms, groom, wetroom, badroom	rooom, inroom, bedroom, cupboard, twin
Breakfast	cereal, eggs, croissant, omelet, buffet, cooked, continental, yoghurt, freshly, cook	breakfats, breakfat, brekfast, breakfeast, breafast, breakast, breakfest, breakdown, bfast, breakout	breakfeast, breakast, breakfest, eggs, cereal, continental, buffet, croissant, omelet, pancake, egg
Bed	mattress, pillow, bedding, duvet, chair, blanket, soundly, couch, silent, armchair	bedbugs, bedded, beds, bedbug, robbed, fobbed, bedeck, bedsheets, grabbed, bedskirt	beds, bedsheets, mattress, pillow, bedding, chair, duvet, topper
Service	sevice, consistently, presentation, skill, focus, approachable, attentiveness, thorough, fulfilling, staff	serviced, servico, serviceminded, seervice, disservice, servicing, roomservice, serviceable, serving, setvice	serviced, serviceminded, servicing, roomservice, serving, presentation, sevice
Bathroom	bathrooms, bath, bathtub, tub, linen, fixtures, vanity, closet, furnishing, dressing	bathrooom, bathrooms, bathrom, bathroon, batrooms, bathrobe, bathrobes, baths, washroom, bathe	bathrooms, bathrobe, bathrobes, baths, washroom, bath, bathtub, plasma, rainfall, furnishing, tub, hairdryer
View	overlook, facing, veiw, veiws, partial, cityscape, overlooked, glimpse, escellent, patio	views, vieuw, vie, viewed, viewing, vienna, overview, viewpoint, vi, vii	views, viewed, viewing, overview, overlook, facing, outlook, glimpse
Food	meal, dish, cuisine, massimo, menu, risotto, steak, ingredient, presentation, seafood	foodies, foodie, seafood, fod, foodhall, meal, menu, hood, oatmeal, menus	foodies, seafood, meal, menu, dish, risotto, presentation, burger, lamb, cuisine
Restaurant	restaurants, boulud, restuarant, resturant, eatery, boloud, pierino, lebanese, cafe, resturants	restaurants, restaurante, restauraunt, restauarant, restauarants, resaurant, restraurants, restaurent, resturant, reataurants	restaurants, resturant, boulud, restuarant, boloud, massimo, kaspers, grill
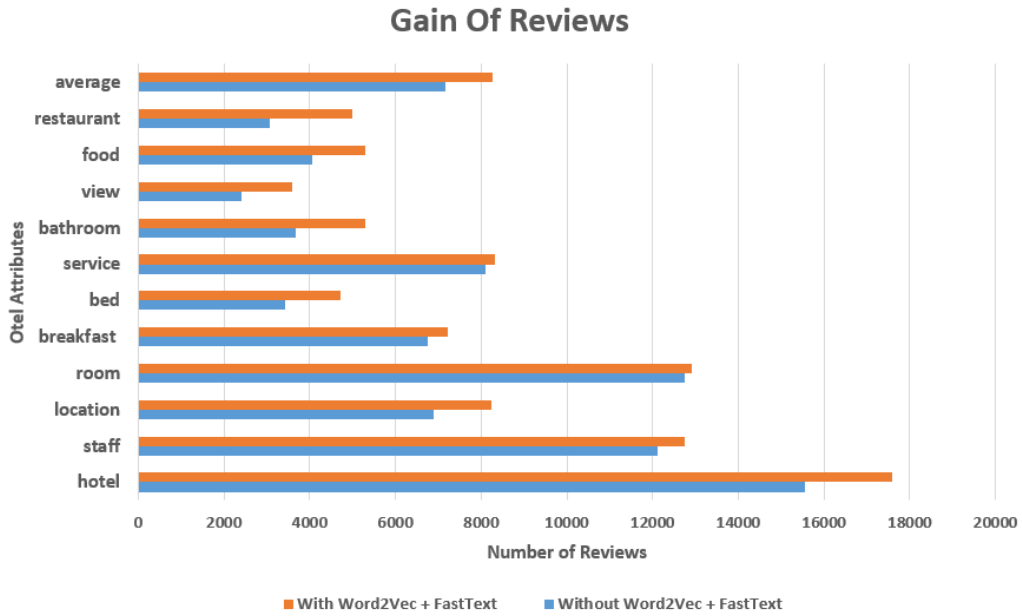
Figure 5. Number of comments that can be analyzed with/without Word2Vec and FastText
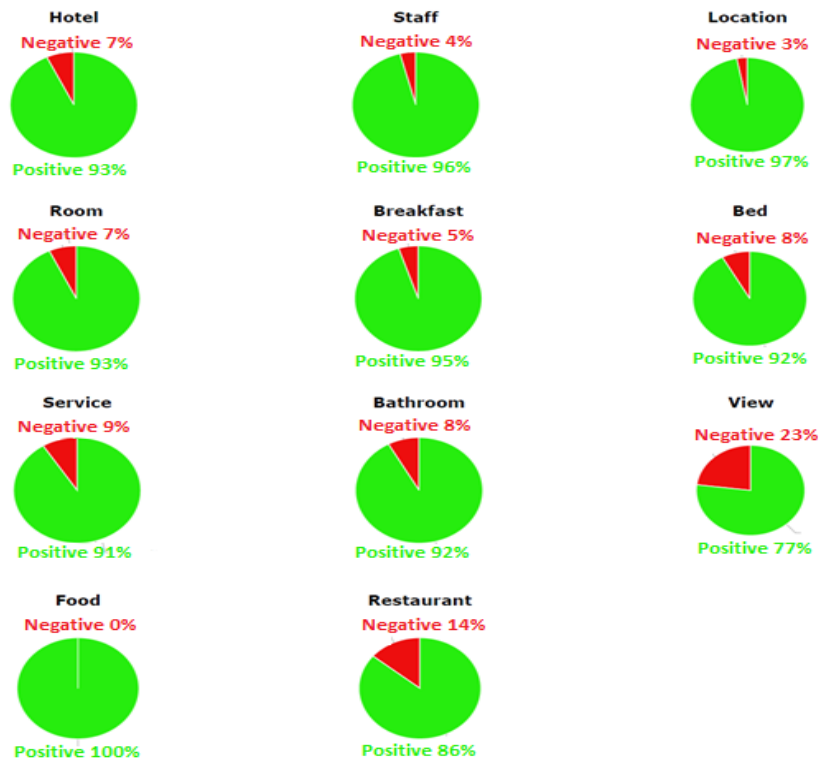

Figure 6. Sample reporting for hotel qualifications

Table 3. Confusion matrix and accuracy between VADER, BERT, and RoBERTa

		VADER			BERT			RoBERTa		
		Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu	Pos
USER	**Neg**	423	74	448	795	105	45	764	97	84
	Neu	169	81	955	441	430	334	407	180	618
	Pos	159	171	19595	501	945	18479	213	197	19515
Success Ratio:		91.05%			89.2%			**92.6%**		

REFERENCES

[1] Sentiment analysis, https://monkeylearn.com/sentiment-analysis/, (2020)

[2] Eroğul, U. (2009). Sentiment analysis in Turkish (Master's thesis).

[3] Vural, A. G., Cambazoglu, B. B., Senkul, P., & Tokgoz, Z. O. (2013). A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In Computer and Information Sciences III (pp. 437-445). Springer, London.

[4] Aytuğ, O. N. A. N. (2018). Sentiment analysis on Twitter based on ensemble of psychological and linguistic feature sets. Balkan Journal of Electrical and Computer Engineering, 6(2), 69-77.

[5] Nizam, H., & Akın, S. S. (2014). Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması. XIX. Türkiye'de İnternet Konferansı, 1-6.

[6] Symeonidis. S, https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html, (2018)

[7] Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971.

[8] Kan. D, Sentiment analysis, https://www.quora.com/What-is-the-difference-between-the-corpus-based-approach-and-the-dictionary-based-approach-in-sentiment-analysis, (2020)

[9] Ling, W., Dyer, C., Black, A. W., & Trancoso, I. (2015). Two/too simple adaptations of word2vec for syntax problems. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 1299-1304).

[10] Wang, H. (2014). Introduction to Word2vec and its application to find predominant word senses. URL: http://compling. hss. ntu. edu. sg/courses/hg7017/pdf/word2vec% 20and% 20its% 20appli cation% 20to% 20wsd. pdf.

[11] Enríquez, F., Troyano, J. A., & López-Solaz, T. (2016). An approach to the use of word embeddings in an opinion classification task. Expert Systems with Applications, 66, 1-6.

[12] What is fasttext? Are there tutorials? , https://fasttext.cc/docs/en/faqs.html, (2020)

[13] Fivez, P., Suster, S., & Daelemans, W. (2017, August). Unsupervised context-sensitive spelling correction of clinical free-text with word and character n-gram embeddings. In *BioNLP 2017* (pp. 143-148).

[14] Pandey, P. (2018). Simplifying sentiment analysis using VADER in Python (on social media text). *Retrieved from Analytics Vidhya website: https://medium. com/analytics-vidhya/simplifying-socialmedia-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f.*

[15] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 8, No. 1).

[16] Horev, R. (2018). BERT Explained: State of the art language model for NLP. *Towards Data Science, Nov*, *10*.

[17] Bert Jadhav, S. A. (2020). Detecting Potential Topics In News Using BERT, CRF and Wikipedia. arXiv preprint arXiv:2002.11402.

[18] https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment

[19] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[20] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

[21] https://www.kaggle.com/PromptCloudHQ/reviews-of-londonbased-hotels

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

A Tertiary Study and Social Network Analysis on Agile Software Development Methodology

Çevik Yazılım Geliştirme Metodolojisi Üzerine Bir Üçüncül Çalışma ve Sosyal Ağ Analizi

Egemen BAYRAM [1] (iD), **Buket DOĞAN** [1] (iD), **Volkan TUNALI** [2] (iD)

¹Marmara University, Faculty of Technology, Department of Computer Engineering, 34722, Kadikoy/ISTANBUL

²Maltepe University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, 34857, Maltepe/ISTANBUL

Abstract

In recent years, there has been an increasing interest in research and application of agile software development. In academic literature, the number of systematic literature reviews (SLRs), systematic mapping studies, and unsystematic reviews has also increased. One of the aims of this article is to create a tertiary study of SLRs on agile software development research topics between 2013 and 2018. Our second goal is to provide an in-depth analysis with data including reference, author, institution, etc. belonging to this tertiary study, and to provide an in-depth analysis for software engineering researchers and practitioners to reveal the relationships of developments in this field. We applied a two-stage research method within the scope of the study. First, we obtained and examined the systematic literature reviews published in the field of agile software development from Web of Science, Science Direct, Scopus, and IEEE academic databases. Second, we visualized the publications examined within the scope of the tertiary study, and revealed the relationships between publications, researchers, institutions, publication sources, and countries with VOSviewer and Gephi social network analysis tools. The findings obtained as a result of the study can be summarized as the fact that developing countries have more studies on the subject, the cited publications are mostly from developed countries, contrary to what is expected, distant countries are more in cooperation, and the number of citations is not directly proportional to the number of publications.
Keywords: Tertiary Study, Agile Software Development, Social Network Analysis, Systematic Literature Review, Bibliometric Analysis.

Öz

Son yıllarda çevik yazılım geliştirme üzerine araştırma ve uygulama geliştirmeye olan ilgi artmaktadır. Akademik literatürde de birçok sistematik literatür taramalarının (SLT), sistematik haritalama çalışmalarının ve sistematik olmayan incelemelerin sayısı da artmıştır. Bu makalenin amaçlarından biri, 2013 ve 2018 yılları arasında çevik yazılım geliştirme araştırma konuları üzerine SLT'lerin üçüncül bir çalışmasını oluşturmaktır. İkinci hedefimiz, referans, yazar, kurum vb. bu üçüncül çalışma ve yazılım mühendisliği araştırmacıları ve uygulayıcıları için bu alandaki gelişmelerin ilişkilerini ortaya çıkarmak için derinlemesine bir analiz sağlamaktır. Çalışma kapsamında iki aşamalı bir araştırma yöntemi uyguladık. İlk olarak, çevik yazılım geliştirme alanında yayınlanmış sistematik literatür taramalarını Web of Science, Science Direct, Scopus, ve IEEE akademik veri tabanlarından elde ederek inceledik. İkinci olarak ise üçüncül çalışma kapsamında incelenen yayınları VOSviewer ve Gephi sosyal ağ analizi araçları ile görselleştirilerek, alandaki yayınlar, araştırmacılar, kurumlar, yayın kaynakları, ve ülkeler arasındaki ilişkileri ortaya koyduk. Çalışma sonucunda elde edilen bulgular gelişmekte olan ülkelerin konu hakkında daha çok çalışması olması, alıntılanan yayınların ise çoğunlukla gelişmiş ülkelerden olması, beklenenin aksine yakın ülkelerin değil uzak ülkelerin daha çok iş birliği içinde olması, çıkarılan yayın sayısı ile alıntılanmanın doğru orantılı olmadığı şeklinde özetlenebilir.
Anahtar Kelimeler: Üçüncül Çalışma, Çevik Yazılım Geliştirme, Sosyal Ağ Analizi, Sistematik Literatür Taraması, Bibliyometrik Analiz.

I. INTRODUCTION

Any scientific research usually begins with a literature review to gain a good understanding of the existing research and the discussions relevant to a particular topic or area of study. A systematic literature review is a powerful tool for identifying, evaluating, and interpreting all current research related to a particular research question or subject area. Its main objective is to provide an unbiased assessment of a research topic using a reliable, rigorous, and auditable methodology [1]. In addition, a systematic literature review provides a repeatable research method with sufficient detail that it can be reproduced by other researchers when properly followed and implemented. Besides, detailed documentation of the steps performed within the systematic literature review allows for an in-depth evaluation of the work carried out [2].

Researchers who want to carry out a systematic literature review need thoroughly to examine all available literature on the subject they are interested in. Individual studies that are subject to a systematic review are called primary studies. A systematic literature review that considers these primary studies is called a secondary study. When these systematic literature reviews are systematically scanned by a higher level methodical literature review, then it is called a tertiary study [2].

In this study, we wanted to determine the recent trends in agile software development by analyzing the related literature by using the tertiary study methodology as our primary tool. After the Agile Manifesto was declared by a group of software development experts in 2001, lots of research have examined the subject of agile software development and produced various publications on it. Agile software development applications such as Extreme Programming (XP) and Scrum have been increasingly adopted to address the challenges of volatile business environments where markets and technologies have been evolving highly rapidly. As seen in Figure 1, number of scientific studies on agile software development, which has a history of about two decades, has been increasing steadily every year.
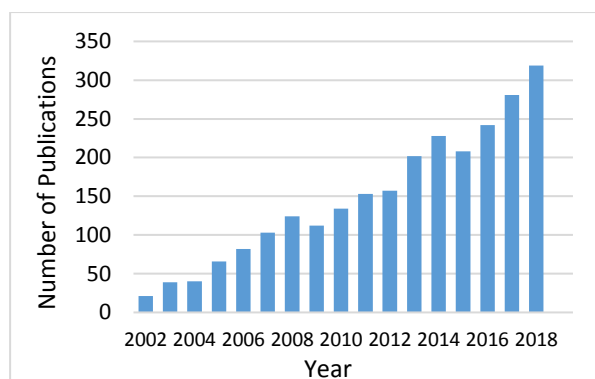
Figure 1. Number of publications on agile software development indexed in Scopus Database.

The second tool we used to analyze literature on agile software development was Social Network Analysis (SNA), which is the process of examining social structures using networks and graph theory. It characterizes network structures in terms of nodes that correspond to some entities, and edges or connections that connect these entities [3]. Applying SNA tools and techniques on the secondary studies, we revealed and analyzed the connections among authors, institutions, and countries of the publications under our consideration.

This paper is organized as follows: In Section 2, we provide some literature review on recent related studies. Section 3 explains the details of our methodology. In Section 4, we present our findings and

discuss about them. Section 5 concludes the study and provides some future directions for further research.

II. RELATED WORK

The term "Agile" is an umbrella concept that includes methods such as Scrum, eXtreme Programming (XP), Crystal, Feature Driven Development (FDD), Dynamic Software Development Method (DSDM), and Adaptive Software Development [4, 5]. Agile Methodologies is a suite of software development methods based on iterative and incremental development. The four main characteristics that are fundamental to all agile methodologies can be summarized as: adaptive planning, iterative and evolutionary improvement, rapid and flexible response to change, and encouraging communication.

Despite agile methods are developed for small and independent projects, they attract large projects and companies due to their benefits. Larger projects are known with a need for more coordination in comparison to small projects. Therefore, in larger projects how to cope with the coordination within teams, departments, users, and even stakeholders occupies the top priority to solve. Regardless of this coordination related challenges, large scale projects are increasingly adopting agile methodologies [6].

We see that the first studies on the agile method focused on issues related to the adoption of these methods [7]. Team dynamics (such as trust, self-organization, and communication) [8], results of test-driven development [9], post-adoption issues [10], and the difficulties of implementing agility in distributed environments [11, 12] are also included in the literature. In addition to primary studies on agile methods, there are also a significant number of secondary studies in the form of literature reviews and mappings published in this area over the past decade [4].

Case studies of current systematic literature reviews on agile software development can be shown as follows. In Jalali and Wohlin's study, it is revealed that the main difficulties for global software development and agile integration between 1999-2009 are differences in time zones, communication, level of trust, knowledge in management, culture and difficulties for personnel [13].

Silva et al.'s study focused on integrating CMMI with agile methods. In this study, they argued that CMMI can be used to reduce the efforts to reach levels 2 and 3, and even showed that there are reports showing that the application of agile methods will achieve level 5 [14].

Methods that provide bibliometric analysis, such as social network analysis, are widely used to provide a comprehensive structure and visualization of a particular literature knowledge structure. In such studies, the data to be analyzed in the social network may also be data groups such as authors, references,

and institutions emerging in the systematic literature studies or the tertiary study in which systematic literature studies are examined [15, 16].

The purpose of any tertiary study is to reveal important researches, subject areas, orientations, knowledge status, important researchers, institutions, and information exchange structure in a particular field [2, 5, 17]. Since the amount of data generated during these studies is enormous, social network analysis offers great opportunities for better understanding, visualization, and in-depth analysis of current relationships [16, 18]. In this current study, subjects, researchers, and reference relationships in the systematic literature studies between 2013-2018 on Agile software development are revealed, summarized and visualized, and the current situation, hidden patterns and future trends in this field are revealed.

III. MATERIAL AND METHOD

3.1. Research Questions
The research questions of our tertiary study were determined by adapting the research questions by Kitchenham et al. [17] to cover the publications between 2013 and 2018. Accordingly, the research questions of our study are as follows:
RQ1: How many SLRs were published related with agile software development between 2013 and 2018?
RQ2: Which subjects of agile software development were focused on in SLR studies published between 2013-2018?
RQ3: Which researchers, institutions, and countries are most active in agile software development SLRs between 2013 and 2018?
RQ4: How is the social network structure of countries and institutions in SLRs between 2013 and 2018?

3.2. Tertiary Study's Search Process
In the first stage, we conducted a literature review for our tertiary study to identify and analyze existing Systematic Literature Reviews (SLR) related to Agile Software Development (ASD). We followed and applied the guidelines proposed by Kitchenham et al. [2] for the tertiary study conducted within the scope of this study. The data of an SLR in this context also served as the basis for social network analysis that we performed subsequently.

In order to select the scientific databases, we considered the databases that contain publications in the fields of Computer Engineering, Software Engineering, and Computer Science. As a result, we selected Scopus, Science Direct, IEEE, and Web of Science databases to search. As stated in the studies of SLRs published in previous years, these databases cover a significant part of the publications on agile software development.

While determining the search texts, we took into consideration the texts used by Kitchenham et al. in [17] and we adapted them to the scope of our study.

We carried out searches on determined texts on title, abstract, and keywords as seen in SLRs published in previous years. Examples of the search strings specified are given in Table 1.

Table 1. Examples of the search strings specified when searching the scientific databases.

Search String
"agile" AND "review of studies"
"agile" AND "structured review"
"agile" AND "systematic review"
"agile" AND "literature review"
"agile" AND "literature analysis"

We determined the search interval as between January 2013 and December 2018, and we researched the publications (research articles and conference papers) published in the last 6 years. We performed our searches on each database specifically on March 16, 2019 for Web of Science, March 18, 2019 for IEEE, and March 30, 2019 for Science Direct and Scopus.

We found 896 publications in total in the searched databases for each search string. We eliminated duplicate publications and the remaining 409 publications were subjected to the inclusion/exclusion criteria at the next stage. We applied the following inclusion/exclusion criteria:

Inclusion:
- Being within the scope of agile software development,
- Including a systematic literature review,
- Availability of full text.

Exclusion:
- Not concentrating on software project management,
- Using agile methodologies as tools, not goals.

After the review for the inclusion/exclusion criteria, 281 publications were eliminated. The remaining 118 publications were examined according to the quality criteria in the next stage. The quality criteria to be applied were chosen as the DARE quality criteria used in tertiary studies [19, 20]. After the evaluation stage made according to the quality criteria, we also eliminated 19 publications with under 2 points. The bibliographic data of five publications that were not indexed in the Scopus database were not available, thus the remaining 94 publications formed the actual dataset of this study for SNA.

3.3. Social Network Analysis
The introduction of the sociogram concept in 1937 by Moreno [21] can be considered as the beginning of SNA (Social Network Analysis) in the social sciences. Sociograms are a representation, where interpersonal relationship patterns can be studied as larger social clusters [3]. The development of social network analysis tools like UCINET, PAJEK, NodeXL, and Gephi with the development of computer technology

has advanced the progress of social network analysis. These social network analysis tools have become attractive for different disciplines such as engineering, primarily sociology, anthropology, economics, and political science.

In this study, we used both VOSviewer [22] and Gephi [23] for SNA and network visualization. VOSviewer is a software tool designed specifically to create and visualize bibliometric maps and to create graphic representations of maps of scientific papers [22]. VOSviewer can create clusters of the nodes in networks. The algorithm used for clustering has several parameters that can be tuned like resolution and minimum cluster size. The resolution parameter determines the level of detail of the clusters and it cannot get a negative value. We used the default value of 1 of the resolution parameter in our research. Gephi is Java based open-source network analysis and visualization software. It can view complex networks in real-time very efficiently [23].

3.4. Data Preprocessing

In order to perform analysis with VOSviewer software, a data file in CSV or RIS format is required. On the English language publications in the Scopus database, with the 2013-2018 year filter, under the title-abstract-keyword tab, the search was carried out with the help of the query *"agile and ("analysis of research" or "literature analysis" or "literature review" or "literature survey" or "meta-analysis" or "past studies" or "review of studies" or "structured review" or "subject matter expert" or "systematic review")"*. As a result of the search, 94 of the 99 publications determined at the beginning were reached and bibliographic data (authors, name of the publication, publication year, source title, number of citations, source type, institutions/countries of the authors, keywords, and resources in the publication) were saved in CSV format for further analysis with social network analysis. The screen of search is shown in Figure 2.

Figure 2. Data search screen in CSV format in Scopus database.

In the data downloaded from the Scopus, the inconsistency problems in the fields such as author name and institution name were resolved at the preprocessing stage. For example, while authors with more than one name can use a single name in some publications, the same person can use more than one name together in a different publication. Additionally, the same conference event can be expressed with different names. Since social network analysis tools cannot detect and automatically resolve this situation, we performed a meticulous pre-processing and deduplication on the data. As a result, the number of publications in our final dataset was reduced from 4,914 to 3,697 after the preprocessing phase.

3.5. Citation Analysis and Bibliographic Coupling

Citation analysis shows the effects of publications, authors, sources, institutions, and countries on the subject through the number of citations [24]. It determines network relationships according to the number of publications the authors cite each other. It not only takes place on the basis of publication, the relations can also be visualized on the basis of the author, the institutions the authors are affiliated with,

the countries, and the source of the publication (scientific journal, book, or conference). Newer publications are disadvantageous in this analysis compared to older ones, since earlier publications will have more citations than newer ones.

Bibliographic coupling was introduced in the literature by Kessler in 1963 [25]. This analysis matches the publications that share common references. It is a useful approach that can be used in all fields of research, as it helps the researcher find related publications in the past.

IV. RESULTS AND DISCUSSION

4.1. Tertiary Study Results

The aim of the first research question investigated within the scope of our study was to determine how many SLRs related to agile software development were published between 2013-2018. A total of 94 publications were examined, 39 (41%) of them were research articles published in scientific refereed journals, and 55 (59%) were papers presented at various scientific conferences and/or congresses. The

distribution chart of these publications according to the years they were published is shown in Figure 3.
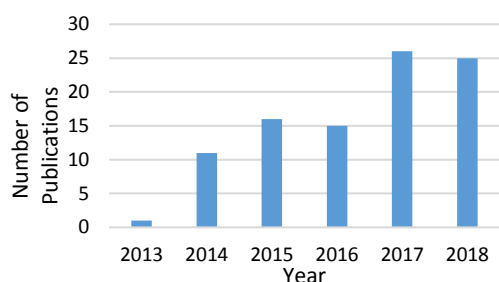

Figure 3. Distribution of publications according to the publication year.

The sources of the 94 publications examined within the scope of the tertiary study can be seen in Table 2. While creating the table, sources with more than one publication were included. Information and Software Technology journal, ACM International Conference Proceeding Series, Lecture Notes in Computer Science are the top three places with the highest number of publications.

The purpose of the second research question investigated within the scope of the study was to determine which subject areas were researched with agile software development between 2013-2018. The main subject areas that 94 publications focused on are shown in Table 3. As seen in Table 3, Agile practices, Agile and management areas and Agile and global or distributed software engineering issues are the most common issues in publications.

One of the aims of the third research question investigated within the scope of the study was to determine which researchers were most active in agile software development SLRs between 2013-2018. There are 3 publications of 6 authors out of 270 authors in 94 publications examined and they are shown in Table 4.

4.2. Citation Analysis

Our dataset consists of publications by 270 researchers and from 133 institutions worldwide. Total number of citations of these publications is presented in [20].

In this study, we present the social network diagram of the institutions and the universities in Figure 4 and 5, respectively. The sizes of the nodes in the figures are proportional to the number of citations. We also present the citation clusters of the institutions and the universities in Table 5 and 6, respectively.

As can be seen in Figure 4 and 5, the selection of a short time interval for analysis negatively affected the number of connections due to the limited time for publications to cite each other. Despite this, although the analysis included a limited number of networks, it is noteworthy that if there were more than one institution belonging to the same country, there would be more than one institution belonging to a country in

the same cluster. This means that researchers are more interested in publications published in their own country, and it is an unexpected result in an era of increasing globalization.

In the citation analysis for instutions, we see that the institutions that were strong according to the connection power had a success above the average compared to the institutions that did not have a relationship in the related figure. Another reason for the low number of links may be the abundance of resources to be used, since agile software development is a very popular topic, because even if the publications do not refer to each other, it was seen that they should not be underestimated in the number of citations.

We also analyzed our data on the basis of the country of the authors, with a total of 33 countries. VOSviewer results of country-level citation analysis can be seen in Figure 6. 7 out of 11 countries that provide a connection between clusters are from Europe, that means the authors refer to each other more in Europe. It can be seen that Brazil does not belong to any cluster and connects the red cluster (Chile, Peru, Spain, UK) to green cluster (Iran, Malaysia, Slovenia) as a kind of bridge.

4.3. Bibliographic Coupling Analysis

Bibliographic coupling is a type of relationship definition used to establish a similarity relationship between publications. It occurs when two studies refer to the same study in their references. The result of bibliometric coupling analysis with VOSviewer for 33 countries created from tertiary study data can be seen in Figure 7, and Table 7 shows bibliographic coupling clusters by country. The size of the nodes in the figure increases in proportion to the number of citations. According to the total connection power calculated as a result of bibliographic coupling analysis, it is seen that the most active five countries are Brazil, Germany, Finland, Malaysia, and Pakistan, respectively. Brazil combines the two largest clusters as the center of this analysis. Five of the six most central countries that follow Brazil are seen to be in Europe.

Figure 8 shows the results of bibliographic coupling analysis for 133 institutions. According to the connection power, the five most active institutions are Pontifical Catholic U. of Rio Grande do Sul (Brazil), Blekinge Institute of Technology from Sweden, University of Sao Paulo (Brazil), Pontifical Catholic University of Peru (Peru), and COMSATS Institute of Inf. Technology (Pakistan).

Institutions in bibliographic coupling include institutions from Brazil, Sweden, Peru, Finland and the Netherlands. On the other hand, Brazil, Germany, Finland, Malaysia, and Pakistan are in the first places in the country-based analysis.

Table 2. Distribution of publications according to sources.

Source of Publication	Type	Total Publications
Information and Software Technology	Journal	12
ACM International Conference Proceeding Series	Conference	9
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	Conference	7
Journal of Software: Evolution and Process	Journal	5
Communications in Computer and Information Science	Journal	4
CEUR Workshop Proceedings	Conference	3
Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE	Conference	3
2016 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 – Proceedings	Conference	2
Advances in Intelligent Systems and Computing	Journal	2
Journal of Systems and Software	Journal	2

Table 3. Distribution of publications according to the subject areas covered.

Subject Area	Total Publications
Agile practices	25
Agile and management areas	13
Agile and global or distributed software engineering	9
Agile and usability	7
Agile transformation	6
Agile methodologies	6
Agile and organization	6
Advantages and disadvantages of the agile method, challenges and critical areas	6
Agile cost / effort calculation	6
Agile and CMMI	5
Agile human and social reviews	3
Agile and sustainability	3
Agile product line engineering	2
Agile and embedded systems	1

Table 4. The most active researchers.

Researcher	Number of Researcher Citations	Number of Publications	Affiliations
Rafael Prikladnicki	20	3	School of Technolog , Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Mirko Perkusich	9	3	Department of Computing and Systems, Federal University of Campina Grande , Brazil
Siffat Ullah Khan	3	3	Department of Computer Science & IT, University of Malakand, Pakistan
Nasir Rashid	3	3	Department of Computer Science and Information Technology, University of Malakand, Lower Dir, Pakistan
Edna Dias Canedo	1	3	Computer Science Department, University of Brasília -, Brazil;
Ruyther Parente da Costa	1	3	Computer Science Department, University of Brasília -, Brazil;

Figure 4. Social network visualization of institutions based on the citation analysis.

Figure 5. Social network visualization of universities based on the citation analysis.

Table 5. Citation clusters by institutions.

Cluster 1	Cluster 2	Cluster 3
Departamento De Ingeniería Informática Y Ciencias De La Computación, **Universidad De Atacama**, Copiapó, Chile	**Agencia Andaluza De Instituciones Culturales**, Junta De Andalucía, Spain	Department Of Computer Science, Chalous Branch, **Islamic Azad University** (Iau), Chalous, Mazandaran, 46615-397, Iran
Department Of Computer Science, **University Of York, York**, United Kingdom	Escuela De Posgrado, **Pontificia Universidad Católica Del Perú**, Lima, Peru	Faculty Of Electrical Engineering And Computer Science, **University Of Maribor**, Smetanova 17, Maribor, 2000, Slovenia
Escuela Politécnica Superior, Universidad **Autónoma De Madrid**, Madrid, Spain	**Kentico**, Brno, Czech Republic	**Msg Life Odateam D.O.O.**, Titova 8, Maribor, 2000, Slovenia
Lappeenranta University Of Technology (Lut), P.O Box 20, Lappeenranta, Fı-53851, Finland	**Masaryk University**, Brno, Czech Republic	**Pontifícia Universidade Católica Do Rio Grande Do Sul**, Porto-Alegrers, Brazil
Universidad Politécnica De Madrid, Campus De Montegancedo, Boadilla Del Monte, 28660, Spain	**Pontificia Universidad Católica Del Perú**, Lima, Peru	**University Of Malaya**, Malaysia
University Of Applied Sciences Emden/Leer, Emden, Germany	**University Of Sao Paulo**, Sao Paulo, Brazil	**University Of Twente**, Enschede, Netherlands
Instituto Tecnológico De Conkal, Km 16 Oldroadmotul Yucatán, Mexico	**University Of Seville**, Spain	
Cluster 4	**Cluster 5**	
Department Of Computer Science, **Aalborg University**, Denmark	**University Of Oulu**, Finland	
Department Of Management, **Aarhus University**, Denmark	**University Of St.Gallen**, Switzerland	
Department Of Information And Computer Science, **Utrecht University**, Utrecht, Netherlands		

Table 6. Citation clusters by universities.

| Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
University of Atacama	Agencia Andaluza de Instituciones Culturales	Islamic Azad University	Aalborg University	University of Oulu
University of York	Pontifical Catholic University of Peru	University of Maribor	Aarhus university	University of St.Gallen
Autonomous University of Madrid	Kentico Software	MSG Life Odateam D.O.O.	Utrecht University	
Lappeenranta University of Technology	Masaryk University	Pontifical Catholic University of Rio Grande do Sul		
Technical University of Madrid	Pontifical Catholic University of Peru	University of Malaya		
University of Applied Sciences Emden/Leer	University of Sao Paulo	University of Twente		
Conkal Institute of Technology	University of Seville			
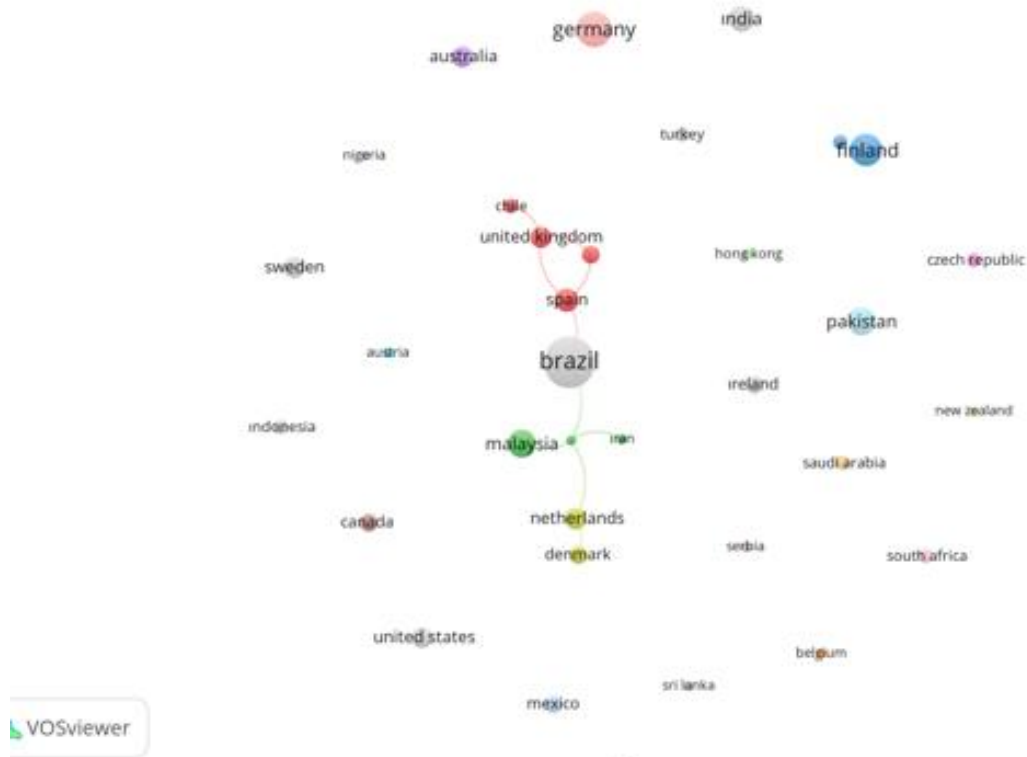
Figure 6. Citation analysis of countries.
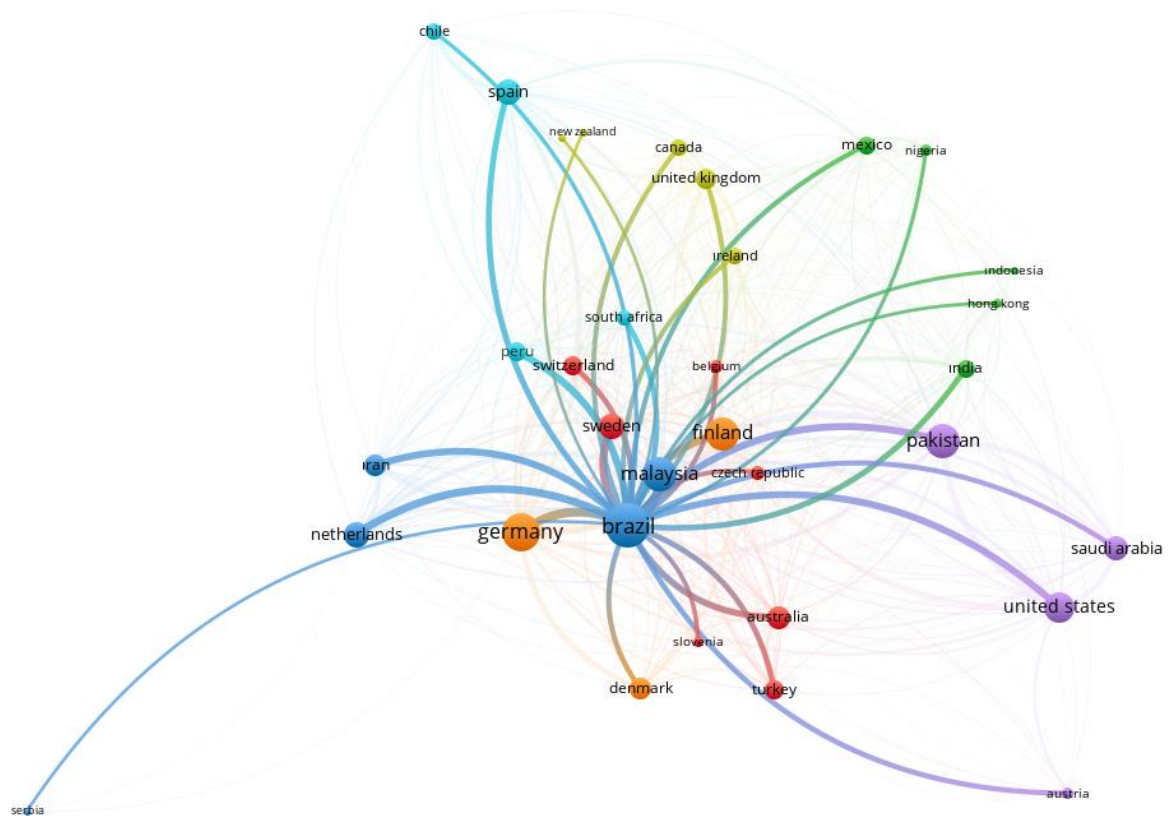
Figure 7. The most active countries of bibliographic coupling analysis.

Table 7. Bibliographic coupling clusters by country.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Belgium	Australia	Chile	Austria	Denmark	Iranian
Canada	Brazil	Peru	Pakistan	Finland	Malaysia
Hong Kong	Czechia	South Africa	Saudi Arabia	Germany	Netherlands
India	Slovenia	Spain	USA		
Indonesia	Sweden				
Ireland	Switzerland				
Mexican	Turkey				
New Zealand					
Nigeria					
Serbia					
Sri lanka					
United Kingdom					

Figure 8. The most active institutions of bibliographic coupling analysis.

V. CONCLUSION

Agile software development, which is increasingly used worldwide, is growing steadily, increasing its reputation and active role in the industry. In this study, a tertiary study from literature reviews on agile software development was conducted and social network analysis was performed with the dataset created by the researchers from the tertiary study.

According to the results of our study, the most active sources according to the number of publications were determined as Information and Software Technology and ACM International Proceeding Series. Information and Software Technology is a journal that deals with systematic literature reviews, and we were inspired by Kitchenham et al.'s 2010 study on it [17]. ACM, on the other hand, is a reputable database that is used by most of the reviewed literature reviews. The scope of the studies reviewed was mostly on agile practices and management areas. Embedded systems, product line engineering, and human and social studies were the least explored topics.

Our study has some limitations that may affect its usefulness. First, our study covered the SLRs on agile software development published between 2013 and 2018. It can further be extended to include the SLRs published since 2018 and can present a more comprehensive view of the field. Second, the term "agile" is an umbrella term that can be too broad to cover the practices applied in software engineering. Rather than using this term as part of the search keys in SLR research, more focused literature reviews can be conducted using related but more specific search terms like "extreme programming", "scrum", "lean", and so on.

The combination of tertiary study and social network increased the power of the study and presented a different perspective for the subject. The findings obtained within the scope of the study can be used for various purposes. For example, they can be used to determine an invited speaker related to the subject at a conference. In addition, they can be used to determine which universities are prominent in a special research field. Moreover, they can be helpful for finding institutions that will partner to work in international university-based studies, or for forming a team on a research subject.

REFERENCES

[1] Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*, Report Department of Computer Science, Keele University, UK.

[2] Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Report Keele University and Durham University.

[3] Tunalı, V. (2016). *Sosyal Ağ Analizine Giriş*. Ankara, Türkiye: Nobel Akademik Yayıncılık.

[4] Hoda, R., Kruchten, P., Noble, J., & Marshall, S. (2010). Agility in context. *Proceedings of the ACM international conference on Object oriented programming systems languages and applications*, Reno, NV, October 17.

[5] Hoda, R., Salleh, N., Grundy, J., & Tee, H. M. (2017). Systematic literature reviews in agile software development: A tertiary study. *Information and Software Technology,* 85, 60-70.

[6] Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and success factors for large-scale agile transformations: A systematic literature review. *Journal of Systems and Software,* 119, 87-108.

[7] Boehm, B. (2002). Get ready for agile methods, with care. *Computer,* 35(1), 64-69.

[8] Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Scrum project. *Information and Software Technology,* 52(5), 480-491.

[9] Janzen, D., & Saiedian, H. (2005). Test-driven development concepts, taxonomy, and future direction. *Computer,* 38(9), 43-50.

[10] Aggarwal, V., & Singhal, A. (2019). Empirical study of test driven development with scrum. *3rd International Conference on Advances in Computing and Data Sciences*, Ghaziabad, India, April 12-13.

[11] Alam, S., Nazir, S., Asim, S., & Amr, D. (2017). Impact and challenges of requirement engineering in agile methodologies: A systematic review. *Int. J. Adv. Comput. Sci. Appl,* 8(4), 411-420.

[12] Schön, E.-M., Winter, D., Escalona, M. J., & Thomaschewski, J. (2017). Key challenges in agile requirements engineering. *18th International Conference on Agile Software Development*, Cologne, Germany, May 22-26.

[13] Jalali, S., & Wohlin, C. (2012). Global software engineering and agile practices: a systematic review. *Journal of software: Evolution and Process,* 24(6), 643-659.

[14] Silva, F. S., Soares, F. S. F., Peres, A. L., de Azevedo, I. M., Vasconcelos, A. P. L., Kamei, F. K., & de Lemos Meira, S. R. (2015). Using CMMI together with agile software development: A systematic review. *Information and Software Technology,* 58, 20-43.

[15] Xu, X., Chen, X., Jia, F., Brown, S., Gong, Y., & Xu, Y. (2018). Supply chain finance: A systematic literature review and bibliometric analysis. *International Journal of Production Economics,* 204, 160-173.

[16] Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2020). A systematic literature review with bibliometric

analysis of big data analytics adoption from period 2014 to 2018. *Journal of Enterprise Information Management*, 34(1), 101-139.

[17] Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering – A tertiary study. *Information and Software Technology,* 52(8), 792-805.

[18] Bartolacci, F., Caputo, A., & Soverchia, M. (2020). Sustainability and financial performance of small and medium sized enterprises: A bibliometric and systematic literature review. *Business Strategy and the Environment,* 29(3), 1297-1309.

[19] Database of Abstracts of Reviews of Effects (DARE). (2008). In: *Encyclopedia of Public Health,* W. Kirch (Ed.). Dordrecht: Springer Netherlands.

[20] Bayram, E., Doğan, B., & Tunalı, V. (2020). Bibliometric Analysis of the Tertiary Study on Agile Software Development using Social Network Analysis. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, October 15-17.

[21] Chinowsky, P., & Taylor, J. E. (2012). Networks in engineering: an emerging approach to project organization studies. *Engineering Project Organization Journal,* 2(1-2), 15-26.

[22] Eck, N. J. v., & Waltman, L. (2014). Visualizing Bibliometric Networks. In: *Measuring Scholarly Impact,* Y. Ding, R. Rousseau & D. Wolfram (Eds.). Switzerland: Springer, Cham, 285-320.

[23] Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*, San Jose, California, USA, May 17-20.

[24] Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. *Journal of Operations Management,* 27(3), 185-202.

[25] Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation,* 14(1), 10-25.

Supervised Learning-Aided Control of a DC-DC Power Converter in Wind Energy Conversion Systems

Rüzgâr Enerjisi Dönüşüm Sistemlerinde bir DC-DC Güç Dönüştürücüsünün Denetimli Öğrenme Destekli Kontrolü

Alper Nabi AKPOLAT[1] ⓘ **, Erkan DURSUN[1]** ⓘ **, Ahmet Emin KUZUCUOĞLU[1]** ⓘ

[1]Marmara University, Department of Electrical-Electronics Engineering, 34854, Istanbul, Turkey

Abstract

Over the last decades, to adopt high penetration of renewable energy sources (RESs) in electrical energy systems, distributed energy resources (DERs) have become prominent. Due to easy attainability status of small wind turbines (WTs), wind energy conversion systems (WECSs) are feasible applications for small customers, especially in windy areas. The next decade is likely to witness a considerable rise in DERs. In this context, WECSs are preferred broadly, thus harvesting wind energy into electrical energy effectively is a substantial issue. WTs can be got involved in the grid-connected or autonomous mode with a variety of topologies. In this paper, we examine to control of DC-DC boost converter of a WECS with the help of artificial intelligence (AI)-aided PI controller based on supervised learning method. Regarding the proposed method, artificial neural networks (ANNs) as a subset of AI are utilized. To test and ensure the applicability of the proposed control method, a small WECS with a permanent magnet synchronous generator (PMSG) connected a DC bus was implemented in MATLAB/Simulink environment. The proposed ANN scheme has reached a high accuracy rate with an overall mean squared error (MSE) equal to 7.4e-08. The results present that dynamic response and less complexity with a high accuracy rate have been obtained under study. The main target of this study is to reduce the number of sensors in the control layer. Thus, a cost-effective and more reliable structure is obtained with fewer sensor requirements.

Keywords: Wind Energy Conversion Systems, Artificial Intelligence, Artificial Neural Networks, DC-DC Power Converters

Öz

Son on yıllarda, elektrik enerjisi sistemlerinde yenilenebilir enerji kaynaklarının yüksek oranda nüfuzunu yaygınlaştırmak için dağıtılmış enerji kaynakları öne çıkmıştır. Küçük rüzgâr türbinlerinin kolay erişilebilirlik durumu nedeniyle, rüzgâr enerjisi dönüşüm sistemleri özellikle rüzgârlı alanlarda küçük müşteriler için elverişli uygulamalardır. Önümüzdeki on yıl muhtemelen dağıtık enerji kaynakları önemli bir artışa tanık olacaktır. Bu bağlamda, rüzgâr enerji dönüşüm sistemleri yaygın olarak tercih edilmektedir, bu nedenle rüzgâr enerjisinin elektrik enerjisine etkin bir şekilde dönüştürülmesi önemli bir konudur. Rüzgâr türbinleri çeşitli topolojilerle şebekeye bağlı veya otonom modda dâhil edilebilirler. Bu makalede, denetimli öğrenme yöntemine dayalı yapay zekâ destekli PI denetleyicisi yardımıyla bir rüzgâr enerji dönüşüm sistemindeki yükselten DC-DC güç dönüştürücüsünün kontrolünü incelemekteyiz. Önerilen yöntemle ilgili olarak, yapay zekânın bir alt kümesi olarak yapay sinir ağları kullanılmaktadır. Önerilen kontrol yönteminin uygulanabilirliğini test etmek ve doğrulamak için, MATLAB/Simulink ortamında bir DC baraya sabit mıknatıslı senkron generatör ile küçük bir rüzgar enerji dönüşüm sistem uygulanmıştır. Önerilen YSA şeması, 7.4e-08'e eşit toplam ortalama karesel hata (MSE) ile yüksek bir doğruluk oranına ulaşmıştır. Sonuçlar, çalışma kapsamında dinamik yanıtın ve daha az karmaşıklığın yüksek doğruluk ile elde edildiğini göstermektedir. Bu çalışmanın ana hedefi, kontrol katmanındaki sensör sayısını azaltmaktır. Böylece daha az sensör gereksinimi ile uygun maliyetli ve daha güvenilir bir yapı elde edilmektedir.

Anahtar Kelimeler: Rüzgâr Enerjisi Dönüşüm Sistemleri, Yapay Zekâ, Yapay Sinir Ağları, DC-DC Güç Dönüştürücüleri

I. INTRODUCTION

1.1. Motivation and Background

The availability of energy is not the only concern for today's power systems. Another critical factor is its impact on consuming resource deficiency. Energy demand figures are expected to increase further due to the growing population, modernization, and globalization. Human beings have produced energy in conventional ways and still resume for years. However, it is inevitable that energy production is gradually abandoned the conventional ways due to both the depletion of resources and environmental concerns. This mandatory transition leads humanity to energy production with renewable energy sources (RESs). For RES-based applications, power electronics have grown into an accepted choice to supply sustainable power, since their complexity and performance have been boosted. Furthermore, several barriers that affect the efficiency of power electronics exist. Accomplishing these barriers demands to handle complicated optimization difficulties, which are challenging to alleviate through conventional methods [1].

One of the most preferred RES-based systems, i.e., distributed energy resources (DERs) is that wind energy conversion systems (WECSs) are operated with the help of power electronic converters according to system topology. The WECSs are largely popular structures, which are preferred to feed electrical appliances in any circumstance [2]. Thereby, a safe and reliable power supply is directly related to the robust control and management of power electronics. In other words, power electronics can be considered as an easy linker of DER applications. However, the coordination of these power electronics interfaces causes many problems in the control layer due to the intermittent nature of RESs and their uncertainty.

The desired characteristics of the control system should comprise tracking the output reference values (e.g., current and voltage), ensuring the presence and reliability of the power balance. Unlike classical control methods, this paper proposes the exploitation of a supervised learning-based multi-layer feedforward artificial neural network (ANN) to accomplish efficiency with high accuracy and to reduce the number of sensors in the control layer. We can implement the model of machine learning in various ways, ANNs are one of them. Thereby, the ANNs are a subset of machine learning and artificial intelligence (AI) targets to facilitate systems with intelligence that is susceptible to human-like learning and reasoning.

Nowadays, AI has palpable advantages by having been successfully applied in innumerable areas such as recognition, classification, computer vision, energy, and vehicle technology, etc [3]. To this end, the proper control of power electronics has crucial for dealing with control difficulties, which should be taken into consideration.

1.2. Related Literature

Upon relevant literature, it is obvious that many researchers have addressed the application of AI methods such as ANN, fuzzy, deep learning, or hybrid methods etc. in power electronics [4]-[6]. In particular, similar topology and controller have been studied as being an experimental setup with a wind turbine (WT) emulator recently [7]. Few researchers have mentioned estimating accurate wind speed to harvest maximum power with sensorless control using AI algorithms in [8]. Similarly, maximizing effective power is discussed profitably with the aid of ANN-based reinforcement learning for a WECS via permanent magnet synchronous generator (PMSG) [9]. The authors in [10], suggest that sliding mode observer outperforms in tracking error than ANN-based observer as a sensorless control of PMSG in WECS; however, sliding mode observer deteriorates with chattering issue. To control the back-to-back power converter of a WECS, space vector-based pulse width modulation (SVPWM) has been developed through a feedforward ANN by obtaining lower total harmonic distortion

(THD) [11]. On the other hand, an ANN-based voltage estimation for the calculation of THD is presented for multi-bus islanded AC microgrids [12]. Especially, a seamless estimated current data is implemented to support the control operation of a small WECS with the help of supervised learning-based ANN [13]. In [14], another RES-based system is that a PV panel is monitored for perceiving in case of the PV panel encounter degradation due to any faults. In other respect, to facilitate the operation of the energy management as reliable an ANN-based sensorless control [15] is proposed to eliminate the implemented sensors.

Various approaches have been put forward to come around the barriers against efficiency and performance. Except for estimating current data, in [16] it is seen that wind speed and torque data are observed properly with sensorless control for small WT clusters throughout a direct torque control algorithm. Much work on the potential of the sensorless control has been carried out [17], yet there are still some critical issues about maximum power point tracking (MPPT) algorithm is applied to harvest maximum power from the wind [18]. As can be clearly seen in Table 1, the most recent studies with similar applications are given based on the related literature.

Table 1. Comparison between related studies.

SCHEME REF.	APPLICATION	SCHEME REF.	APPLICATION
ANN-Based MPC [19]	Control of power converters	Artificial neural network [23]	Intelligent long-term performance analysis
FCS-MPC-based NN Classifier [20]	Voltage sag classification in PV system	Artificial Neural Networks [24]	Cyberattack detection & mitigation for microgrids
Long short-term memory NN (LSTM) [21]	Power prediction for virtual plants	MPC Using ANN [25]	Control of DC-DC Converters
Proposed ANN-Based PI [13]	Sensorless control of DC-DC converter	Batch normalization neural network (BN-NN) [26]	Circuit parameter design of the converter
Multilayer Perception ANN [22]	Intelligent energy management	Lagrange programming neural network (LPNN) [27]	Obtaining optimal scheduling for microgrid

1.3. Contributions and Organization

Estimating current data enables power estimation as well. We do not only assert controlling a power converter of the WECS via ANN but also provide more reliable operation with less measured data in the control layer. This paper proposes to control a DC-DC boost converter with the aid of ANNs. The strong

aspects of this paper can be divided into three matters: 1) the applied ANN possesses high accuracy to estimate the desired output, which is easy to apply, thanks to the proper input selection, 2) after an offline (training) phase of the application, the results validate that the online (exploitation) phase is conducted effectively, 3) this proposed AI-aided control via ANN increase the reliability of the system by reducing the used sensor in the control layer, as the measured sensor may corrupt, 4) the proposed method decreases the cost of ownership of sensors in the control layer and lastly 5) the estimated sensor's data communication challenge is removed.

The reminder of this paper is organized as follows. Section II explains about the WECS under study. Section III elucidates system control strategy for online and offline operation. Section IV evaluates the simulation results of the paper. Finally, conclusion remarks are stated in Section V.

II. WIND ENERGY CONVERSION SYSTEM UNDER STUDY

The mentioned WECS model can be split up into four parts regarding WT, PMSG, three-phase full-bridge diode rectifier, and lastly DC-DC boost converter.

Variable wind speed profile is used for power extraction from the wind. The WT has a vertical axis turbine of three blades with fixed pitch angle (β) is equal to 0°. The WT has coupled with a three phase PMSG to generate power and then dispatch that power to the DC-DC converter with a full bridge diode rectifier. The synchronous generators are broadly preferred for variable wind speed in WT applications due to their low rotation synchronous speeds which track grid frequency. Therefore, it is worth using PMSG in these applications where the wind speed has too much fluctuation [28].

The configuration of the WECS can be illustrated in Figure 1. The maximum value of phase voltage is V_m, average output of the rectifier voltage V_o is V_{DC} and it is expressed as

$$V_{DC} = \frac{1}{T} \int_0^T V_0(t)\,dt. \tag{1}$$

Using equation (3), the average voltage of the output V_{DC-out} can be found as

$$V_{DC-out} = \frac{6}{2\pi} \int_{\pi/3}^{2\pi/3} \sqrt{3} V_m \sin\theta\,d\theta, \tag{2}$$

or

$$V_{DC-out} = V_m \frac{3\sqrt{3}}{\pi} = 1.654 V_m. \tag{3}$$
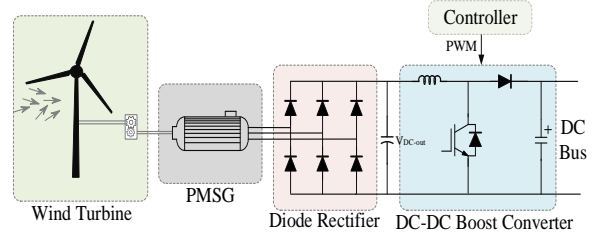
Figure 1.   Configuration of WECS under study.

The WT power characteristics versus turbine speed can be expressed in Figure 2. The turbine output power can be seen that it coincides 320 W for 0.8 per unit of nominal mechanical power at base wind speed, so the nominal output power corresponds to 400 W. The output power and mechanical torque of the WECS can be aligned by following equations. The output power of the turbine is defined as

$$P_m = c_p(\lambda, \beta) \frac{\rho A}{2} v_{wind}^3, \tag{4}$$

where, P_m is Mechanical output power of the turbine (W), c_p is the performance coefficient of the turbine, ρ is air density (kg/m^3), A is the turbine swept area (m^2), V_{wind} the wind speed (m/s), λ is tip speed ratio of the rotor blade tip speed to wind speed, and lastly β the blade pitch angle (degree). An equation is used to model $c_p(\lambda,\beta)$ that based on the modeling turbine characteristics [29] as follows:

$$c_p(\lambda, \beta) = c_1(c_2/\lambda_i - c_3\beta - c_4)e^{-c_5/\lambda_i} + c_6\lambda, \tag{5}$$

with

$$\frac{1}{\lambda} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1}, \tag{6}$$

where, coefficients are located as follows: c_1=0.5176, c_2=116, c_3=0.4, c_4=5, c_5=1, and c_6=0.0068. The turbine characteristics can vary for different values of the pitch angle β.

The model in d-q frame of the PMSG is based on (7). Besides, the electromagnetic torque (T_{em}) of the generator can be expressed as equation (11) unless L_q is not equal to L_d,

$$\begin{bmatrix} V_d \\ V_q \end{bmatrix} = \begin{bmatrix} R_c & -\omega L_c \\ \omega L_c & R_c \end{bmatrix} * \begin{bmatrix} i_d \\ i_q \end{bmatrix} + L_c \frac{d}{dt}\begin{bmatrix} i_d \\ i_q \end{bmatrix} + \begin{bmatrix} e_d \\ e_q \end{bmatrix}, \tag{7}$$

$$T_{em} = P\varphi i_q, \tag{8}$$

where, V_d and V_q are direct and quadratic stator voltages, i_d and i_q are direct and quadratic currents, e_d and e_q are direct and quadratic magnetic motive forces, respectively. R_c is resistance of each stator

phase, φ is the permanent magnetic flux, $\square$ is the rotor electrical angular speed and L_c is inductance of each stator phase. As to the model of PMSG in (7), taking into consideration the Clarke transformation, the PMSG model in α-β frame can be written by

$$\begin{bmatrix} \dot{i}_\alpha \\ \dot{i}_\beta \end{bmatrix} = \begin{bmatrix} \dfrac{-R_s}{L_s} & 0 \\ 0 & \dfrac{-R_s}{L_s} \end{bmatrix} * \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} + \begin{bmatrix} \dfrac{1}{L_s} & 0 \\ 0 & \dfrac{1}{L_s} \end{bmatrix} \begin{bmatrix} V_\alpha - e_\alpha \\ V_\beta - e_\beta \end{bmatrix}, \tag{9}$$

$$e_\alpha = -\omega\varphi\sin\theta, \tag{10}$$

$$e_\beta = \omega\varphi\cos\theta, \tag{11}$$

where, V_α and V_β are stator voltages vectors, i_d and i_q are stator current vectors, e_α and e_β are direct and quadratic magnetic electromotive forces, respectively. R_c is resistance of each stator phase, and L_c is inductance of each stator phase.
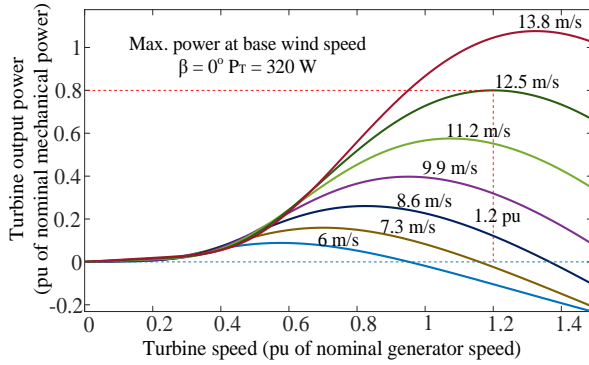
Figure 2. Turbine power characteristic versus wind speed.

III. SYSTEM CONTROL STRATEGY

3.1. Conventional Control
To control the power electronics converters which do not only use traditional multiple feedback loops but also pulse width modulation (PWM) is required. The controller is the key component in achieving a well-controlled and high-performance system. The DC-DC boost unidirectional converter is triggered by the duty cycle (D) of the power semiconductor MOSFET thanks to the PWM signal. Equation (12) describes the output and input voltages relationship as

$$V_{DC\_Bus} = \frac{1}{1-D} \times V_{DC-out}. \tag{12}$$

A block diagram of the configuration and proportional integral (PI)-cascaded conventional control structure of the WECS is seen in Figure 3.
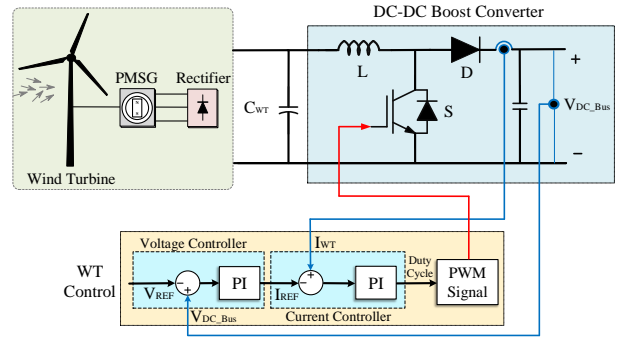
Figure 3. Configuration and control structure of WECS by PI-cascaded current and voltage controllers.

Before proceeding, it is important to mention about obtaining the necessary transfer function for any controller after simplifying and linearizing the modeled system. Where $R_{WT}=V_{WT}/I_{WT}$ is output resistance of the WECS, the transfer function between the converter input voltage (V_{WT}) and the duty cycle (D) can be expressed as follows with the help of the small-signal model as follows:

$$\frac{V_{WT}(s)}{D(s)} = \frac{\dfrac{-V_{DC\_Bus}}{LC_{WT}}}{s^2 + \dfrac{s}{R_{WT}C_{WT}} + \dfrac{1}{LC_{WT}}}. \tag{13}$$

3.2. ANN-Based Control
Essentially, AI tries to mimic the mindset in the human brain. The brains contain a lot of neurons, and these neurons are associated with biological neural networks that facilitate the behavior process. An ANN aims to mimic a sub-simulation of the biological network using electronic circuits. AI databases are applied in many fields such as software, economy, industry, and engineering [12]. Neural, statistical, and evolutionary learning are among the artificial intelligence-based multiple learning theories that ANNs are the most basic technique of neural learning [30]. The ANN was claimed by McCulloch and Pitts based on the mathematical model of a primitive cell of the human brain [31]. Roughly, a neuron is triggered by the weights of inputs that exceed the threshold limit, causing the response of activated output functions, and thus produce an appropriate output. Prediction is a sort of filtering, where past values of one or more-time series are utilized to estimate future values. Dynamic ANNs contain tapped delay lines are ready for nonlinear prediction. They are also feasible for system identification, impending failure detection, dynamic modeling of a physical model. AI can be used for simulation, analysis, visualization, and control of a variety of systems.

Another point worth mentioning is that elements of feedforward neural networks are separated layers. The signs from the input layer to the output layer are transmitted by a one-way connection. While being a

link from one layer to the next, no link exists in the same layer. In feedforward networks, the outputs of the cells in the layers are the input of the next layer. If the layer by layer network is examined, the input layer transmits the information from the external environment to the cells in the intermediate (hidden) layer without making any changes. The output of the network is calculated by processing the input data within the hidden layers and the output layer.

A multi-layer applicator consists of one or more hidden layers and one output layer. All cells in one layer are in contact with all cells in the next layer. Feedforward is derived from the forward flow of information on behalf of networks. There is no information processing in the input layer. The number of cells in the input layer depends on the number of entries in the applied problem. Similarly, the hidden layer and the number of cells in this layer differ with the problem structure but is found by trial-and-error. The number of cells in the output layer is related to the number of outputs in the applied problem.

Feedforward networks can generally respond to problems such as classification, generalization, and recognition by applying the Delta learning rule. For the input presented to the network, the output of the network is compared with the actual result. The difference resulting from this comparison reveals the error value. The purpose of calculating backpropagation is to produce a good output by reducing the error. The error will be distributed to the weight ratings of the network every iteration.

Since nonlinear problems cannot be learned with single-layer perception, and most of the problems encountered in daily life are not linear, we have used this multi-layer feedforward neural network with Levenberg-Marquardt backpropagation in this study. This backpropagation algorithm approaches second-order training speed regardless of computing the Hessian matrix likewise the quasi-Newton methods. When the performance function has the form of a sum of squares, the Hessian matrix and the gradient can be expressed as

$$H = J^T J, \tag{14}$$

with

$$g = J^T e, \tag{15}$$

where, J is the Jacobian matrix that comprises of first derivatives of the network errors regarding the weights and biases, and e is a vector of network errors. The Jacobian matrix can be extracted with a standard backpropagation technique which is less complicated than computing the Hessian matrix. The Levenberg-

Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update as

$$X_{k+1} = X_k - [J^T J + \mu I]^{-1} J^T e, \tag{16}$$

where, the scalar μ is zero, that is precisely Newton's method, using the approximate Hessian matrix. If μ is large, this inclines gradient descent with a small step size. Newton's method is quicker and more accurate near an error minimum, so the aim is to shift toward Newton's method as rapidly as possible.

An auxiliary network training function named "Trainlm" is generally the fastest backpropagation algorithm in the Simulink toolbox, which revises weight and bias states according to Levenberg-Marquardt optimization and also is recommended as a supervised algorithm, even although it occupies more memory but less time than others. By the way, a supervised learning allows one to acquire data and then to produce a new output from previous experience i.e., correct output. It provides to improve performance criteria using this experience. Also, supervised machine learning helps us solving a variety of computational problems in the life [30].

The training process is stopped once the generalization finishes improving, as same as being in the MSE of validation process. Mentioning the operation, an ANN estimates data series of $Y(t)$, while obtaining past values up to delay (d) pieces of $X(t)$ series as

$$Y(t) = f\left(X(t-1), ..., X(t-d)\right). \tag{17}$$

Additionally, supposing that the forms of error data series as $e_t = \{e\}_t$; where, $t = 1, ..., n$, the neural network can be designed under four sections such as proper input selection, defining the paradigms, estimation, and lastly implementation. As stated previously, the backpropagation is a way for training the weights in a multilayer neural network [13, 32]. The generic structure of the backpropagation neural network can be expressed as follows:

$$b_h = \{b_{h,j} | 1 \le j \le m\}, \tag{18}$$

$$v_{t,j} = \left(\sum_{i=1}^{n} (w_{t,ij} X_{t,i}) + b_{z,j}\right); \quad \begin{cases} i = 1,2,...,n \\ j = 1,2,...,m \end{cases}, \tag{19}$$

$$Q_{t,j} = f_{hidden}(v_{t,j}), \tag{20}$$

If an ANN consists of n inputs, one hidden layer with m neurons, and one output, so the output signal can be calculated as

$$Y_t = f_{output}\left(\sum_{i=1}^{n} w_{t,j} Q_{t,j} + b_k\right), \tag{21}$$

where, $X_{t,j}$ is input, $v_{t,i}$, $Q_{t,i}$, and $b_{h,j}$ are input, output of hidden layer (i.e, ith node of hidden layer), and the bias factor for the ith neurons of the hidden layer respectively, b_y is the bias factor of the neuron in the output layer, $w_{t,ij}$ and $w_{t,i}$ are connection weights, and Y_t is output. Then the ANN generates $(m+1)^{th}$ error points, thereby mean absolute percentage error rate, which is the average squared difference between outputs and targets, can be computed as

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X(t)-Y(t)}{X(t)}\right|.100. \qquad (22)$$

The structure of implemented neural network possesses one hidden layer, and its sigmoid activation function is given by

$$Q_{t,j} = \frac{1}{1+e^{-v_{t,j}}}, \qquad (23)$$

also can be defined as

$$Q_{t,j} = \begin{cases} 0 & v_{t,j} < 0, \\ 1 & v_{t,j} \ge 0. \end{cases} \qquad (24)$$

The schematic diagram of elucidated and studied of multi-layer feedforward neural network with Levenberg-Marquardt backpropagation can be illustrated in Figure 4. Neural Net Time Series application exists in Machine Learning Toolbox of MATLAB/Simulink was used in order to implement the mentioned ANN. Implemented ANN has 3 inputs, 1 output, one hidden layer with 10 hidden neurons, and 2 delays. The node size of the hidden layer is selected based on satisfactory results. If it is selected less than average, the network cannot be trained well, otherwise, selecting more nodes causes more unnecessary complexity. On the other hand, the output depends on the historical value of the inputs, though the number of delays may result in a better performance when the system is more dynamic; however, a great deal of delays makes the training process slow. With respect to stages in the implementation phase, the data was shared to having divided into training, validation, and testing part was selected 70%, 15%, and 15%, respectively. In the training part, the network is adjusted according to the error, then is met to adapt network generalization until generalization stops training with the help of using Levenberg-Marquardt. Due to selection of simulation sample time (T_s) as 5e-6 secs and simulation period (ΔT) as 4 secs, the number of elements can be calculated regarding $(1/T_s)*(\Delta T)$. In the testing, both during and after training the network is tested for an independent measure of network performance. Besides, we embedded the input as an 800001x3 matrix, representing dynamic data 800001-time steps of 3 elements and targeted 800001x1 matrix,

representing dynamic output data 800001-time steps of 1 element [13].
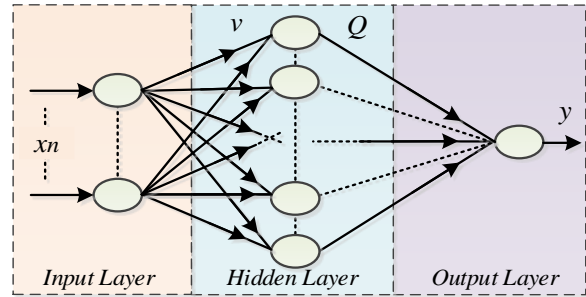
Figure 4. Schematic structure of ANN to estimate output [12].

In this paper, we have implemented a nonlinear input-output neural network which is simple and easy to apply. It is worth mentioning that applying ANN-aided control to the WECS can be split up into two main phases regarding the offline and online part. The offline mode is met to gather the required dataset for training ANN that will be an estimator as can be depicted in Figure 5. After that phase, ANN is well-trained and prepared to estimate the desired output for the online mode. That approach alleges preparing a well-trained and tuned network via a large amount of input data (related output) in the training phase.

With the help of depth of dataset, the ANN targets to predict required inputs of the converter controller without using sensor measurements to eliminate the sensor data of the mentioned input. In the online phase, the well-tuned ANN carries out to estimate the output whose sensor needs to be eliminated in the control layer. In this context, Figure 6 depicts a general picture of the control structure that is aided by ANN.
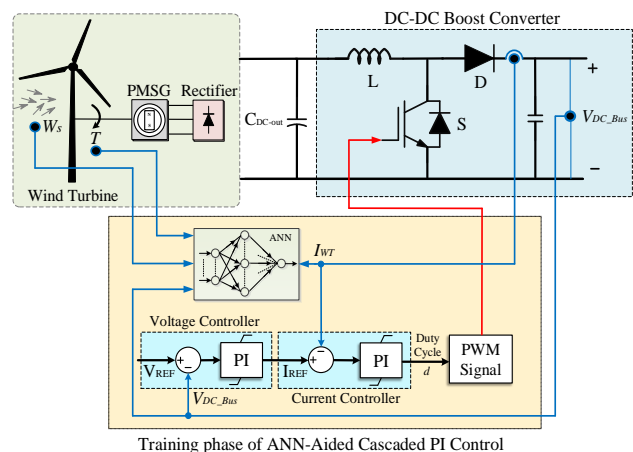

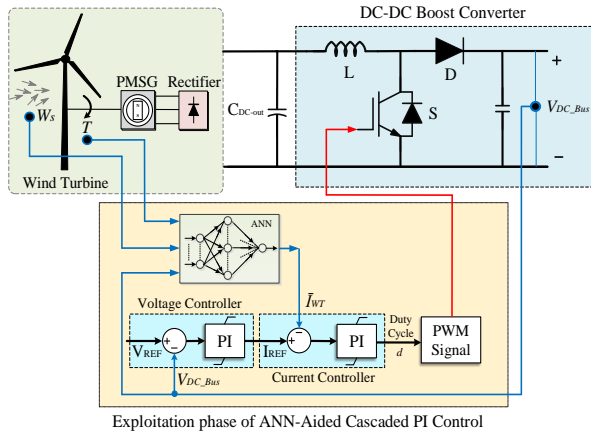**Figure 5.** Control structure of WECS by ANN-Aided PI-cascaded current and voltage controllers in offline phase.

Figure 6. Control structure of WECS by ANN-Aided PI-cascaded current and voltage controllers in online phase.

IV. RESULTS

The results of this paper has been achieved under two different variable wind speed conditions. Training of ANN yields different results because of different initial conditions and sampling. Acquiring lower values are always better, namely close to zero. In this study, best validation performance was obtained 8.1564e-11 at 1000 epochs. In Figure 7, the variable wind speed profile which considerably affects the output is given. In terms of accurate operation, regression (R) values assess the correlation between outputs and targets. If R value converges 1 that means a close relationship exists in Table 2, otherwise, 0 is a random relationship that needs to be edited. On the other hand, mean absolute percentage error ($MAPE$) is the average squared difference between outputs and targets. The advantages of the proposed neural network (NN)-based methods are expressed in Table 3 based on the literature. Estimating the data with high accuracy which influences the system positively is profitable. Furthermore, it is clear that the amount of percent error is less than 10%, it is considerable that is a high accuracy estimator at every moment of the simulation.

Table 2. Performance criteria of ANN under study.

| TRAINING PART | CRITERIA | |
	BEST PERFORMANCE OF *MSE*	REGRESSION-R
Training	5.1e-08	0.99
Validation	9.9e-08	0.99
Test	3.4e-07	0.98
Overall	7.4e-08	0.99

Table 3. Performance criteria of ANN under study.

SCHEME REF.	APPLICATION	BENEFIT	OVERALL ACCURACY
ANN-Based MPC [19]	Control of power converters	Less computational effort and better attenuation	98.25 %
FCS-MPC-based NN Classifier [20]	Voltage sag classification in PV system	Enabling voltage support	98.60 %
Long short-term memory NN (LSTM) [21]	Power prediction for virtual plants	Combining the concepts of ANN and ML	91 %
Proposed ANN-Based PI [13]	Sensorless control of DC-DC converter	Less cost, fewer needed sensors, larger reliability tolerance	98.85 %
Multilayer Perception ANN [22]	Intelligent Energy Management	Minimization of operation costs and emission	96.01 %
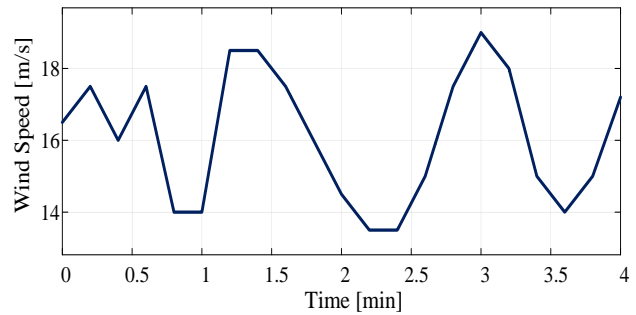
Figure 7. Variable wind speed profile-I.
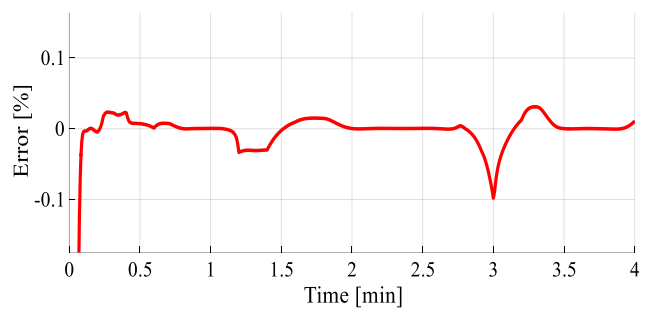
Figure 8. Percentage error values of estimated current.

The estimated value can be considered that is relatively low as seen in Figure 8. Also, Figure 9 shows one of the best validation performance of *MSE* results during the training process for 1000 epoch.
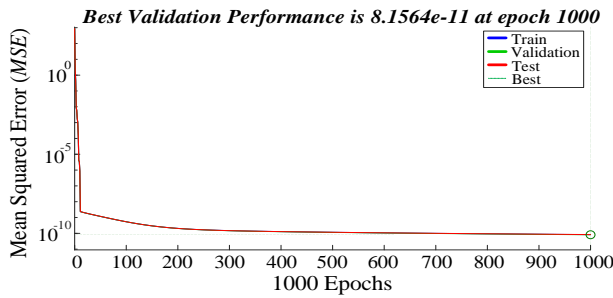
Figure 9. *MSE* value for output until 1000 epoch size.

Upon considering the performance criteria of the trained ANNs, the equations show the metrics in equations (25) and (26) to obtain high accuracy rates. As mentioned before, *R* should reflect convergence to 1, and mean squared error *MSE* should be near 0 as possible as much.

$$R = \frac{1}{m} \sum_{t=1}^{m} \left(\frac{Y_{ref}(t) - \mu_{Y_{ref}}}{\sigma_{Y_{ref}}} * \frac{Y(t) - \mu_Y}{\sigma_Y} \right), \quad (25)$$

$$MSE = \sqrt{\frac{1}{m} \sum_{t=1}^{m} (Y_{ref}(t) - Y(t))^2}, \quad (26)$$

where, $Y_{ref}(t)$ is the desired value, $Y(t)$ is the estimated value of the proposed method, and also μ and σ are the mean and standard deviation values, respectively. As shown in Figure 10, since the estimated value set up bears a very close resemblance to the measured data, our findings validate the usefulness of ANN-aided control for power converter with a complementary and easy way. It is clear that the estimated value can perform instead of the measured value by eliminating the relevant sensor data. For power estimation, measured and estimated powers can be shown in Figure 11.
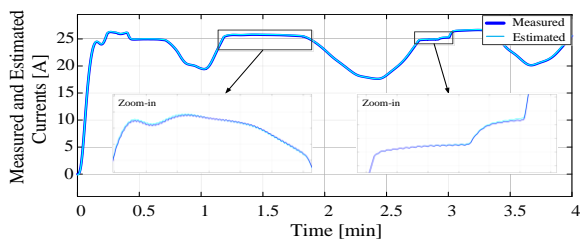
Figure 10. Measured and estimated currents.
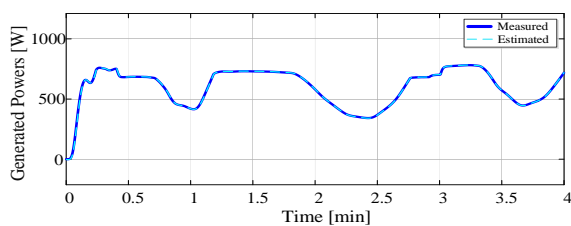
Figure 11. Measured and estimated powers regarding generated power in WECS.
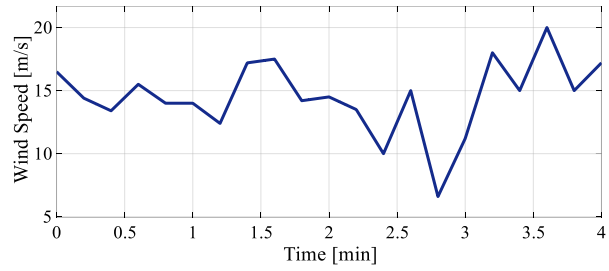
Figure 12. Variable wind speed profile-II.
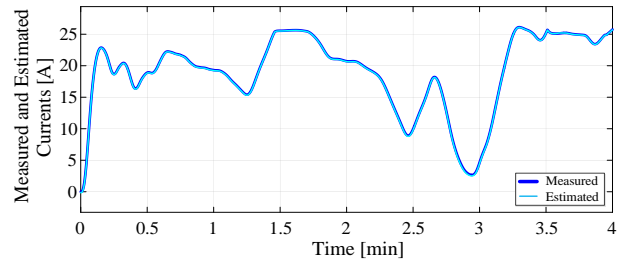
Figure 13. Measured and estimated currents.
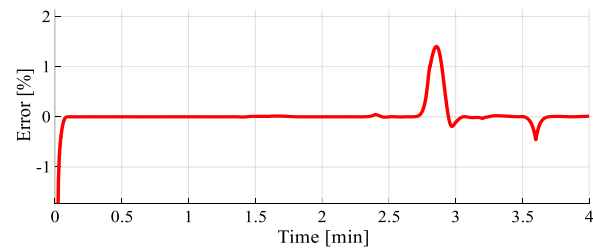
Figure 14. Percentage error values of estimated current.

For a better illustration, the second variable wind speed profile has been defined to the system to test under different conditions, as seen in Figure 12. Subsequently, measured and estimated currents in WECS is given in Figure 13. It is clear that abrupt changes in wind speed makes the system more vulnerable; however, the MAPE of this operation is still within an acceptable band in Figure 14. Once changing the percentage allocation of the target time steps for training, validation, and testing parts, the trained network achieves reaching accuracy rates as well due to having a static nonlinear relationship between inputs and output.

V. CONCLUSION
AI-aided namely ANN-based control of a DC-DC boost power converter in WECS was implemented in this work. The results of this work show us that the proposed method can perform like a conventional controller with a high accuracy rate. Additionally, that method enables reducing sensor requirements from the main control layer, which makes the system more reliable and cost-effective. Also, the communication challenges with sensors are diminished.

That research could possibly support the people who are interested in applied AI in energy systems. On the other hand, some better features of the system were

obtained such as quick response and less complexity without steady-state oscillation. Depending on the gathered dataset, it is possible to estimate the instantaneous power prediction from the WECS. In terms of the findings, our results are encouraging and should be applied by real-time experimental system.

REFERENCES

[1] Kolar, J. W., Biela, J., Waffler S., Friedli, T. & Badstuebner, U., (2011). Performance trends and limitations of power electronic systems. *2010 6th International Conference on Integrated Power Electronics Systems*, Nuremberg, pp. 1-20.

[2] Melicio, R., Mendes, V. M. F. & Catalao, J. P. S., (2010). Power converter topologies for wind energy conversion systems: Integrated modeling, control strategy and performance simulation. *Renewable Energy*, 35 (10), pp. 2165-2174.

[3] Hannan, M. A., et al. (2019). Power electronics contribution to renewable energy conversion addressing emission reduction: Applications, issues, and recommendations. *Applied Energy*, 251, p. 113404.

[4] Soliman, M. A., Hasanien, H. M., Azazi, H. Z., El-Kholy, E. E., & Mahmoud, S. A., (2019). An Adaptive Fuzzy Logic Control Strategy for Performance Enhancement of a Grid-Connected PMSG-Based Wind Turbine. *IEEE Transactions on Industrial Informatics*, 15 (6), pp. 3163-3173.

[5] Zhang, Y., Wang, Z., Wang, H., & Blaabjerg, F., (2020). Artificial Intelligence-Aided Thermal Model Considering Cross-Coupling Effects. *IEEE Transactions on Power Electronics*, 35 (10), pp. 9998-10002.

[6] Bayhan, S., Demirbaş, S., & Abu-Rub, H., (2016). Fuzzy-PI-based sensorless frequency and voltage controller for doubly fed induction generator connected to a DC microgrid. *IET Renewable Power Generation*, 10 (8), pp. 1069-1077.

[7] Mesbahi, A., Aljarhizi, Y., Hassoune, A., Khafallah, M., & Alibrahmi, E., (2020). Boost Converter implementation for Wind Generation System based on a variable speed PMSG," *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, pp. 1-6.

[8] Deng, X. et al. (2019). Sensorless effective wind speed estimation method based on unknown input disturbance observer and extreme learning machine. *Energy*, 186 (115790).

[9] Wei, C., Zhang, Z., Qiao, W., & Qu, L., (2016). An Adaptive Network-Based Reinforcement Learning Method for MPPT Control of PMSG Wind Energy Conversion Systems. *IEEE Transactions on Power Electronics*, 31 (11), pp. 7837-7848.

[10] Chatri, C., & Ouassaid, M., (2018). Sensorless Control of the PMSG in WECS Using Artificial Neural Network and Sliding Mode Observer. *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Rabat, Morocco, pp. 1-6.

[11] Jday, M., & Haggège, J., (2017). Modeling and neural networks based control of power converters associated with a wind turbine. *2017 International Conference on Green Energy Conversion Systems (GECS),* Hammamet, Tunisia, pp. 1-7.

[12] Adineh, B., Habibi, M. R., Akpolat, A. N., & Blaabjerg, F., (2021). Sensorless Voltage Estimation for Total Harmonic Distortion Calculation using Artificial Neural Networks in Microgrids. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 68 (7), pp. 2583-2587.

[13] Akpolat, A. N., Dursun, E., & Kuzucuoğlu, A. E., (2020). AI-Aided Control of a Power Converter in Wind Energy Conversion System. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, pp. 1-6, doi: 10.1109/ASYU50717.2020.9259877.

[14] Samara, S., & Natsheh, E., (2019). Intelligent Real-Time Photovoltaic Panel Monitoring System Using Artificial Neural Networks. *IEEE Access*, 7, pp. 50287-50299.

[15] Akpolat, A. N., Habibi, M. R., Dursun, E., Kuzucuoğlu, A. E., Yang, Y., Dragičević, T., & Blaabjerg, F., (2021). Sensorless Control of DC Microgrid Based on Artificial Intelligence. *IEEE Transactions on Energy Conversion*, 36 (3), pp. 2319-2329.

[16] Egea-Àlvarez, A., Aragüés-Peñalba, M., Gomis-Bellmunt, O., Rull-Duran, J., & Sudrià-Andreu, A., (2016). Sensorless control of a power converter for a cluster of small wind turbines. *IET Renewable Power Generation*, 10 (5), pp. 721-728.

[17] Teiar, H., Chaoui, H., & Sicard, P., (2015). Almost parameter-free sensorless control of PMSM. *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, Yokohama, pp. 004667-004671.

[18] Syskakis. T., & Ordonez, M., (2019). MPPT for Small Wind Turbines: Zero-Oscillation Sensorless Strategy. *2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, Xi'an, China, pp. 1060-1065.

[19] Akpolat, A. N., Habibi, M. R., Baghaee, H. R. Dursun, E., Kuzucuoğlu, A. E., Yang, Y., Dragičević, T., & Blaabjerg, F., (2021). Dynamic Stabilization of DC Microgrids using

ANN-Based Model Predictive Control. *IEEE Transactions on Energy Conversion*, Early Access, doi: 10.1109/TEC.2021.3118664.

[20] Khan, M. A., Haque, A., Kurukuru, V. S. B., & Saad, M., (2020). Advanced Control Strategy with Voltage Sag Classification for Single-Phase Grid-Connected Photovoltaic System," *IEEE Trans. Emerg. Sel. Topics Ind. Electron.*, Early Access, doi: 10.1109/JESTIE.2020.3041704.

[21] Rosato, A., Panella, M., Araneo, R., & Andreotti, A., (2019). A Neural Network Based Prediction System of Distributed Generation for the Management of Microgrids. *IEEE Trans. Ind. Appl.*, 55 (6), pp. 7092-7102.

[22] Chaouachi, A., Kamel, R. M., Andoulsi, R., & Nagasaka, K., (2013). Multiobjective Intelligent Energy Management for a Microgrid. *IEEE Trans. Ind. Electron.*, 60 (4), pp. 1688-1699.

[23] Peyghami, S., Dragicevic, T., & F. Blaabjerg., (2021). Intelligent long-term performance analysis in power electronics systems. Sci. Rep., 11 (1), pp. 1–18.

[24] Habibi, M. R., Baghaee, H. R., Blaabjerg, F., & Dragičević, T., (2021). Secure Control of DC Microgrids for Instant Detection and Mitigation of Cyber-Attacks Based on Artificial Intelligence. Early Access, *IEEE Systems Journal*, doi: 10.1109/JSYST.2021.3119355.

[25] Farooq, Z., Zaman, T. M., Khan, A., Nasimullah, Muyeen, S. M., & Ibeas A., (2019). Artificial Neural Network Based Adaptive Control of Single Phase Dual Active Bridge With Finite Time Disturbance

Compensation. *IEEE Access*, 7, pp. 112229-112239.

[26] Li, X., Zhang, X., Lin, F., & Blaabjerg, F., (2021). Artificial-intelligence based design (AI-D) for circuit parameters of power converters. *IEEE Trans. Ind. Electron.*, Early Access, doi: 10.1109/TIE.2021.3088377.

[27] Wang, T. He, X., & Deng, T., (2019). Neural networks for power management optimal strategy in hybrid microgrid. *Neural Comput & Applic* 31, pp. 2635–2647.

[28] Onar, O. C., & Khaligh, A., (2015). Alternative Energy in Power Electronics. Chapter 2 - Energy Sources, 1st ed., UK: Butterworth-Heinemann, Elsevier, 2015, pp. 81-154.

[29] Heier, S., (2014). Wind Energy Conversion System. *Grid Integration of Wind Energy: Onshore and Offshore Conversion Systems,* 3rd ed., Germany: John Wiley & Sons, pp. 31-117.

[30] Akpolat, A.N., (2021). Management And Control of Distributed Energy Generation Systems via Artificial Intelligence Techniques. Doktora Tezi, Marmara Üniversitesi, Türkiye, pp. 37-38.

[31] Piccinini, G., (2004). The first computational theory of mind and brain: A close look at McCulloch and Pitts's Logical Calculus of Ideas Immanent in Nervous Activity. *Synthese*, 141, pp. 75–215.

[32] Rathnayaka, R. M. K. T., & Seneviratna, D. M. K. N., (2019). A Novel Hybrid Back Propagation Neural Network Approach for Time Series Forecasting Under the Volatility. *Communications in Computer and Information Science*, Singapore.

A Comparative Study of Point-Based Deep Learning Techniques for Semantic Classification in Search and Rescue Arenas

Arama ve Kurtarma Alanlarında Anlamsal Sınıflandırma için Nokta Tabanlı Derin Öğrenme Tekniklerinin Karşılaştırmalı Bir Çalışması

Kaya TURGUT [1] , **Burak KALECİ** [1]

[1]*Eskisehir Osmangazi University, Electrical and Electronics Engineering Department, Eskisehir, Turkey*

Abstract

Post-disaster indoor environments, which occur after calamities such as floods, fires, and poisonous material spread, could include serious risks for search and rescue teams. For example, the building's structural integrity could be corrupted, and some harmful substances for humans and animals could exist. Exploiting robots could prevent search and rescue teams from these risks. Nevertheless, robots need to possess advanced techniques to produce high-level information from raw sensor data in these harsh environments. This study aims to explore the positive and negative aspects of point-based deep learning architectures for the semantic classification of ramps in search and rescue test arenas, which are proposed by the National Institute of Standards and Technology (NIST). Also, we take into account walls and terrain since they can provide crucial information for robots. In this study, we opted to utilize point cloud data that is robust against lousy illumination conditions, which is frequently encountered in post-disaster environments. We used the ESOGU RAMPS dataset that contains point cloud data captured from a simulated environment similar to NIST search and rescue arenas. We selected PointNet, PointNet++, Dynamic Graph Convolutional Neural Network (DGCNN), PointCNN, Point2Sequence, PointConv, and Shellnet point-based deep learning architectures to analyze their performance for semantic classification of ramps, walls, and terrain. The test results indicate that accuracy of semantic classification is over 90% for all architectures.

Keywords: point-based deep learning, semantic classification, NIST ramps, point cloud data.

Öz

Sel baskını, yangın ve zehirli madde yayılımı gibi felaketlerden sonra meydana gelen afet sonrası iç ortamlar, arama ve kurtarma ekipleri için ciddi riskler barındırabilir. Örneğin, binanın yapısal bütünlüğü bozulmuş ve insanlar ve hayvanlar için bazı zararlı maddeler mevcut olabilir. Arama ve kurtarma ekiplerinin bu risklerden korunmasını sağlayabilmek için robotlardan yararlanılabilir. Bununla birlikte, robotların bu zorlu ortamlarda ham algılayıcı verilerinden üst düzey bilgi üretmek için gelişmiş tekniklere sahip olması gerekir. Bu çalışma, Ulusal Standartlar ve Teknoloji Enstitüsü (NIST) tarafından önerilen arama kurtarma test alanlarında bulunan rampaların anlamsal sınıflandırması için nokta tabanlı derin öğrenme mimarilerinin olumlu ve olumsuz yönlerini araştırmayı amaçlamaktadır. Ayrıca robotlar için çok önemli bilgiler sağladıklarından dolayı duvarlar ve zeminde dikkate alınmıştır. Bu çalışmada, afet sonrası ortamlarda sıklıkla karşılaşılan kötü aydınlatma koşullarına karşı dayanıklı olan nokta bulutu verilerini kullanmayı tercih ettik. NIST arama ve kurtarma alanlarına benzer bir ortamdan alınan nokta bulutu verilerini içeren ESOGU RAMPS veri kümesini kullandık. Rampaların, duvarların ve zeminin anlamsal sınıflandırma performanslarını analiz etmek için PointNet, PointNet ++, Dinamik Grafik Evrişimli Sinir Ağı (DGCNN), PointCNN, Point2Sequence, PointConv ve Shellnet nokta tabanlı derin öğrenme mimarilerini seçtik. Test sonuçları, anlamsal sınıflandırma doğruluğunun tüm mimariler için %90'ın üzerinde olduğunu göstermektedir.

Anahtar Kelimeler: nokta tabanlı derin öğrenme, anlamsal sınıflandırma, NIST rampları, nokta bulutu verisi.

I. INTRODUCTION

In recent years, attempting to use robots for search and rescue missions in post-disaster environments, which occur after calamities such as floods, fires, and poisonous material spread, is among the hot topics in the robotic community. The main reasons for that are happening new collapses or leakage of harmful substances for humans and animals while search and rescue teams perform their tasks. Therefore, exploiting robots in these missions could prevent undesirable accidents and new casualties. However, post-disaster environments consist of some challenges even for robots such as uneven terrain and lousy illumination conditions due to these environments' dusty nature and power outages. To cope with these difficulties, robots need to possess advanced techniques to produce high-level information (semantic information of scenes or objects) from raw sensor data. Fortunately, robots have begun to utilize these advanced techniques owning to incoming perception technologies and developments in computer vision algorithms. Nevertheless, the advantages and disadvantages of these technologies and algorithms should be observed regularly for determining new research areas in the search and rescue domain. To achieve this, competitions related to search and rescue missions have been organized yearly by RoboCup society since 2001.

Corresponding Author: Burak KALECİ, Tel: +905557375966, e-posta: burakaleci@gmail.com
Submitted: 15.03.2021, **Revised:** 07.07.2021, **Accepted:** 07.07.2021

The RoboCup rescue competitions mainly aim to improve robots' abilities in autonomous navigation, mapping, and finding victims. Besides, these competitions monitor the performance of new technologies in software and hardware and encourage the researchers to introduce new challenges for robots. Kitano and Tadokoro [1] assessed the first RoboCup rescue competition considering the initial standards and evaluation metrics. They examined requirements that the robots need to have, and they projected future works according to these requirements. One of them was to evaluate robot performance in standard test environments. For that reason, NIST introduced reference test arenas for search and rescue missions in 2003 [2]. Figure 1 shows an example of these test arenas. In the first years of competitions, participant teams generally preferred to use visual and 2D laser range data to perform search and rescue missions. For example, the PoAReT team [3], the winner of the Virtual Robot Competition at RoboCup 2012, applied the simultaneous localization and mapping (SLAM) technique to 2D laser scans for building a map of reference test arenas. They also produced semantic information while determining rooms and corridors through the detection of doorway locations. This was the pioneering study in the search and rescue missions that considers semantic information to the best of our knowledge.

Sheh et al. [4] evaluated the RoboCup rescue competitions in the 16th year. They noticed that participant teams of competitions improved the abilities of robots while integrating new sensors into robot systems such as RGB-D cameras, LiDARs, and 3D laser scanners over the years. Besides, they observed that robots were able to perform search and rescue missions adequately in challenging environments. Robot Operating System (ROS) [5] have been begun to use in the RoboCup rescue competitions since 2017. This could be a milestone for competitions because ROS provides numerous packages that researchers can easily integrate their systems to achieve more complex tasks. The participant teams utilized GMapping [6] and Hector SLAM [7] packages to represent the environments with 2D maps. Then, OctoMap [8] and Real-Time Appearance-Based Mapping (RTAB-Map) [9] approaches were used to produce 3D information about the environment. OctoMap is an octree-based representation method, and it does not take into account semantic information. On the other hand, RTAB-Map exploited feature extractors to generate semantic information while recognizing frequently encountered objects in daily life. However, RTAB-Map is not able to extract semantic information for ramps at reference test arenas since ramps are specific pieces for search and rescue missions.

Figure 1. An example reference test arena [10]

Although participant teams of the RoboCup rescue competitions generally do not concentrate on producing semantic information for search and rescue missions, semantic classification of walls, ramps, and terrain via point cloud data has been addressed in previous studies. These approaches could be separated into two categories. In the first category, the studies handled the segmentation problem, which clusters points into a segment considering their features such as x, y, z coordinates, point normals, and colors. The reviews proposed by Nguyen and Le [11], Grilli et al. [12], and Xie et al. [13] mainly divide the segmentation approaches into five groups: Edge-based [14], region growing [15], model-based [16, 17], clustering-based [18], and graph-based [19]. These approaches have been frequently applied for segmentation problems since their implementations are available on Point Cloud Library (PCL) [20]. Besides, Eruyar et al. [21] examined the performance of segmentation approaches situated in PCL for structural planar surfaces. However, only region growing and RANSAC approaches were employed to classify walls, ramps, and terrain. Region growing begins with determining seed points and searches points within a predetermined radius or a number of neighbors around the seed points to identify points that have similar features with seed points. The main disadvantages of region growing approaches are high time complexity and sensitivity to rapid changes on point features. In contrast, RANSAC requires less computational load, and robust against noise, in other words, rapid changes. However, it does not utilize local information for segmentation, and this can cause clustering points that have similar mathematical models into a segment, although they belong to different planes. The approaches placed in the first category classified points considering plane equations to obtain semantic classes of points.

In the second category, machine and deep learning techniques were used to determine the semantic class of points. Machine learning techniques firstly determine suitable descriptors for the problem, and then they extract features. This step is probably the most challenging part because the selected descriptors directly affect the success of these techniques. Besides, 3D descriptors generally tend to overfit since

they are significantly high dimensional. On the other hand, deep learning architectures have become popular in recent years since they are able to provide useful features according to the problem. For example, Deng et al. [22] applied Convolutional Neural Network (CNN) approach to visual data and depth images acquired from a post-disaster environment similar to NIST reference test arenas to determine the semantic class of points.

The first attempts to exploit deep learning techniques with point cloud data are to use CNN because of success of CNNs with image data. However, CNN approaches cannot be employed to point cloud data directly since its permutation invariant and unstructured characteristic. For that reason, the deep learning techniques that have utilized point cloud data can be separated into two groups: direct (point-based) and indirect approaches [23]. Point-based approaches accept raw point cloud data, while indirect approaches convert unstructured point cloud data into a 2D or 3D structured form before receiving the data. It can be observed from the previous studies that indirect approaches have some drawbacks such as quantization artifact, loss of the geometric details, and computational cost of conversion. In consequence, the researchers have generally preferred to develop point-based approaches [23, 24]. Guo et al. [24] gave detailed information about 3D deep learning approaches and categorized them for classification, segmentation, object detection, and tracking problems. They addressed the point-based methods for semantic classification of point cloud data into four groups: point-wise Multi-Layer Perceptrons (MLP), point convolution, graph-based, and Recursive Neural Networks (RNN) based.

PointNet [25] was the first point-based deep learning architecture accepted as a milestone since it works directly on unordered and permutation-invariant point cloud data. PointNet utilizes the point-wise MLP followed by maximum pooling to summarize the global feature. However, PointNet does not extract the local relationship because it considers all points in the point cloud as individually. Many point-wise architectures like PointNet++ [26], ShellNet [27], and PointWeb [28] were proposed based on PointNet to encode local neighborhood information because of simplicity and performance.

Applying the convolution kernel to the point cloud is not straightforward because data may have missing parts and does not have a regular pattern. Some studies applied continuous convolution [29, 30] or discrete convolution [31] to point cloud data. KPConv [29] learns the weights of kernel points defined in Euclidean space. Linear correlation is applied between kernel points and the points around the kernel points. In the PointConv [30], convolution kernels defined as nonlinear weighting function on 3D space. The kernel

weights are learned with MLP layers. PointCNN [31] learns the transformation matrix to order canonically for applying the discrete convolutional operator.

In the graph-based methods, the points are considered as nodes, and the distance between nodes is treated as edges [32, 33]. Landrieu et al. [32] extended the superpixel term of images as superpoint for point clouds to partition into homogeneous parts. They introduced the superpoint graph to expose the contextual information between object parts. DGCNN [33] uses each point as a node and defines the graph for each local region. The features are extracted over edges, and the graph is updated in all layers.

RNN-based architecture aims to capture contextual information of local parts [34-36]. 3P-RNN [34] utilized a two-directional RNN structure to exploit the long-range relationship of uniformly-spaced blocks along x and y directions. Point2Sequence [35] used the RNN structure to extract the correlation of multi-scale local areas. An attention mechanism is used to highlight the critical feature of multi-scale local areas. RSNet [36] proposed the slice pooling layer, which slices the point cloud x, y, and z coordinate independently to project unordered point cloud onto a sequential form.

In this study, semantic classification of point cloud data as walls, ramps, and terrain is aimed. Producing semantic information about walls, ramps, and terrain could promote the mapping and navigation tasks of robots in a post-disaster environment. An example of that is the robot could augment the environment's representation with semantic information to generate maps that first-responders easily read. In this way, the robot can also improve its path plan while adding appropriate targets according to ramps. The robot can navigate without losing its balance when it passes over the ramps considering these target points. Besides, the robot can regulate its speed when being aware position and orientation of ramps.

In this study, we prefer to use point cloud data because visual and/or 2D laser range data could not be adequate for post-disaster environments. The success of studies that utilized visual data highly depends on the illumination condition of the environment, and the post-disaster environments have lousy illumination conditions due to these environments' dusty nature and power outages. Although 2D laser range data robust against these conditions, it only captures information about the plane at the height that the laser scan is placed. Besides, 2D laser range data cannot provide any information about the ramps below that height. On the other hand, point cloud data could describe 3D characteristics of walls, ramps, and terrain and cannot be influenced by the environment's illumination condition. After that point, we analyzed point-based and indirect deep learning approaches for semantic

classification of point cloud data, and we decided to use point-based approaches when disadvantages of indirect approaches are thought. In our previous works [37, 38], we showed that PointNet, PointNet++, DGCNN, and PointCNN point-based deep learning architectures are classified walls, ramps, and terrain with over 90% accuracy. This study aims to extend our previous works by considering Point2Sequence, PointConv, and Shellnet point-based deep learning architectures. In this way, these architectures' positive and negative aspects could be revealed for semantic classification of walls, ramps, and terrain in reference test arenas. To train and test these architectures, we used the ESOGU RAMPS dataset [39]. The architectures were evaluated with recall, precision, Intersection over Union (IoU), Mean Intersection over Union (MIoU), and accuracy metrics.

The rest of the paper is organized as follows: Section 2 briefly explains the point-based deep learning architectures. Section 3 describes the ESOGU RAMPS dataset. Section 4 and 5 presents experimental works and concludes the paper, respectively.

II. METHODS

The point cloud data is defined as a set of points that contain x, y, and z coordinates. Besides, other features such as point normal and color could be attached to each point. Although point cloud has a simple nature, a variety of architectures were proposed due to its unstructured, permutation, and rotation invariant characteristic. In this section, notable attributes of PointNet, PointNet++, DGCNN, PointCNN, Point2Sequence, PointConv, and Shellnet point-based deep learning architectures are concisely described.

2.1. PointNet

The first architecture that receives raw point cloud data, in other words, point-based deep learning architecture, is PointNet. It employs successive Multi-Layer Perceptron (MPL) layers to point's x, y, and z coordinates to learn weight matrices. These matrices are shared among points for each feature channel as similar to CNN structure. Besides, normal, curvature, and color features can be used in the feature extraction process. It is important to note that PointNet assesses each point individually and also independently from other points. Namely, neighbor points of a point do not take into account for feature extraction. This could be the most characteristic feature of the PointNet architecture. PointNet provides a maximum pooling method named as the symmetric function to obtain the global feature. The symmetric function receives the feature vector of all points and determines each feature channel's maximum values to summarize a single global feature vector. The classification scores for each category of points are acquired by successive MLP layer again after the global and local features are concatenated.

2.2. PointNet++

PointNet++ architecture derives from the PointNet. The main difference between PointNet and PointNet++ is that PointNet++ considers neighbors of points to extract the feature vector of each point. The architecture first selects center points to construct local regions. These center points are specified according to a predefined number of nearest neighbors or a predefined radius. After local regions are determined, the feature vector of each point that belongs to a local region is extracted by employing PointNet to point's x, y, and z coordinates. The symmetric function then takes the feature vector of all points in a local region and determines the region feature vector. This process is repeated for each local region. At that point, PointNet++ continues with the successive layer while extending the local regions hierarchically based on x, y, and z coordinates. In this way, local regions cover large portions, and the region feature vectors begin to approximate the scene's characteristics. After successive layers, similar to PointNet, the symmetric function summarizes local region feature vectors to obtain a global feature vector. In successive feature encoding layer, point number is subsampled. However, all original points feature is required to classify points semantically. A distance based interpolation technique is proposed to propagate the features of sampled points to original points. After interpolated feature are concatenated skip linked feature from feature encoding layer, PointNet is used to update these features. Eventually, the semantic class of each point is determined.

2.3. DGCNN

DGCNN architecture is categorized as graph-based, according to the review presented by Guo et al. [24], because it constructs local regions similar to PointNet++, and it builds a graph for each local region. The main difference between PointNet++ and DGCNN is that DGCNN does not expand local regions hierarchically. DGCNN first selects center points to construct local regions. Then, it searches a predefined number of nearest neighbors of these center points to determine the points that belong to local regions. To do that, DGCNN uses distances between a point and center points, which are calculated with point's x, y, and z coordinates in the first layer, while feature space distances are utilized in successive layers. At that point, the graphs for each local region are formed. The nodes of these graphs are the points in the local regions. However, it is crucial to notice that the edges only exist for the center point and other points. The edge weights are calculated considering spatial and feature space distances in the first and successive layers, respectively. This is the second difference between PointNet++ and DGCNN. After edge weights are determined, they are used in the feature extraction process by employing MLPs. The authors are named these steps as EdgeConv operator. In the DGCNN architecture, first EdgeConv

operator estimates features, and then PointNet accepts these features to classify points.

2.4. PointCNN

In contrast to PointNet, PointNet++, and DGCNN architectures that consider points individually while extracting features, PointCNN calculates features by applying convolution to a point and its neighbors. Unfortunately, convolution approaches, just like CNN, cannot be directly employed to point cloud data due to its permutation invariant nature. For that reason, the authors proposed a convolution operator, which is named as X-Conv. PointCNN first constructs local regions similar to previously mentioned architectures. Then, for each local region, a transformation matrix is learned through X-Conv operator, and according to the matrix, points in a local region are weighted and canonically ordered. Lastly, convolution is employed local regions to extract features. Similar to PointNet++ local regions are hierarchically expanded in the successive layers. PointCNN propagate summarized global into point feature with skip linked X-Conv.

2.5. Point2Sequence

Point2Sequence is an RNN-based architecture. Apart from the previously explained architectures, Point2Sequence does not have successive layers since it utilizes Long Short-Term Memory (LSTM) structure to extract features. Point2Sequence constructs local regions just like PointNet++. The main difference between Point2Sequence and PointNet++ is that Point2Sequence constructs more than one local region with different radius values around a center point, namely multi-scale concentric local regions. Then, features are determined separately for each different-scale local region. After features corresponding to multi-scale concentric local regions are obtained for a center point, RNN is employed to learn the correlation between these features. In this way, Point2Sequence intends to reveal contextual information about local regions. The relationship between features that belong to a local region is stored in LSTM, and the attention approach puts forward hidden states of LSTM. Consequently, the local region features are calculated through a context vector, which is built consolidation of different scale features. The global feature vector is propagated from shape level to point level by using the interpolation techniques in PointNet++.

2.6. PointConv

The continuous convolution-based PointConv applies the convolution operator to each point in the point cloud in a similar way as traditional 2D image convolution. Firstly, the local region is defined around each point, and the convolution kernel weight is learned by using weight-shared MLP over the relative position of points in the local neighborhood. Because it uses the weight-shared kernel across all points, permutation invariance is satisfied. The density function is used to re-weight convolution filter weight to handle non-uniform density in local regions. The learned convolution kernel is applied to the feature of the local points, and the encoded feature of each point is obtained. Similar to PointNet++, the feature encoding module consists of sampling, grouping, and PointConv layer to learn feature hierarchically. Also, the PointDeconv layer is introduced to increase the point number, which is decreased in the feature encoding layers by applying interpolation and PointConv convolution.

2.7. Shellnet

ShellNet architecture consists of the ShellConv convolution operator based on point-wise MLP and convolution approaches. In this architecture, firstly, the local region is constituted in a similar way as PointNet++. However, kNN neighborhoods of representative points are divided into multi-scale concentric shells. A fixed number of points is assigned to each shell. In each shell, MLP is used to encode relative point coordinates to higher-dimensional features. After the features came from the previous layer and encoded features are concatenated, maximum pooling is used to summarize the points feature in each shell. Thanks to maximum pooling, point order ambiguity is resolved in each shell. Also, by nature of the shell, it is ordered from inside to outside. Then, 1D convolution is applied to the shell features. Shellconv layers are added hierarchically to extract the abstract feature of the entire point cloud. Deconvolution layers with ShellConv are used to increase point size to the original point number for the point-wise classification.

III. THE ESOGU RAMPS DATASET

Gazebo [40] simulation environment and ROS [5] were utilized to construct the ESOGU RAMPS dataset. An Asus Xtion Pro RGB-D camera was emplaced on a Pioneer 3-AT mobile robot, and 681 scenes were captured. For training, 581 scenes were randomly selected, and the remaining 100 scenes were separated for testing. In each scene, there are four classes: wall, terrain, inclined, and flat ramps. Examples for the ESOGU RAMPS dataset are presented in Figure 2. The point cloud data is shown in the left column. Wall, terrain, inclined, and flat ramps classes are described with red, yellow, blue, and magenta. The dataset also provides RGB images, which are shown in the right column. The details about the dataset are given in [37, 38] and you can download the dataset from [39].

IV. EXPERIMENTAL WORKS

4.1. Experimental Setup

PointNet, PointNet++, DGCNN, PointCNN, Point2Sequence, PointConv, and Shellnet point-based deep learning architectures were implemented with Tensorflow library [41] in Python programming

language for the semantic classification of scenes placed in the ESOGU RAMPS dataset. Before applying these architectures, we preprocessed the data, which is a requirement for point-based deep-learning architectures. First, the NaN values that correspond to unmeasured points were removed. Besides, using the whole scene to feed these architectures is making difficult to identify local features and causes losing data. Therefore, the point cloud data was separated into 1 m² blocks in xy plane to avoid these drawbacks. Lastly, the number of points that belong to a block must be a fixed number, and it was selected as 4096 in this study. The farthest point algorithm was employed for upsampling and downsampling to fix the number of points in blocks. Also, we discard the blocks that have less than 500 points. Although these architectures are able to process color and normal features of points, we only used point's x, y, and z coordinates. In the training process, we used default parameters for all architectures.Recall, precision,

Intersection over Union (IoU), Mean Intersection over Union (MIoU), and accuracy (Acc) metrics were used to analyze the architectures' efficiency. True positive (TP) and false positive (FP) are defined as a correctly and incorrectly classified sample, which is owned to a positive class, respectively. On the other hand, false negative (FN) is an incorrectly classified sample to a positive class while the sample belongs to a negative class. The recall value of a class is the ratio of the true positive and total number of samples of that class (Equation (1)). The ratio of the true positive and total number of classified samples of a class is called the class's precision value (Equation (2)). The Intersection over Union of a class is the ratio of the true positive and the summation of the total number of samples and incorrectly classified samples (Equation (3)). Mean Intersection over Union (MIoU) is the mean of IoU of classes (Equation (4)). Lastly, accuracy is defined as the ratio of the total number of true positive for all classes and the total number of samples (Equation (5).
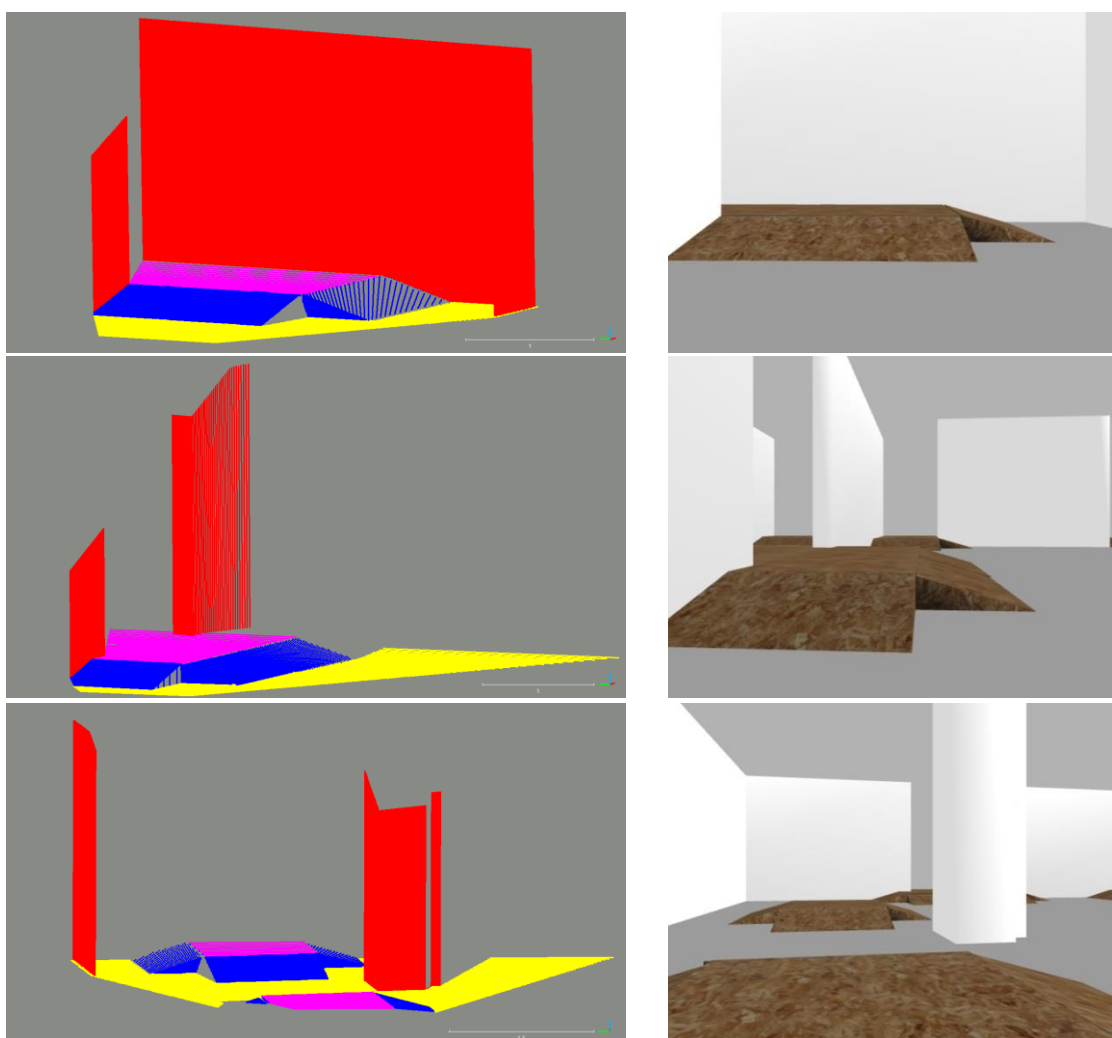
Figure 2. Examples for the ESOGU RAMPS dataset. The left column shows the point cloud data. The right column indicates the corresponding RGB images.

$$recall = \frac{TP}{TP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

$$MIoU = \frac{\sum_{i=1}^{n} IoU_i}{n} \tag{4}$$

$$Acc = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i + FP_i)} \tag{5}$$

4.2. Experimental Results

The metric results obtained on the test dataset are given in Table 1. Besides, four example scenes are selected to examine the positive and negative aspects of architectures, and they are shown in Figure 3. In the figure, rows describe the ground truth and architectures, and columns are depicted selected example scenes. Besides, white ellipses are used to emphasize the incorrectly classified regions of architectures. It has been observed that for all classes, the DGCNN architecture produces results over 99% in all metrics, and it is the most successful among architectures. In contrast to other architectures, DGCNN does not expand local regions. Besides, it utilizes x, y, and z coordinates in the feature extraction process in the first layer while it considers feature space in the successive layers. These facts are the main reasons for the success of DGCNN.
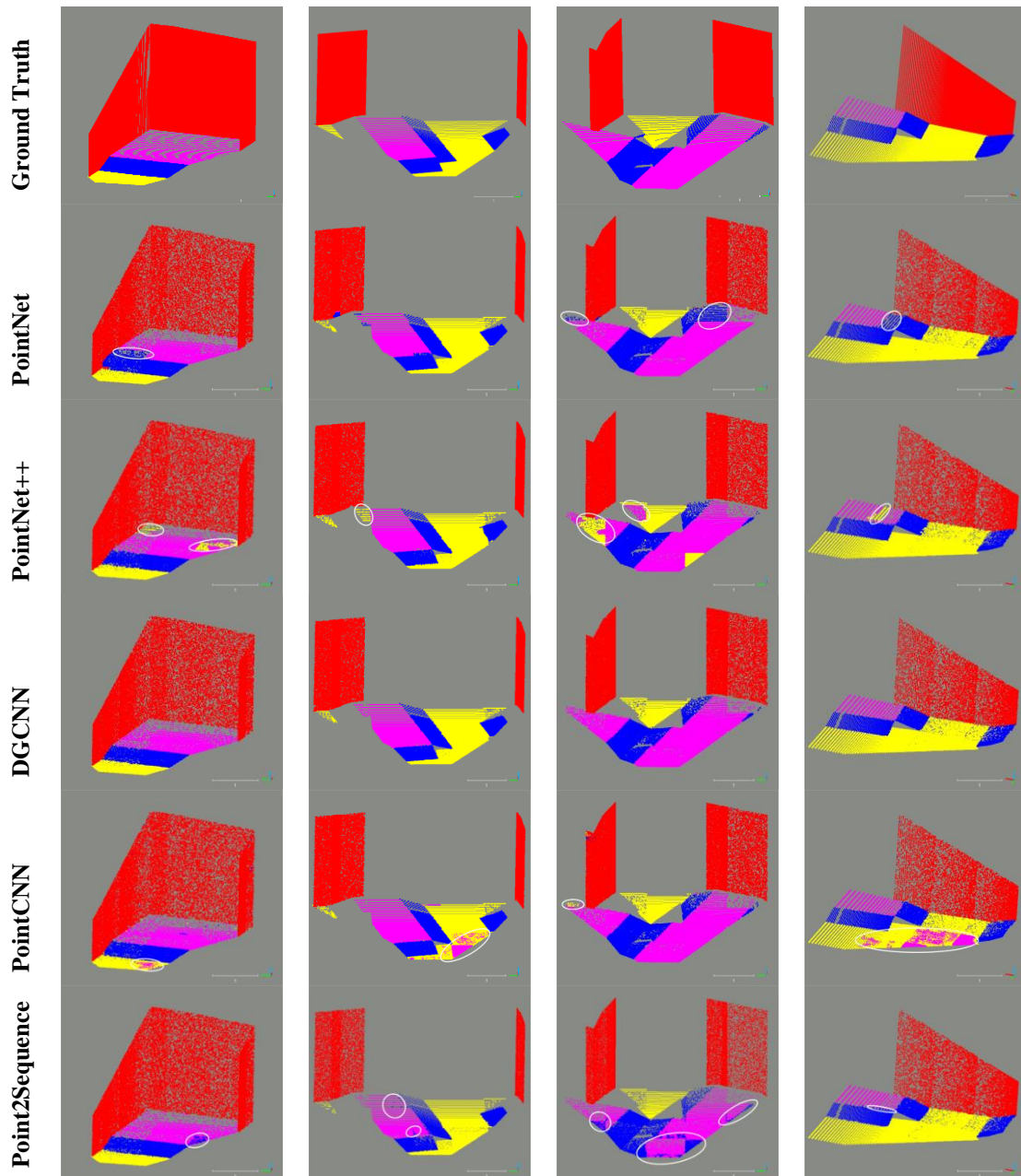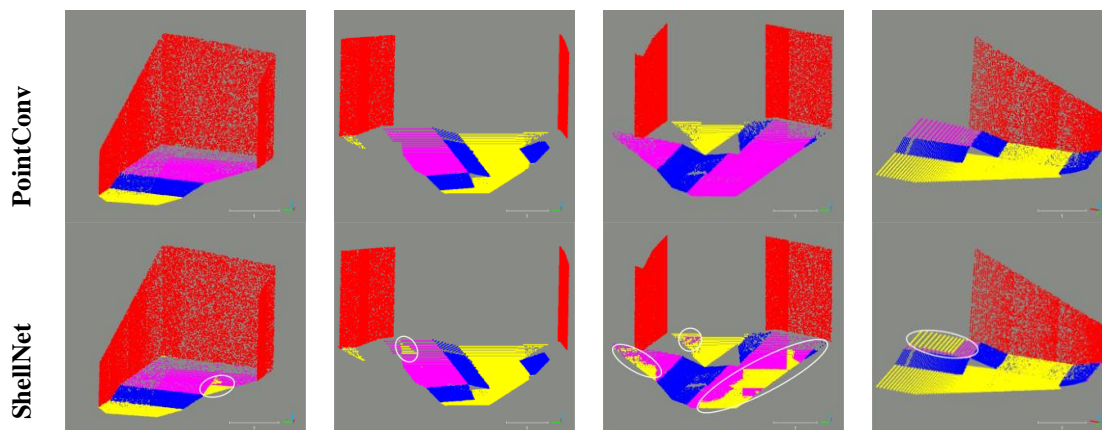
Figure 3. Visual Results

Table 1. Metric Results

	Inclined Ramp			Wall			Flat Ramp			Terrain				
	R	P	IoU	R	P	IoU	R	P	IoU	R	P	IoU	MIoU	Acc
PointNet	98.3	96.1	94.6	99.1	99.9	99.0	96.8	98.1	95.0	**99.9**	99.8	**99.8**	97.1	98.9
PointNet++	99.5	98.9	98.5	99.9	99.9	99.8	89.0	93.4	83.8	95.9	93.6	90.0	93.0	96.8
DGCNN	**99.8**	99.7	**99.6**	99.9	99.9	**99.9**	99.7	**99.8**	**99.6**	**99.9**	99.9	**99.8**	**99.8**	**99.9**
PointCNN	99.7	99.4	99.2	99.5	99.9	99.4	99.6	64.6	64.4	67.0	99.4	66.8	82.5	90.4
Point2Sequence	98.7	85.9	84.9	99.8	**100.0**	99.8	81.1	**99.8**	80.9	99.5	99.0	98.5	91.0	96.3
PointConv	97.9	**99.9**	97.8	**100.0**	99.7	99.7	**99.9**	98.3	98.2	**99.9**	99.7	99.7	98.8	99.5
ShellNet	99.7	**99.9**	**99.6**	**100.0**	**100.0**	**99.9**	71.2	93.5	67.9	97.1	84.7	82.7	87.5	94.2

PointConv also yielded successful results similar to DGCNN. PointConv learns the convolution kernel weight through weight-shared MLP over the relative position of points in the local neighborhood. Then, it re-weight the convolution kernel according to the density of points. In this way, PointConv understands the general characteristic of a scene while avoiding suppressed features that correspond to dense local regions.

PointNet assesses each point individually and also independently from other points. Therefore, it may produce erroneous results where the regions in which the points that belong to different classes are neighbors to each other. Examples for these erroneous regions are given in Figure 3. It is seen that wall and terrain classes are recognized successfully when the metric results are examined. However, PointNet may confuse flat and inclined ramp classes since they are generally placed together in scenes.

PointNet++ classifies wall and inclined ramp classes with over 98% in all metrics. However, in some cases, it may not distinguish between terrain and flat ramp classes, as shown in Figure 3. PointNet++ exploits local information when it extracts point features similar to DGCNN. However, it expands the local regions in successive layers. This may cause confusing terrain and flat ramp classes since PointNet++ generally considers exact feature of points while the DGCNN utilizes the difference feature (in other words edges) of points.

Point2Sequence, similar to PointNet, can identify wall and terrain classes. Point2Sequence constructs more than one local region with different radius values around a center point. Then, the features for these multi-scale concentric local regions are aggregated with the LSTM mechanism. Point2Sequence may yield erroneous results for the flat and inclined ramp classes. The reason for that, these ramps are generally situated together, and the

features for different radius can describe these different classes. As a result, the LSTM mechanism may not consolidate a feature vector to distinguish these classes. Also, because of attention approaches, descriptive features may be suppressed.

ShellNet behaves similar to PointNet++, which means that it can classify wall and inclined ramp, but it may confuse between terrain and flat ramp classes. The main difference between ShellNet and PointNet++ is that ShellNet applies maximum pooling and then 1D convolution to the shells' points, while PointNet++ directly employed maximum to all points in the local regions. In this way, ShellNet suppresses the dominant features, making it challenging to distinguish terrain and flat ramp classes.

In contrast to other architectures, except PointConv, that consider points individually while extracting features, PointCNN calculates features by applying discrete convolution to a point and its neighbors. On the other hand, PointConv employs continuous convolution, and it does not order the points canonically. In this way, it preserves the general

characteristic of the scene. However, PointCNN confuses terrain and flat ramp classes since it only considers the transformation matrix learned through the X-Conv operator.

V. CONCLUSION AND FUTURE WORKS

This study aimed to classify walls, ramps, and terrain in NIST reference test arenas. To achieve that, PointNet, PointNet++, DGCNN, PointCNN, Point2Sequence, PointConv, and Shellnet point-based deep learning architectures were implemented. The tests were conducted using ESOGU RAMPS dataset. The test results showed that DGCNN and PointConv architectures are capable of classifying all classes. Besides, all architectures successfully identified wall class. However, it was observed that each architecture produce erroneous results depending on its own characteristic. Thus, by revealing the positive and negative aspects of these architectures, it was aimed to create a reference point for researchers who will use them for the classification of planar surfaces in NIST test reference areas.

For future works, it is planned to classify walls, ramps, and terrain with data captured from an environment similar to NIST test reference areas in Eskisehir Osmangazi University Electrical and Electronics Engineering Artificial Intelligence and Robotics laboratory. At this point, the idea that is using the transfer learning method of the models trained with the ESOGU RAMPS dataset obtained from the Gazebo simulation environment and a small number of training data from the real environment comes to the fore.

REFERENCES

[1] Kitano, H. and Tadokoro, S. (2001). RoboCup Rescue: A Grand Challenge for Multiagent and Intelligent Systems. *AI Magazine.*, 22(1), 39-52.

[2] Jacoff, A., Messina, E., Weiss, B. A., Tadokoro, S., & Nakagawa, Y. (2003, October). Test arenas and performance metrics for urban search and rescue robots. *In Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)* (pp. 3396-3403).

[3] Amigoni, F., Visser, A., & Tsushima, M. (2012, June). Robocup 2012 rescue simulation league winners. *In Robot Soccer World Cup* (pp. 20-35). Springer, Berlin, Heidelberg.

[4] Sheh, R., Schwertfeger, S., & Visser, A. (2016). 16 years of robocup rescue. *KI-Künstliche Intelligenz.*, 30(3), 267-277.

[5] Robot Operating System (ROS), Open source robotics foundations (OSRF), https://www.ros.org/, (March,2021).

[6] Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE transactions on Robotics.*, 23(1), 34-46.

[7] Kohlbrecher, S., Von Stryk, O., Meyer, J., & Klingauf, U. (2011, November). A flexible and scalable SLAM system with full 3D motion estimation. In 2011 *IEEE international symposium on safety, security, and rescue robotics* (pp. 155-160).

[8] Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., & Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous robots.*, 34(3), 189-206.

[9] Labbé, M., & Michaud, F. (2011, September). Memory management for real-time appearance-based loop closure detection. *In 2011 IEEE/RSJ international conference on intelligent robots and systems* (pp. 1271-1276).

[10] Example reference test arena, Rescue Robot League https://en.wikipedia.org/wiki/Rescue_Robot_League, (March,2021).

[11] Nguyen, A., & Le, B. (2013, November). 3D point cloud segmentation: A survey. *In 2013 6th IEEE conference on robotics, automation and mechatronics (RAM)* (pp. 225-230).

[12] Grilli, E., Menna, F., & Remondino, F. (2017). A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences.*, XLII-2/W3, 339-344.

[13] Xie, Y., Tian, J., & Zhu, X. X. (2020). Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and Remote Sensing Magazine.*, 8(4), 38-59.

[14] Kaleci, B., & Turgut, K. (2019, October). Plane Segmentation of Point Cloud Data Using Split and Merge Based Method. *In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-7).

[15] Rabbani, T., Van Den Heuvel, F., & Vosselmann, G. (2006). Segmentation of point clouds using smoothness constraint. *International archives of photogrammetry, remote sensing and spatial information sciences.*, 36(5), 248-253.

[16] Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM.*, 24(6), 381-395.

[17] Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition.*, 13(2), 111-122.

[18] Melzer, T. (2007). Non-parametric segmentation of ALS point clouds using mean shift. *Journal of Applied Geodesy Jag.*, 1(3), 159-170.

[19] Golovinskiy, A., & Funkhouser, T. (2009, September). Min-cut based segmentation of point clouds. *In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (pp. 39-46).

[20] Rusu, R. B., & Cousins, S. (2011, May). 3d is here: Point cloud library (pcl). *In 2011 IEEE international conference on robotics and automation* (pp. 1-4).

[21] Eruyar, E. E., Yılmaz, M., Yılmaz, B., Akbulut, O., Turgut, K., & Kaleci, B. A Comparative Study for Indoor Planar Surface Segmentation via 3D Laser Point Cloud Data. *Black Sea Journal of Engineering and Science*., 3(4), 128-137.

[22] Deng, W., Huang, K., Chen, X., Zhou, Z., Shi, C., Guo, R., & Zhang, H. (2020). Semantic RGB-D SLAM for Rescue Robot Navigation. *IEEE Access.*, 8, 221320-221329.

[23] Zhang, J., Zhao, X., Chen, Z., & Lu, Z. (2019). A review of deep learning-based semantic segmentation for point cloud. *IEEE Access.*, 7, 179118-179133.

[24] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*.

[25] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learningon point sets for 3D classification and segmentation. InCVPR.

[26] Qi, C. R., Yi, L., Su, H. & Guibas, L. J. (2017). PointNet++: Deephierarchical feature learning on point sets in a metric space. InNeurIPS.

[27] Zhang, Z., Hua, B. S. & Yeung, S. K.(2019). ShellNet: Efficient Point Cloud Convolutional Neural Networks using Concentric Shells Statistics. International Conference on Computer Vision (ICCV).

[28] Zhao, H., Jiang, L., Fu,, C.-W. & Jia, J. (2019). PointWeb: Enhancing lo-cal neighborhood features for point cloud processing, InCVPR.

[29] Thomas, H., Qi, C. R., Deschaud,, J.-E., Marcotegui, B., Goulette, F., & Guibas, L. J. (2019). KPConv: Flexible and deformable convolutionfor point clouds. InICCV.

[30] Wu, W., Qi, Z., & Fuxin, L. (2019). PointConv: Deep convolutionalnetworks on 3D point clouds. InCVPR.

[31] Li, Y., Bu, R., Sun, M., Wu, W., Di, X. & Chen, B. (2018). PointCNN:Convolution on x-transformed points. InNeurIPS.

[32] Landrieu , L. & Simonovsky, M. (2018). Large-scale point cloud seman-tic segmentation with superpoint graphs. InCVPR.

[33] Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics.

[34] Ye, X., Li, J., Huang, H., Du, L., & Zhang, X. (2018). 3D recurrentneural networks with context fusion for point cloud semanticsegmentation. InECCV.

[35] Liu, X., Han, Z., Liu , Y. S., & Zwicker, M. (2019). Point2Sequence:Learning the shape representation of 3D point clouds with anattention-based sequence to sequence network. InAAAI.

[36] Huang, Q., Wang, W., & Neumann, U. (2018). Recurrent slice networksfor 3D segmentation of point clouds. InCVPR.

[37] Turgut, K., & Kaleci, B. (2019). A PointNet Application for Semantic Classification of Ramps in Search and Rescue Arenas. *International Journal of Intelligent Systems and Applications in Engineering.*, 7(3), 159-165.

[38] Turgut, K., & Kaleci, B. (2020, October). Comparison of Deep Learning Techniques for Semantic Classification of Ramps in Search and Rescue Arenas. *In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-6).

[39] The ESOGU RAMPS dataset, ESOGU, https://ai-robotlab.ogu.edu.tr/Sayfa/Index/11, (March,2021).

[40] Gazebo, Open source robotics foundations (OSRF), https://gazebosim.org/, (March,2021).

[41] Tensor Flow Library, https://www.tensorflow.org/, (March,2021).

İnsansız Sualtı Aracı Hareketinin Kalman Filtre ile Kestirimi ve Makine Öğrenmesi ile İyileştirilmesi

Estimation of Unmanned Underwater Vehicle Motion with Kalman Filter and Improvement by Machine Learning

Berna EROL[1] , **Recep Fatih CANTEKİN**[1] , **Seda KARADENİZ KARTAL**[1] ,
Rıfat HACIOĞLU[1] , **Kurtuluş Sedar GÖRMÜŞ**[2] , **Şenol Hakan KUTOĞLU**[2] ,
Mehmet Kemal LEBLEBİCİOĞLU[3]

[1]Zonguldak Bülent Ecevit Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, 67100, Zonguldak, Türkiye
[2]Zonguldak Bülent Ecevit Üniversitesi, Geomatik Mühendisliği Bölümü, 67100, Zonguldak, Türkiye
[3]Ortadoğu Teknik Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, 06800, Ankara, Türkiye

Öz

Bu çalışmada bir insansız sualtı aracının altı serbestlik dereceli doğrusal olmayan matematiksel modeli elde edilmiştir. Aracın matematiksel model cevabından aracın konum ve yönelim bilgileri elde edilmiştir. Elde edilen konum ve yönelim bilgilerine gürültü eklenerek navigasyon sensör verileri üretilmiştir. Üretilen gürültülü sensör verilerinin kestirimi için kokusuz ve genişletilmiş Kalman filtre algoritmaları kullanılmıştır. Kokusuz Kalman filtresinde, sistem modeli için insansız sualtı aracının doğrusal olmayan modeli kullanılmıştır. Genişletilmiş Kalman filtresinde ise sualtı aracının doğrusal olmayan modeli belirli denge noktalarında doğrusallaştırılmıştır. Kokusuz ve genişletilmiş Kalman filtresi kestirim sonuçları karşılaştırılmıştır. Kokusuz Kalman filtre ve genişletilmiş Kalman filtre kestirimlerine makine öğrenmesi olan Destek Vektör Makinesi algoritması uygulanarak, gürültünün fazla olduğu durumlar için, kestirimler iyileştirilmiştir. Buna ek olarak, aracın verilen bir kare yolu takip ettiği hareketi için kokusuz Kalman filtre ve genişletilmiş Kalman filtre kestirimleri iyileştirilmiştir. Tüm çalışma MATLAB/Simulink ortamında yapılmıştır.
Anahtar kelimeler: Genişletilmiş Kalman filtre, kokusuz Kalman filtre, destek vektör makinesi, insansız sualtı aracı, matematiksel modelleme

Abstract

In this study, the nonlinear mathematical model of unmanned underwater vehicle is obtained in 6 degrees of freedom. Position and orientation data of the vehicle are obtained from the mathematical model response of the vehicle. The navigation sensor data are generated by adding noise to the obtained position and orientation information. Extended Kalman filter and unscented Kalman filter algorithms are used to estimate noisy sensor data.. For the extended Kalman filter, nonlinear model is linearized around the equilibrium points. For the unscented Kalman filter, nonlinear system model is used. The estimation performance of extended Kalman filter and unscented Kalman filter are compared. Estimates data comes from Extended and Unscented Kalman filter are improved by applying support vector machine (SVM) which is machine learning for situations with high noise. In addition, unscented Kalman filter and extended Kalman filter estimates are improved for the given square path. All this study is modeled in MATLAB/Simulink environment.
Keywords: Extented Kalman filter, unscented Kalman filter, support vector machine, unmanned underwater vehicle, mathematical modelling

I. GİRİŞ

Uzun yıllardır insansız araçların takibinde doğru konum belirleme için Kalman filtresi kullanılmaktadır. Kalman filtresi konum belirleme algoritmalarının temelini oluşturmaktadır [1]. İnsansız araçların gerçek pozisyon bilgisi, nesne konumu ve hareket kontrolü için çok önemlidir [2]. Günümüzde konum takibi için en iyi yöntem olarak küresel navigasyon uydu sistemleri kullanılmaktadır [3,4]. Ancak konum belirlemede sensörler ile alınan ölçümlerde gürültü ya da dış bozucu etki olması sebebi ile ölçümlerin doğruluğu düşmektedir [5,6]. Kalman filtresi ile ölçümler kestirilerek daha doğru konumlandırma bilgisine ulaşılmaktadır [7]. Geleneksel Kalman filtresi (KF), sistem modeli tam olarak biliniyor ve sistem doğrusal bir sistem ise uygulanabilmektedir [8]. Doğrusal olmayan sistemlerde kestirim performansı iyi değildir [4]. Doğrusal olmayan sistemler doğrusallaştırılarak konum bilgisi kestirimi için genişletilmiş Kalman filtresi kullanılmıştır [9,10]. Genişletilmiş Kalman filtresi (EKF) doğrusal olmayan sistemler için, geçiş matrisini ve ölçüm matrisini Taylor seri açılımlarını kullanarak doğrusallaştırır [5]. EKF gerçek konum bilgisine yaklaşmaktadır. Fakat tam olarak gerçek değeri kestirememektedir. EKF'nin bu dezavantajının sebebi, doğrusal olmayan sistemlerin

doğrusallaştırılmasında yüksek dereceli terimlerin ihmal edilerek hataya sebep olmasıdır [11]. EKF'de doğrusallaştırma yaparken matrisini hesaplar [9]. Ancak yüksek dereceli sistemlerde Jacobian matrislerin elde etmek zordur. Bu nedenle EKF' de en iyi kestirimi veremez. Doğrusal olmayan sistemlerde en iyi konum kestirimi için doğrusal olmayan sistemlere doğrusallaştırma yapılmadan uygulanabilecek kokusuz Kalman filtre (UKF) algoritması geliştirilmiştir [10].

UKF, EKF' ye nazaran rastgele bir dağılımı tahmin etmekten ise bilinen bir olasılık dağılımını tahmin etmenin daha kolay olabileceğini savunur [2]. UKF, doğrusallaştırma yapmak yerine Kokusuz Dönüşüm tekniğini kullanarak sigma noktalarını hesaplar. Buna rağmen EKF ile aynı işlem yoğunluğuna sahiptir [12]. Dış bozucu etkisinin fazla olduğu ortamlarda UKF performansı azalabilmektedir. UKF performansını iyileştirmek için birçok farklı çalışma yapılmıştır [13,14]. Mevcut çalışmalardan farklı olarak KF çeşitlerinin kestirimlerini iyileştirmek için günümüzde popüler olan makine öğrenmesi algoritmaları kullanılabilir [15]. Makine öğrenme algoritmaları tahmin problemlerinde kullanılmaktadır. Makine öğrenmesi, oluşturulan veri setindeki giriş ve çıkış değerlerine göre tahmin yapabilen bilgisayar algoritmalarının genel adıdır [16]. Birçok makine öğrenmesi algoritması vardır. Bunlardan biri Destek Vektör Makinesi (SVM)'dir. SVM, Vapnik ve ark. tarafından 1995'te geliştirilmiştir [17]. SVM algoritması hem sınıflandırma hem de regresyon problemlerini çözmek için kullanılan güçlü bir makine öğrenmesi algoritmasıdır [18]. SVM şimdiye kadar pillerin kalan ömrünü tahmin etme, enflasyon tahmini, konut fiyat tahmini gibi birçok farklı alanlarda kullanılmıştır [19-22]. Genel olarak sınıflandırma problemlerinde kullanılan SVM, regresyon için kullanılması Smola ve ark. tarafından ileri sürülmüştür [23]. Bu yöntem Destek Vektör Regresyonu (SVR) olarak adlandırılmaktadır. SVR algoritması, elektrik yük tahmini [24], gemi hareketini tahmin etmek için [25], filtre kestiriminin iyileştirilmesi [15] ve günümüzde birçok alanda kullanılmaktadır [26,27].

Bu çalışmada insansız sualtı aracının navigasyon sensör verisini, simülasyon bazlı üretmek için, aracın 6 serbestlik dereceli doğrusal olmayan matematiksel modeli kullanılmıştır. Elde edilen sensör verilerine sıfır ortalamalı Gauss gürültüsü eklenmiş ve KF çeşitleri uygulanarak durumlar kestirilmiştir. Doğrusallaştırılmış araç modeline EKF, doğrusal olmayan modele ise UKF uygulanmıştır. UKF ve EKF kestirim performansları karşılaştırılmıştır. Çok gürültülü (dış bozucunun fazla olduğu ortamlarda) sensör verilerinde de filtre performansları incelenmiştir. Çok gürültülü sensör verilerinde UKF ve EKF kestirimlerinin performansı kötüleşebilmektedir. Gürültülü verilerde daha doğru konum kestirimi için UKF ve EKF kestirimlerine

makine öğrenmesi SVM algoritması uygulanmıştır. Uygulanan algoritma sonrasında EKF kestirimlerinin değişiklik gösterdiği, UKF kestirimlerinin ise iyileştiği gözlemlenmiştir.

II. MATEMATİKSEL MODELLEME VE SENSÖR VERİSİ OLUŞTURMA

Bir insansız sualtı aracının doğrusal olmayan matematiksel modeli **Eşitlik (1)** ve **Eşitlik (2)** de tanımlanmıştır [28].

$$M(\dot{v}) + C(v)v + D(v)v + g(\eta) = \tau = u \qquad (1)$$

$$\dot{\eta} = J(\eta)v \qquad (2)$$

M: Aracın kütle matrisi,
C: Merkezkaç kuvveti ve Coriolis kuvveti matrisi,
D: Sönümleme matrisi,
g: Yerçekimi ve suyun kaldırma kuvveti matrisi,
τ: Girdi vektörü,
v: Aracın doğrusal (v_1) ve açısal hız (v_2)vektörü,
η: Aracın doğrusal (η_1) ve açısal konum (η_2) (yönelim) vektörü,
J: Koordinat çevirim matrisi.

İnsansız sualtı aracının matematiksel modeli için, aracın 6 serbestlik dereceli (DOF) doğrusal olmayan modeli için kullanılan vektörler **Eşitlik (3), (4)** ve **(5)**'te verilmiştir [28].

$$\vec{\eta} = [\eta_1^T, \eta_2^T]^T \qquad \eta_1^T = [x, y, z]^T \qquad \eta_2^T = [\emptyset, \theta, \psi]^T \qquad (3)$$

$$\vec{v} = [v_1^T, v_2^T]^T \qquad \vec{v_1} = [u, v, w]^T \qquad \vec{v_2} = [p, q, r]^T \qquad (4)$$

$$\vec{\tau} = u = [\tau_1^T, \tau_2^T]^T \qquad \vec{\tau_1} = [X, Y, Z]^T \qquad \vec{\tau_2} = [K, M, N]^T \qquad (5)$$

İticilerin ürettikleri moment ve kuvvetlerden oluşan u matrisi sistem girdisidir. Şekil 1'de bu çalışmada kullanılan insansız sualtı aracı verilmiştir. Şekil 1'de gösterilen araçta x ekseni yönündeki (yatay konum) hareketi ve z ekseni etrafındaki dönmeyi (sapma açısı) sağlayan 2 tane yatay konumda itici bulunmaktadır, y ekseni etrafındaki dönmeyi sağlayan 1 tane dikey konumda itici bulunmaktadır [29].

Şekil 1. Bu çalışmada kullanılan insansız sualtı aracı (Zonguldak Bülent Ecevit Üniversitesi havuz deneyinden)

Bu çalışmada, insansız sualtı aracının x ve y eksenlerindeki konum bilgisi akustik bazlı konum ölçüm sistemi olan çok kısa taban hatlı, (Ultra short base line) USBL'den (GPS entegreli), yönelim bilgisi ataletsel ölçüm sistemi, (Inertial Measurement Unit) IMU' dan alındığı, aracın z eksenindeki konumu araca entegre edilen basınç sensöründen de alındığı düşünülmüştür. Su üstü aracına USBL/GPS in GPS modülü yerleştirilerek USBL'in 4 adet alıcı antenleri, yerleri bilinen su üstü aracının 4 köşesine yerleştirilmiştir (P1, P2, P3, P4).

USBL'in alıcı vericisi insansız sualtı aracına entegre edilmiştir. İnsansız sualtı aracındaki alıcı-vericiden gönderilen akustik ses sinyali su üstü aracındaki 4 adet alıcı tarafından cevaplanarak tekrar insansız sualtı aracındaki alıcı-verici tarafından alınmıştır. Böylelikle ses sinyal hızı ve sinyalin seyahat süresinden her bir anten ile insansız sualtı aracı arasındaki mesafeler elde edilmektedir. Yine bu mesafeler aynı zamanda SLAM, eş zamanlı konum belirleme algoritmasından elde edilebilmektedir. Böylelikle d1, d2, d3, d4 mesafelerinin aynı zamanda akustik sinyalin hız ve seyahat süresinin çarpımına eşit olduğu düşünülerek insansız sualtı aracının bilinmeyen konumu P(x,y,z) elde edilir.
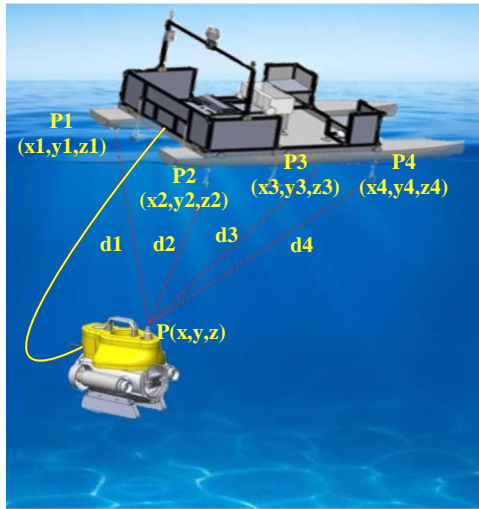
Şekil 2. İnsansız sualtı aracı navigasyonu için tasarlanan sistemin şematik gösterimi

İnsansız sualtı aracına yerleştirilen verici-alıcıdan giden sinyal su üstündeki alıcılar tarafından cevaplanarak tekrar sualtı aracına ulaşmaktadır. Bu geçen süreden ve ses hızından faydalanılarak iki araç arasındaki mesafe ve yönelim ölçülür. Aracın konum bilgisinin elde edilmesi için en az 3 tane su üstü alıcısına ihtiyaç vardır. Bu ölçülen bilgilerden faydalanılarak su altındaki aracın konumu hesap edilir. Bu çalışmada aracın x eksenindeki konumu ile z ekseni etrafındaki dönüş açısını oluşturmak için insansız sualtı aracının matematiksel modeli kullanılmıştır. Modelden elde edilen gerçek konum ve

açı bilgilerine sıfır ortalamalı gauss gürültüsü eklenerek sensör verileri üretilmiştir.

III. SENSÖR VERİLERİNİN KESTİRİMİ
Bu çalışmada insansız sualtı aracının model cevabından üretilen gürültülü sensör verilerini kestirmek için UKF ve EKF algoritmaları kullanılmıştır. İnsansız sualtı aracının doğrusal olmayan modeli için UKF ve doğrusallaştırılmış modeli için EKF algoritmaları kullanılmıştır. Gürültünün fazla olduğu durumda, makine öğrenme algoritması, KF algoritmalarına entegre edilerek iyileştirilmiştir.

3.1. Genişletilmiş Kalman Filtresi
Geleneksel KF doğrusal olmayan sistemlere uygulanamamaktadır. Doğrusal olmayan sistemlere KF uygulayabilmek için doğrusallaştırma yapmak gerekir [10]. KF'nin doğrusal olmayan sistemlere uygulanabildiği ilerletilmiş formuna "Genişletilmiş Kalman Filtresi (EKF)" denir. Doğrusallaştırma yapmanın birçok metodu vardır [30]. Bunlardan bir tanesi diferansiyelleme metodudur. Bu çalışmada, EKF için sistem modeli olarak, insansız sualtı aracının doğrusal olmayan modeli belirli denge noktalarında doğrusallaştırılarak kullanılmıştır.

3.1.1. *İnsansız sualtı aracının doğrusal hareket denklemleri*
İnsansız sualtı aracının doğrusal hareket denklemleri, **Eşitlik (1)** ve **Eşitlik (2)** de belirtilen doğrusal olmayan hareket denklemlerinin zamana göre değişen bir referans yörünge veya denge noktası etrafında doğrusallaştırılmasıyla elde edilir. Burada referans noktaları η_0 (konum için) ve v_0 (hız için) olarak alınırsa **Eşitlik (1)** diferansiyellenerek, aşağıdaki gibi doğrusallaştırılır [28]. Aynı zamanda doğrusallaştırma yapılabilmesi için aracın sabit hızda bir çalışma noktası seçilip, bu noktada yerçekimi kuvveti ve suyun kaldırma kuvveti terimleri ihmal edilir.

$\Delta_v = v - v_0 = x_1$ (referans noktasından sapmayı gösterir)   (6)

$\Delta_\eta = \eta - \eta_0 = x_2$ (referans noktasından sapmayı gösterir)   (7)

$$M\Delta_{\dot{v}} + \frac{\partial C(v)v}{\partial v}\Big|_{v_0} \Delta_v + \frac{\partial D(v)v}{\partial v}\Big|_{v_0} \Delta_v + \frac{\partial g(\eta)}{\partial \eta}\Big|_{\eta_0} \Delta_\eta = \Delta_\tau \quad (8)$$

Zamana bağımlı doğrusallaştırılmış dinamik hareket denklemi **Eşitlik (9)** da verildiği gibidir.

$$M\dot{x}_1 + C(t)\,x_1 + D(t)\,x_1 + G(t)\,x_2 = \tau \quad (9)$$

Burada; $C(t) = \frac{\partial C(v)v}{\partial v}\Big/_{v_{0(t)}}$, $D(t) = \frac{\partial D(v)v}{\partial v}\Big/_{v_{0(t)}}$ ve $G(t) = \frac{\partial g(\eta)}{\partial \eta}\Big/_{\eta_{0(t)}}$ şeklindedir.

Eşitlik (2) aşağıdaki gibi doğrusallaştırılır:

$$\Delta_{\dot\eta} = J(\eta_0)\ \Delta_v + J^*(v_0,\eta_0)\ \Delta_\eta \tag{10}$$

$$J^*(v_0,\eta_0)\ \Delta_\eta = [\ J(\eta_0 + \Delta_\eta) - J(\eta_0)\ ]\ v_0 \tag{11}$$

Zamana bağımlı doğrusallaştırılmış kinematik hareket denklemi aşağıdaki gibidir:

$$\dot X_2 = J(t) x_1 + J^*(t)\ x_2 \tag{12}$$

Burada; $J(t) = J(\eta_0(t))$ ve $J^*(t) = J^*(v_0(t),\ \eta_0(t))$ şeklindedir. $J^* = 0$ alınarak zamandan bağımsız model **Eşitlik (13)** da, $\dot x = Ax + Bu$ formunda olur.

$$\begin{vmatrix} \dot x_1 \\ \dot x_2 \end{vmatrix} = \begin{vmatrix} -M^{-1}[C+D] & -M^{-1}G \\ J & 0 \end{vmatrix} \cdot \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} + \begin{vmatrix} -M^{-1} \\ 0 \end{vmatrix} \cdot u \tag{13}$$

EKF tasarımında Jacobian matrisleri hesaplanır. Bu adım, ölçüm matrisi (H) ve geçiş matrisinin (F) diferansiyellenmesiyle doğrusallaştırılmasıdır [30].

$$H_k = \frac{\partial h(x_k,k)}{\partial x}\Big|_{x=\hat x_k^-} \tag{14}$$

$$F(k+1,\ k) = \frac{\partial f(x,k)}{\partial x}\Big|_{x=\hat x_k} \tag{15}$$

EKF uygulamasında, insansız sualtı aracı hareket denklemlerinin zamandan bağımsız formları dikkate alınmıştır. Durum tahmini ayrık zaman geçiş matrisi ile **Eşitlik (16)** daki gibi ilişkilendirilir.

$$x_{k+1} = \phi_k \cdot x_k \tag{16}$$

KF'nin ayrık zaman formu **Eşitlik (17)** de sunulmuştur [1].

$$\phi_k = e^{F\Delta t} = I + F\Delta t + \frac{(F\Delta t)^2}{2} + \frac{(F\Delta t)^3}{6} + \dots \tag{17}$$

Burada:

$$\Delta t = t_{k+1} - t_k \tag{18}$$

Zaman aralığı olarak belirlenir [1]. Doğrusallaştırmada ikinci derece terimler ihmal edilmektedir [28]. Bu yüzden **Eşitlik (17)** için ikinci ve daha yüksek dereceli terimler ihmal edilir.

3.1.2. Genişletilmiş Kalman filtresi algoritması
Sistem modeli için **Eşitlik (13)** ve **Eşitlik (15)** kullanılmıştır. Durum vektörü x **Eşitlik (19)** daki gibidir.

$$x = |\eta_1 \quad \eta_2 \quad v_1 \quad v_2|^T \tag{19}$$

Q matrisi, sistem gürültü kovaryans matrisidir ve köşegen matristir [1]. Bu çalışmada 12×12'lik köşegen matris alınmıştır. İnsansız sualtı aracı sistem gürültü kaynakları; insansız sualtı aracının matematiksel modelinden kaynaklı konumda ve hızda meydana gelen rastgele hatalardır. Her bir terim sistem durumlarının standart sapmaları ile orantılıdır.

$$Q = \begin{vmatrix} Q_{11} & 0 & \dots & 0 \\ 0 & Q_{22} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & Q_{1212} \end{vmatrix} \tag{20}$$

Burada;

- Q_{11}, Q_{22} ve Q_{33} sırasıyla x, y ve z yönündeki hareketin sistem gürültüsüdür.
- Q_{44}, Q_{55} ve Q_{66} sırasıyla x, y ve z ekseni etrafındaki dönme için sistem gürültüsüdür.
- Q_{77}, Q_{88} ve Q_{99} sırasıyla x, y ve z yönündeki hareket için doğrusal hız sistem gürültüsüdür.
- Q_{1010}, Q_{1111} ve Q_{1212} sırasıyla x, y ve z ekseni etrafındaki dönme için açısal hız sistem gürültüsüdür.

Ölçüm modeli için z ölçüm vektörü **Eşitlik (21)** ve H ölçüm matrisi **Eşitlik (23)** daki gibi tanımlanmıştır.

$$z_k = H_k \cdot x_k \tag{21}$$

$$z = |konum \quad yönelim|^T \tag{22}$$

H matrisi ölçüm matrisidir ve birim matristir [1]. Bu çalışmada 6×12'lik birim matris alınmıştır.

$$H = \begin{vmatrix} I_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & I_3 & 0_3 & 0_3 \end{vmatrix} \tag{23}$$

R matrisi ölçüm gürültü kovaryansıdır ve köşegen matristir. Bu çalışmada 6x6'lık köşegen matris alınmıştır. R matrisindeki köşegen terimleri, her bir ölçümün varyanslarıdır. İnsansız sualtı aracı ölçüm gürültü kaynakları; USBL/GPS kaynaklı konumdaki rastgele hatalar ve yönelim ölçümünde meydana gelen rastgele hız hatalarıdır.

$$R = \begin{vmatrix} R_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & R_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & R_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & R_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & R_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & R_{66} \end{vmatrix} \tag{24}$$

Burada;

- R_{11}, R_{22} ve R_{33} sırasıyla x, y ve z yönündeki hareketin ölçüm gürültüsüdür.
- R_{44}, R_{55} ve R_{66} sırasıyla x, y ve z ekseni etrafındaki dönme için ölçüm gürültüsüdür.

EKF algoritması da geleneksel KF algoritma adımlarına göre işler. EKF algoritmasındaki fark, sistem modeli için yukarıda belirtilen diferansiyelleme ile doğrusallaştırma yapılması ve Jacobian matrislerinin türetilmesidir. EKF algoritması aşağıda sunulduğu gibidir. Bu algoritmada 1. ve 2. adım zaman güncelleme (sistem yayılımı) aşamasıdır. Sonraki adımlar ölçüm güncelleme aşamasıdır.

1. Durum tahmini hesapla.

$$\hat{x}^-_{k+1} = \phi \cdot \hat{x}^+_k \tag{25}$$

2. Hata kovaryansı hesapla.

$$P^-_{k+1} = \phi_k \cdot P_k \cdot \phi^T_k + Q_k \tag{26}$$

3. Kalman kazancını hesapla.

$$K_k = P^-_k \cdot H^T_k \cdot (H_k \cdot P^-_k \cdot H^T_k + R_k)^{-1} \tag{27}$$

4. Durum tahminini güncelle.

$$\hat{x}^+_{k+1} = \hat{x}^-_{k+1} + K_k \cdot [z_k - h \cdot (\hat{x}^-_{k+1}, t_k)] \tag{28}$$

5. Hata kovaryansını güncelle.

$$P^+_{k+1} = (I - K_k \cdot H_k) \cdot P^-_{k+1} \tag{29}$$

3.2. Kokusuz Kalman Filtresi

EKF, doğrusal olmayan sistemlere uygulanabilir. Ancak doğrusallaştırma yapmak iki dezavantaja neden olur [28]. Bunlardan birisi yüksek dereceli terimlerin ihmal edilmesidir. Bunun sonucunda gerçek değerden istenilmeyen derecede uzaklaşan filtreler oluşabilir. İkincisi, Jacobian matrislerinin bazı uygulamalarda türetilmesine gerek yoktur. Ancak her uygulama için Jacobian matrislerini türetmek uygulamada önemli zorluklara yol açar [8]. EKF' nin bu dezavantajlarını, UKF ortadan kaldırmıştır. UKF' nin çalışma prensibi, rastgele bir olasılık dağılımını tahmin etmektense bilinen bir olasılık dağılımını (Gaussian) tahmin etmenin daha kolay olduğudur [31]. UKF, doğrusal olmayan sistemlere doğrudan uygulanabilmektedir. UKF, kokusuz dönüşüm metodu kullanılarak uygulanır. Kokusuz dönüşüm doğrusal olamayan bir dönüşüm geçiren rastgele değişkenin ortalama ve kovaryansları hesaplamak için kullanılır [32]. Birçok kokusuz dönüşüm tekniği vardır. Bu çalışmada, karekök kovaryans tipi kullanılmıştır. UKF' deki ikinci önemli husus ise sigma noktalarıdır. Sigma noktaları UKF kestiriminde bir önceki ölçümün istatistiğinden elde edilen aracın olabileceği muhtemel konumlardır. Örnek olarak, sigma noktaları Şekil 3'de gösterilmektedir.

Sigma noktalarından oluşan S matrisi $2n+1$ boyutundadır. Burada n durum matrisinin boyutudur. Sigma noktalarının ortalamasının ağırlığı w_m ve kovaryansının ağırlığı w_c' dir. Ayrıca sigma noktalarının ortalaması durum vektör tahminini,

varyansı ise hata kovaryans matrisini ifade eder [1]. Sigma noktaları **Eşitlik (30), Eşitlik (31)** ve **Eşitlik (32)** gibi elde edilir:
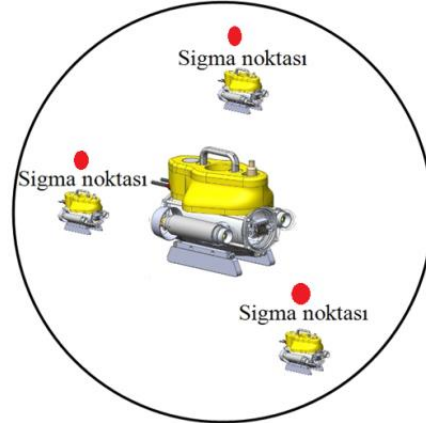
Şekil 3. Sigma noktaları

$$S_0 = \bar{x}, \ i = 0 \tag{30}$$

$$S_i = \bar{x} + \sqrt{(n + K\_sig) \cdot P_{k-1}} \,, \ i = 1, 2, \dots, n \tag{31}$$

$$S_i = \bar{x} - \sqrt{(n + K\_sig) \cdot P_{k-1}} \,, \ i = n+1, \dots, 2n \tag{32}$$

Kokusuz dönüşümde, ölçekleme parametreleri bulunur ve bu ölçekleme paremetlerinin modele uygun bir şekilde belirlenmesi filtre performansını artırır. K_sig, ölçekleme parametresi, sigma noktasının orijine konumlandırılmış x rastgele değişkeninden ne kadar uzaklaşabileceğini kontrol eder [2].

$$K\_sig = \alpha^2 \cdot (n + \lambda) - n \tag{33}$$

Burada, α, birincil ölçekleme parametresidir. Sigma noktalarının $\bar{x}$ etrafındaki yayılımını belirler. Genellikle $10^{-4} \leq \alpha \leq 1$ şeklinde küçük ve pozitif bir sayı seçilmelidir. Çünkü istenilen durum, $\bar{x}$ değerinden uzaklaşmamaktadır. λ, üçüncül ölçeklendirme parametresidir. Genellikle 0 veya 3-n şeklinde belirlenir [32].

$$w_m = w_c = \frac{1}{2 \cdot (n + K\_sig)} \tag{34}$$

$$w^c_0 = \frac{K\_sig}{(K\_sig + n)} + (1 - \alpha^2 + \beta) \tag{35}$$

$$w^m_0 = \frac{K\_sig}{(K\_sig + n)} \tag{36}$$

Burada, β, İkincil ölçeklendirme parametresidir ve x'in önceki dağılımlarının bilgisini dahil etmek için kullanılır ve Gauss dağılımları için en uygun 2 alınır [32]. UKF'deki ilk adım hata kovaryans matrisinin (P) karekökünün alınmasıdır. Ardından sigma noktaları, hata kovaryansının karekökü ve ölçeklendirme parametreleri doğrultusunda hesaplanır. Daha sonra

durum kestirimi hesaplanmasında kullanılarak algoritma işleyişi KF ve EKF'deki gibi devam eder.

3.2.1. Kokusuz Kalman filtresi algoritması

UKF tasarımında, insansız sualtı aracının **Eşitlik (1)** ve **Eşitlik (2)** deki doğrusal olmayan hareket denklemlerinden yola çıkılarak aşağıdaki doğrusal olmayan durum uzay modeli ede edilir. Durum seçimi ve ölçüm modeli EKF'deki gibidir.

$$\begin{vmatrix} \dot{x}_1 \\ \dot{x}_2 \end{vmatrix} = \begin{vmatrix} -M^{-1}[C(v)+D(v)] & 0 \\ J(\eta) & 0 \end{vmatrix} . \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} + \begin{vmatrix} -M^{-1} \\ 0 \end{vmatrix} . u + \begin{vmatrix} -g(\eta) \\ 0 \end{vmatrix} \quad (37)$$

Doğrusal olmayan durum uzayı gösterimi aşağıdaki gibi olur:

$$\dot{x} = A(x)x + B(x)u + f(x) \quad (38)$$

Burada; $f = \begin{vmatrix} 0 \\ -g(\eta) \end{vmatrix}$ olur. UKF algoritması aşağıda sunulduğu gibidir. Bu algoritmada 3. 4. ve 5. adım zaman güncelleme (sistem yayılımı) aşamasıdır. Sonraki adımlar ölçüm güncelleme aşamasıdır.

1. Başlangıç değerlerini belirle.

$$\hat{x}_0 = E[x_0] \quad (39)$$

$$P_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \quad (40)$$

2. Sigma noktalarını belirle.

$$S_{k-1} = [\hat{x}_{k-1}^-, \ \hat{x}_{k-1}^- + \sqrt{(n + K\_sig) . P_{k-1}}, \hat{x}_{k-1}^- - \sqrt{(n + K\_sig). P_{k-1}}] \quad (41)$$

3. Durum tahmini hesapla.

$$\hat{x}_{k+1}^* = \phi_k . S_{k-1} \quad (42)$$

$$\hat{x}_k^- = \sum_{i=0}^{2n} w_i^{(m)} . \hat{x}_{i,k+1}^* \quad (43)$$

4. Hata kovaryansını hesapla.

$$P_k^- = \sum_{i=0}^{2n} w_i^{(c)} . \left(\hat{x}_{i,k+1}^* - \hat{x}_k^-\right)\left(\hat{x}_{i,k+1}^* - \hat{x}_k^-\right)^T + Q \quad (44)$$

5. Ölçümü formüle et.

$$\hat{y}_{k-1} = H . (\hat{x}_{k+1}^*) \quad (45)$$

$$\hat{y}_k = \sum_{i=0}^{2n} w_i^{(m)} . \hat{y}_{i,k-1} \quad (46)$$

6. Kalman kazancını hesapla.

$$P_{yy} = \sum_{i=0}^{2n} w_i^{(c)} . \left(\hat{y}_{i,k-1} - \hat{y}_k\right)\left(\hat{y}_{i,k-1} - \hat{y}_k\right)^T + R \quad (47)$$

$$P_{xy} = \sum_{i=0}^{2n} w_i^{(c)} . \left(\hat{x}_{i,k+1}^* - \hat{x}_k^-\right)\left(\hat{y}_{i,k-1} - \hat{y}_k\right)^T \quad (48)$$

$$K_k = P_{xy} . P_{yy}^{-1} \quad (49)$$

7. Durum tahminini güncelle.

$$\hat{x}_k^+ = \hat{x}_k^- + K_k . (z - \hat{y}_k) \quad (50)$$

8. Hata kovaryansını güncelle.

$$P_k^+ = P_k^- - (K_k . P_{yy} . K_k^T) \quad (51)$$

3.3. Destek Vektör Makinesi

Destek vektör makinesi (SVM), tahmin ve sınıflandırma problemlerinde kullanılan popüler bir makine öğrenmesi algoritmasıdır [33]. Sınıflandırma problemlerinde daha sık kullanılan SVM, regresyon problemlerinde kullanılması Smola ve ark. tarafından önerilmiştir [23]. SVM, Şekil 4'teki gibi giriş-çıkış ilişkisi doğrusal ve doğrusal olmayan problemlerde kullanılabilir.
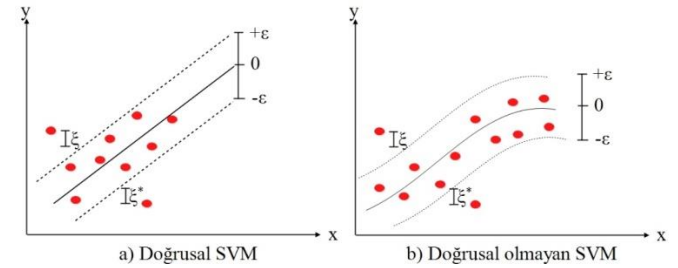
Şekil 4. Tek boyutlu SVM

$D = \{(x1, y1), (x2, y2), ..., (xi, yi)\}$ den oluşan bir eğitim seti olsun. Burada xi giriş değişkenleri, yi ise bu girişlere denk gelen çıkış değişkenleridir.

SVM'de amaç verilen giriş değişkenleri ile çıkış değişkenleri arasındaki bağlantıyı kurmaktır. Yani SVM algoritması xi'ler ile yi'lerin ilişkisini en doğru şekilde oluşturulacak fonksiyonu bulmaktadır. SVM bu bağlantıyı kurarken **Eşitlik (52)** da verilen J maliyet fonksiyonunu minimize eden optimizasyon problemini çözmektedir.

Maliyet fonksiyonu (J):

$$J = \frac{1}{2} \|w\|^2 + C . \sum_{i=1}^{m} (\xi_i + \xi_i^*) \quad (52)$$

Kısıtlar:

$$y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \qquad (53)$$

$$(w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \qquad (54)$$

$$\xi_i, \xi_i^* \geq 0, \quad i=1,....,m \qquad (55)$$

Burada, *C* parametresi sıfırdan büyük bir sayı olup Şekil 4'deki kesikli çizgilerin aralığının boyutunu belirlemek için kullanılmaktadır. *C* parametresinin belirlenmesi veri setine göre değişiklik göstermektedir. Bu yüzden deneme yanılma yoluyla tespit edilmektedir. ξ_i ve ξ_i^* ise aykırı gözlemlerin kesikli çizgiye olan mesafelerini belirtmektedir. *ε* sıfırdan büyük bir sayı olup kullanıcı tarafından belirlenmektedir. **Eşitlik (53)**' deki *b* katsayısı oluşturulan fonksiyonda algoritmanın belirlediği bir sabit sayıdır. *w* ise verilen *xi* giriş değişkenleri ile çarpılacak sayı olup algoritma tarafından belirlenmektedir. Şekil 5'te görüldüğü gibi UKF kestirimlerini iyileştirmek için SVM algoritmasının eğitimi, UKF kestiriminden elde edilen veriler ile yapılmıştır. SVM eğitiminde kullanılan veri seti 3 adet sigma noktası, UKF kestirimi ve insansız sualtı aracının matematiksel modelinden alınan gerçek değer olmak üzeri her bir örnekleme aralığı için 5 farklı veriden oluşmaktadır.
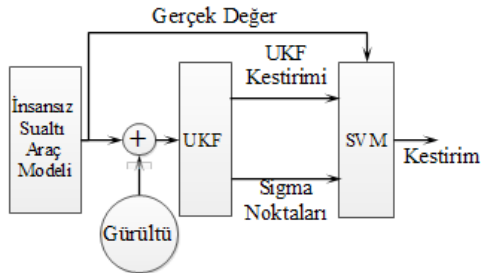
Şekil 5. SVM/UKF blok şeması

Şekil 6'da gösterildiği gibi EKF kestirimlerini iyileştirmek için ise SVM algoritmasının eğitimi, EKF kestiriminden elde edilen veriler ile yapılmıştır. SVM eğitiminde kullanılan veri seti, EKF kestirimi ve insansız sualtı aracının matematiksel modelinden alınan gerçek değer olmak üzeri her bir örnekleme aralığı için 2 farklı veriden oluşmaktadır.
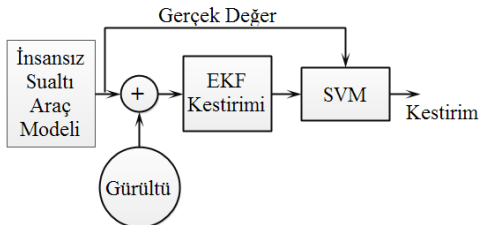
Şekil 6. SVM/EKF blok şeması

IV. SİMÜLASYON SONUÇLARI

Bu çalışmada kullanılan insansız sualtı aracının sağ ve sol iticilerine eşit kuvvet uygulayarak ileri yön hareketi analizi, sağ ve sol iticilere farklı kuvvet uygulayarak ise sapma açısı (yaw) analiz edilmiştir. Buna ek olarak, insansız sualtı aracının belirtilen kare bir yolu takip etmesi sağlanmıştır ve bu kare yol takip hareketi için, oluşturulan filtrelerin başarımları karşılaştırılmıştır.

Bu çalışmada insansız sualtı aracının farklı gürültülerdeki konum ve yönelim hareketi incelenmiştir. Şekil 7 ve Şekil 9'da farklı gürültülerde aracın ileri yön hareketi için, matematiksel model cevabından elde edilen x eksenindeki gerçek konum bilgisi ile EKF (sarı çizgi) ve UKF (mavi çizgi) ile SVM (kırmızı ve yeşil çizgi) algoritmalardan elde edilen konum kestirimi gösterilmiştir. Şekil 8 ve Şekil 10'da x konum kestirimindeki elde edilen mutlak hata grafikleri UKF (mavi çizgi), EKF (sarı çizgi), SVM/UKF (kırmızı çizgi) ve SVM/EKF (yeşil çizgi) için verilmiştir. Mutlak hata kestirilen veri ile aracın model cevabından elde edilen gerçek veri arasındaki fark olarak tanımlanmıştır. Şekil 8 ve 10' da görüldüğü gibi az gürültülü ve çok gürültülü durumda en düşük hata SVM/UKF algoritması ile elde edilmiştir.

Şekil 11 ve Şekil 13'de farklı gürültülerde aracın sapma yön hareketi için, matematiksel model cevabından elde edilen sapma açısı (siyah çizgi) ile EKF (sarı çizgi) ve UKF (mavi çizgi) ile SVM algoritmalardan (kırmızı, yeşil çizgi) elde edilen sapma açısı kestirimleri sunulmuştur. Şekil 12 ve Şekil 14'te sapma açısı kestiriminde elde edilen mutlak hata grafiği verilmiştir. Sapma yönelim hareketi için, az gürültülü ve çok gürültülü durumlarda SVM/UKF algoritmanın en düşük hataya sahip olduğu Şekil 12 ve 14 'te görülmektedir.
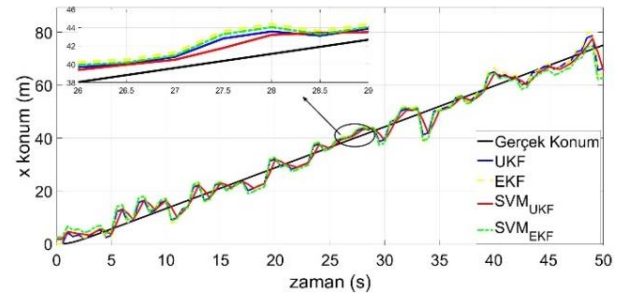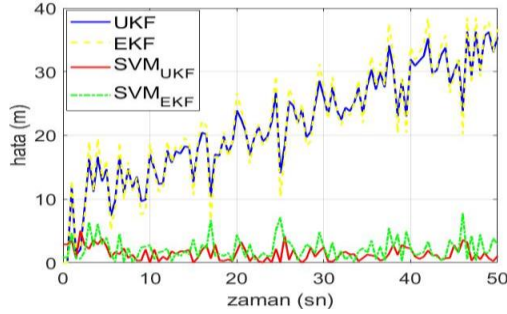

Şekil 7. Zamana bağlı gerçek x konumu ve kestirimleri (az gürültülü durum)
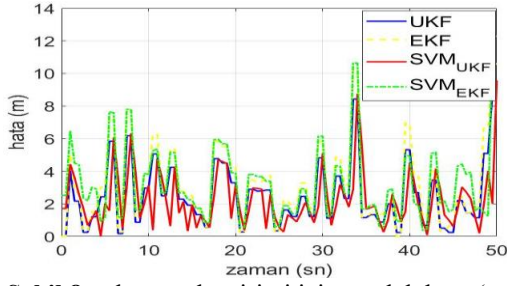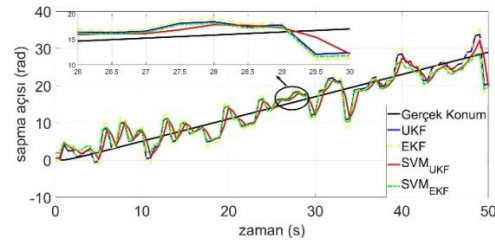
Şekil 8. x konum kestirimi için mutlak hata (az gürültülü durum)
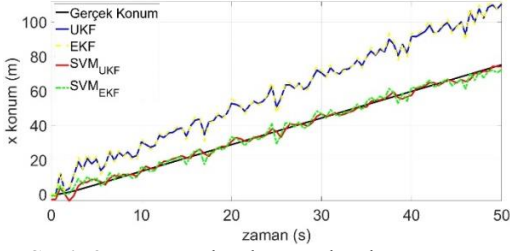
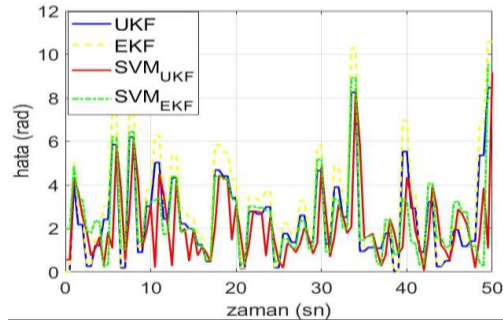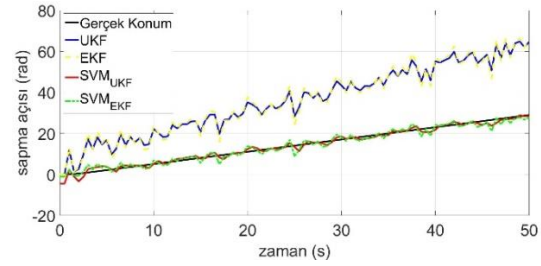**Şekil 9.** Zamana bağlı gerçek x konumu ve kestirimleri   (çok gürültülü durum)

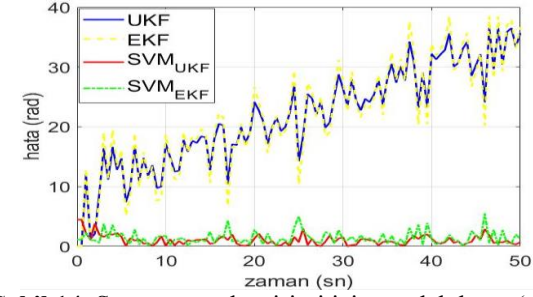**Şekil 10.** x konum kestirimi için mutlak hata   (çok gürültülü durum)

**Şekil 11.** Zamana bağlı sapma açısı ve kestirimleri (az gürültülü durum)

Şekil 12. Sapma açısı kestirimi için mutlak hata (az gürültülü durum)

Şekil 13. Zamana bağlı gerçek sapma açısı ve kestirimleri
(çok gürültülü durum)

Şekil 14. Sapma açısı kestirimi için mutlak hata (çok gürültülü durum)

Az gürültülü ve çok gürültülü ortamlarda her bir algoritmanın mutlak hatası analiz edilmiştir. Mutlak hata her bir algoritma tarafından kestirilen veri ile aracın model cevabından üretilen gerçek veri arasındaki farktır. Tablo 1 de, az gürültülü ve çok gürültülü durumlarda EKF, UKF, SVM/UKF ve SVM/EKF algoritmaları ile aracın kestirilen x konumu ve gerçek x konumu arasındaki mutlak hata verilmiştir. Buna ek olarak, EKF, UKF, SVM/UKF ve SVM/EKF algoritmaları ile aracın kestirilen sapma açısı ve gerçek sapma açısı arasındaki fark (mutlak hata) verilmiştir. Tablo 1'dende görüldüğü gibi hem az gürültülü hem de çok gürültülü durumda en doğru sonuç SVM algoritmasının UKF ye entegre edildiği durumdur.

Tablo 1. Filtre kestirimler için mutlak hata değerleri

ALGORİTMA	MUTLAK HATA		
	x-Konum	**y-Konum**	**Sapma Açısı**
SVM$_{UKF}$	1,529 m	2,0977 m	0,1254 rad
SVM$_{EKF}$	2,6819 m	2,3262 m	0,1509 rad

Bu çalışmada SVM ile iyileştirilmiş filtrelerin başarımları x ve y eksenlerinde $75m \times 75m$ olarak verilen kare yol hareketi için karşılaştırılmıştır. Aracın verilen kare yolu takip edebilmesi kontrolcüler ile uygun girdi bilgileri üretilerek sağlanmıştır. Verilen kare yol için iyileştirilmiş filtre kestirimleri Şekil 15'te verilmiştir. Şekil 16 ve 17'de sırasıyla x eksninde alınan yol ile y ekseninde alınan yola ait mutlak hatalar verilmiştir. Sapma açısı için iyileştirilmiş filtre kestirimleri Şekil 18 verilmiştir. Sapma açısına ait mutlak hatalar SVM/UKF ve SVM/EKF için Şekil 19'da verilmiştir.
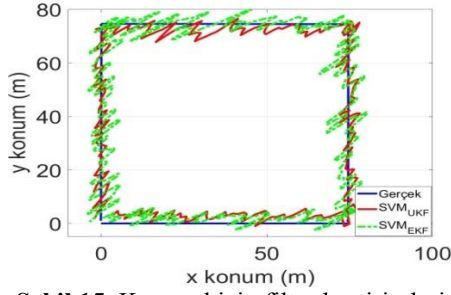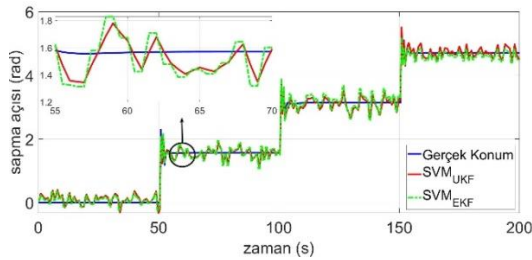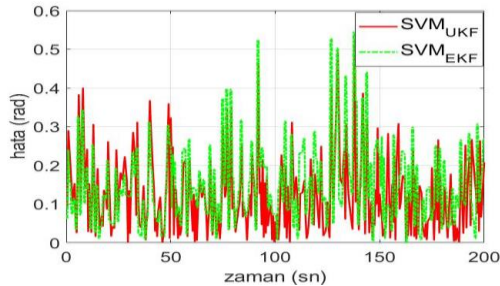
Şekil 15. Kare yol için filtre kestirimleri
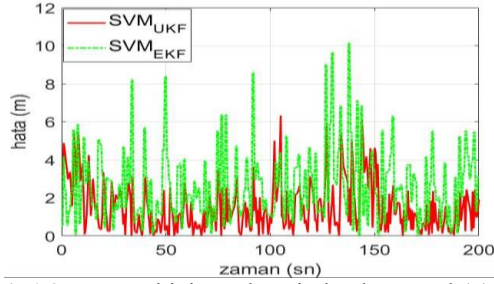
Şekil 16. Kare yol için x ekseninde alınan yol (x) için mutlak hata
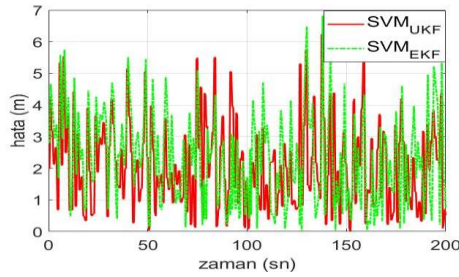
Şekil 17. Kare yol için y ekseninde alınan yol (y) için mutlak hata

Şekil 18. Kare yol için sapma açısı kestirimleri

Şekil 19. Kare yol için sapma açısının mutlak hatası

Tablo 2 de aracın takip ettiği kare bir yola ait SVM/UKF ve SVM/EKF algoritmalarına kestirilen konum (x, y) ve sapma açısına ait mutlak hatalar verilmiştir. Mutlak hata kestirilen veri ile aracın model cevabından elde edilen gerçek veri arasındaki farktır. Tablo 2'den de görüldüğü gibi SVM/UKF'nin doğruluğu SVM/EKF'ye göre daha büyüktür.

Tablo 2. Kare yol kestirimi için mutlak hata değerleri

ALGORİTMA	MUTLAK HATA		
	konum (x)	konum (y)	sapma açısı
SVM$_{UKF}$	1,529 m	2,0977 m	0,1254 rad
SVM$_{EKF}$	2,6819 m	2,3262 m	0,1509

V. SONUÇ

Matematiksel modeli elde edilen insansız sualtı aracının gerçek konum ve yönelim bilgilerine gürültü eklenerek navigasyon sensör verisi üretilmiştir. Bu sensör verilerinden konum ve sapma açısı kestirimi için UKF, EKF algoritmaları uygulanmıştır. Sonuçlardan UKF'nin kestirim performansı daha iyi olduğu görülmektedir. Çok gürültülü ortamlardaki UKF ve EKF kestirimlerini iyileştirmek için farklı makine öğrenmesi algoritmaları test edilmiştir. Test edilen algoritmalardan en iyi sonucu veren SVM algoritması UKF ve EKF kestirimlerine uygulanmıştır. İyileştirilmiş filtre kestirimlerinden başarımı en yüksek olan SVM$_{UKF}$ (SVM algoritması ile iyileştirilmiş UKF) algoritması olduğu gözlemlenmiştir. Bunun nedeni, SVM algoritmasının veri setini oluşturmak için veri çeşitliliği UKF algoritmasında mevcuttur. SVM veri setini UKF algoritması ile oluştururken UKF kestirimi ve sigma noktaları kullanılmıştır. Ancak SVM veri setini EKF ile oluştururken sadece EKF filtre kestirimi kullanılabilir. Bu nedenle çok gürültülü ortamlarda bile SVM$_{UKF}$ algoritması gerçeğe yakın konum ve yönelim kestirimi yapabilmektedir. İyileştirilmiş filtre başarımlarını daha iyi karşılaştırmak için insansız sualtı aracının verilen kare bir yolu takip etmesi sağlanmıştır. Verilen bu kare yolda da SVM$_{UKF}$ algoritmasının gerçeğe daha yakın olduğu görülmektedir. Tüm çalışma MATLAB/Simulink ortamında gerçekleştirilmiştir.

Gelecek çalışmada, sualtı aracı ile denizde gerçek navigasyon deneyleri yapılarak alınan ölçümlere bu çalışmadaki algoritmalar uygulanarak karşılaştırılacaktır.

KAYNAKLAR

[1] Groves, P. D. (2013). Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems., Artech House.

[2] Liu , K.-z., Li , J., Guo, W., Zhu , P.-q., ve Wang , X.-h.(2014) Navigation system of a class of underwater vehicle based on adaptive unscented Kalman fiter algorithm, Springer, 550-557.

[3] Duymaz, E., Oğuz, A. E., ve Temeltaş, H. (2017). Eş zamanlı konum belirleme ve haritalama probleminde yeni bir durum tahmin yöntemi olarak parçacık akış filtresi. DergiPark, 1255-1270.

[4] Jwo, D. J., Hu, C. W., ve Tseng, C. H. (2013). Nonlinear Filtering with IMM Algorithm for Ultra-Tight GPS/INS Integration. SAGE, 1-16.

[5] St-Pierre, M., Gingras, D. (2004). Comparison between the unscented Kalman filter and the extended Kalman filter for the position estimation module of an integrated navigation information system, Proc. of IEEE Intelligent Vehicles Symposium, 831-835.

[6] Makavita, C. D., Jayasinghe, G. S., Nguyen, D. H., ve Ranmuthugala, D. (2019). Experimental Study of Command Governor Adaptive Control for Unmanned Underwater Vehicles. IEEE, 332 – 345.

[7] Daum, F. (2005). Nonlinear filters: beyond the Kalman filter. IEEE, 57 – 69.

[8] Wan, E. A., Merwe, R. V. (2000). The unscented Kalman filter for nonlinear estimation. Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, Lake Louise: IEEE, 153-158.

[9] Kandepu R., Foss, B., Imsland, L. (2008). Applying the unscented Kalman filter for nonlinear state estimation, J. Process Control, 753-768.

[10] Xiong, K., Zhang, H. Y. ve Chan, C. W. (2006). Performance evaluation of UKF-based nonlinear filtering. ELSEVIER, 261-270.

[11] Dini, D. H., Mandic, D. P. ve Julier, S. J. (2011). A Widely Linear Complex Unscented Kalman Filter. IEEE, 623 - 626.

[12] Menegaz, H. M. T., Ishihara, J. Y., Borges, G. A. ve Vargas, A. N. (2015). A Systematization of the unscented Kalman filter theory. IEEE Trans. Automat. Control, 60 (10), 2583-2598.

[13] Holmes, S. A., Klein, G. ve Murray, D. W. (2008) An O(N2) Square Root Unscented Kalman Filter for Visual Simultaneous Localization and Mapping[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 31 (7), 1251-1263.

[14] Huang, M., Li, W., Yan, W. (2010). Estimating parameters of synchronous generators using square-root unscented Kalman filter Electr. Power Syst. Res., 80 (9), 1137-1144

[15] Erol, B., Cantekin, R.F., Kartal, S.K., Hacıoğlu, R., Görmüş, K.S., Kutoğlu, Ş.H. ve Leblebicioğlu, M.K. (2020). Improvement of filter estimates based on data from unmanned underwater vehicle with machine learning. Innovations in Intelligent Systems and Applications Conference (ASYU). Istanbul, Turkey. 15-17 October. IEEE.

[16] Solomatine, D. P., Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. Water Resources Research, 45(12), 1–16.

[17] Cortes, C., Vapnik, V. (1995). Support-vector networks Machine Learning, 273-297

[18] Zhang, Z., Ding, S., Sun, Y., (2021), MBSVR: Multiple birth support vector regression. Information Sciences, 552, 65-79.

[19] Zhao, Q., Qin, X., Zhao, H., ve Feng, W. (2018). A novel prediction method based on the support vector regression for the remaining useful life of lithium-ion batteries. Microelectronics Reliability, 85, 99-108.

[20] Li, X., Shu, X., Shen, J., Xiao, R., Yan, W., ve Chen, Z. (2017). An On-Board Remaining Useful Life Estimation Algorithm for Lithium-Ion Batteries of Electric Vehicles. Energies, 10(5), 691.

[21] Oktanisa, I., Mahmudy, W. F., ve Maski, G. (2020). Inflation Rate Prediction in Indonesia using Optimized Support Vector Regression Model. Journal of Information Technology and Computer Science, 5(1), 104-114.

[22] Manasa, J., Gupta, R., ve Narahari, N. S. (2020). Machine Learning based Predicting House Prices using Regression Techniques. 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 624-630.

[23] Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

[24] Dong, Y., Zhang, Z., Hong, W.C. (2018). A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. MDPI. 1-21.

[25] Li, M.W., Geng, J., Hong, W.C., Zhang, L.D. (2019). Periodogram estimation based on LSSVR-CCPSO compensation for forecasting ship motion. Nonlinear Dynamics, 97 (4), 2579-2594.

[26] Cheng, K., Lu, Z. (2021). Active learning Bayesian support vector regression model for global approximation. Information Sciences, 544, 549-563.

[27] Zhang, Z., Ding, S., Sun, Y. (2020). A support vector regression model hybridized with chaotic krill herd algorithm and empirical

mode decomposition for regression task. Neurocomputing, 410, 185-201.

[28] Fossen, T. I. (1999) Guidance and Control of Ocean Vehicles. 1. baskı, Wiley.

[29] Kartal, S.K, Leblebicioğlu, M.K., ve Ege, E. (2016). Bir İnsansız Sualtı Gözlem Aracı (SAGA) Yer Tespitinin Deneysel Testi ve Sistem Tanılaması. Otomatik Kontrol Ulusal Toplantisi (TOK), 1-13.

[30] Crassidis, J. L., Junkins, J. L. (2011). Optimal Estimation of Dynamic Systems (Chapman & Hall/CRC Applied Mathematics & Nonlinear Science). Chapman and Hall/CRC.

[31] Merwe, R. V., Wan, E. A. (2001). The square-root unscented Kalman filter for state and parameter-estimation. 2001 IEEE International Conference on coustics, Speech, and Signal Processing, Salt Lake City: IEEE, 3461-3464.

[32] Haykın, S. (2001). Kalman Filtering and Neural Networks, 1. Baskı. Wiley-Interscience, New York, USA.

[33] Vapnik, V. (2000). The Nature of Statistical Learning Theory. 2. baskı. Springer-V.

Low-Latency SoC Design with High-Level Accelerators Specific to Sound Effects

Ses Efektlerine Özel Yüksek Seviye Hızlandırıcılarla Düşük Gecikmeli SoC Tasarımı

Yunus Emre ESEN[1] 🔘 , **İsmail SAN[1]** 🔘

[1] *Eskişehir Technical University, Department of Electrical and Electronics Engineering, 26555, Eskişehir, Türkiye*

Abstract
High-quality sound processing requires hardware acceleration in order to reduce the processing latency of the applied sound effects. Computational latency of producing enhanced sound from the audio input is an important delay component and affects the performance especially in artists' live performance or high-quality sound generation. Artists want to apply a sound effect in real-time on their music and latency is the main problem when these systems running in real-time. CPU-based systems present flexibility, but introduce a high amount of latency while processing, which in fact affects the artist negatively. In this study, to get the flexibility through software and the acceleration via hardware specialization, we present a system-on-chip (SoC) solution with HW/SW co-design methodology for some sound-effects. We reduce the latency and increase the frequency by applying pipelining through MATLAB. The system is implemented and tested on a programmable SoC platform, ZedBoard, which contains ZC7020 Zynq chip with a dual-core ARM-Cortex-A9 processor and a reconfigurable FPGA part. The ARM processor enables the management of sound-effect hardware accelerator running on FPGA and provides communication with user. A sound effect is designed with block models provided by MATLAB & Simulink at high-level. MATLAB HDL Coder then converts these blocks into RTL-level hardware designs. The followed design methodology provided by MATLAB & Simulink enables high-level block design that can be embedded into FPGA at RTL-level to benefit from the speed provided by high-speed hardware registers and to have an AXI interconnect interfacing with software in order to utilize the software flexibility. The study shows that latency is reduced significantly.
Keywords: hardware accelerator; system-on-chip; pipelining; sound effect.

Öz
Yüksek kaliteli ses işleme, ses üzerinde uygulanan efektin sebep olduğu işleme gecikmesini düşürmek için donanım hızlandırması gerektirir. Ses girişinden gelişmiş ses oluşurken hesaplanan gecikme, önemli bir gecikme bileşenidir ve özellikle sanatçıların canlı performansında veya yüksek kaliteli ses üretiminde performansı etkiler. Sanatçılar, müziklerine gerçek zamanlı bir ses efekti eklemek ister ve bu efektler gerçek zamanlı sistemlerde kullanıldığında gecikme ana problem haline gelir. CPU tabanlı sistemler esneklik sunar, ama işleme esnasında oluşan büyük miktardaki gecikme sanatçıları olumsuz olarak etkiler. Bu çalışmada, esnekliğin yazılımla ve hızlandırmanın donanım özelleştirmesi ile elde edildiği, bazı ses efektleri için yazılım/donanım ortak tasarım yöntemi ile bir sistem üzerinde çip (SoC) çözümü sunuyoruz. MATLAB üzerinden ardışık düzen uygulayarak gecikmeyi azaltıyor ve frekansı artırıyoruz. Sistem, çift çekirdekli ARM Cortex A9 işlemcisi ve yeniden yapılandırılabilir FPGA'e sahip ZC7020 Zynq çipi barındıran programlanabilir bir SoC platformu olan ZedBoard'da çalıştırılmış ve test edilmiştir. ARM işlemcisi, FPGA üzerinde çalıştırılan ses efekti donanım hızlandırıcısının yönetilmesine ve kullanıcı ile haberleşilmesine olanak sağlar. Bir ses efekti MATLAB & Simulink tarafından sağlanan yüksek seviyede blok modeller ile tasarlanmıştır. MATLAB HDL Coder, bu blokları RTL seviyesinde donanım tasarımlarına dönüştürür. MATLAB & Simulink tarafından sağlanan bu tasarım yöntemi, yüksek hızlı donanım yazmaçlarının sağladığı hızdan yararlanmak için FPGA'ya RTL seviyesinde gömülebilen yüksek seviyeli blok tasarımı yapılmasına ve yazılım esnekliğinden yararlanmak için AXI ara bağlantısı ile arayüz oluşturulmasına olanak verir. Bu çalışma gecikmenin önemli bir ölçüde düşürüldüğünü gösterir.
Anahtar Kelimeler: donanım hızlandırıcısı; çip üzerinde sistem; ardışık düzen; ses efekti.

I. INTRODUCTION

Digital signal processing (DSP) is a key technology in many signal processing applications such as speech and image compression, audio filtering, classification and coding where processing operations are performed in digital domain. Many real-life examples which are possible with DSP technology have already been in our daily life such as video streaming, computer applications, audio effects etc. In real life, signals are in continuous form such as a sound signal that is heard. In order to process a signal in a conventional computer, continuous signal must be converted to digital format that computer can understand and perform the processing. First, real-world signals such as temperature, pressure, audio, or video are transformed into digital domain, then signal processing algorithms run on digitized signal samples and manipulate the digital signal data to get the desired transform.

Finally, the digitized data needs to be transformed back to real-world continuous signal again [1]. The most important problem here is relatively long processing delays. There are many delay components in a processor-based system. Some applications can tolerate long delays; however, some delay-intolerant applications require a special attention in order to reduce the overall delay introduced by computation or communication. Hence, specialized hardware architecture with DSP-enabled components that is tailored to the signal processing algorithm is a powerful solution.

The demand of people on the low-latency applications is increasing gradually in every year. Especially in last 30 years, with evolution of technology, microprocessors and specialized hardware are used and developed with newer techniques in order to obtain better throughputs and better execution times. Some of the solutions are being developed with high-level software programming language for CPU-based solutions, which gives more design productivity and development opportunities, but incurs a cost of processing time for application being developed [2]. Application-specific hardware design provides very low latency for the selected application and gives better processing times with very high-throughput values (Ahmed Elhossini, Shawki Areibi, Robert Dony, 2006), but this costs the designer to lose design flexibility that comes from higher level programming languages. This article presents a case study to use the advantages of the two mentioned design flows in a system-on-chip (SoC) design with a HW/SW codesign methodology: (1) designing a DSP block with a high-level programming tool in MATLAB & Simulink, (2) converting this application-specific design to a specialized hardware with MATLAB HDL Coder, (3) managing and controlling the generated this special hardware design from a software application running on ARM-based processor and (4) demonstrating how pipelining can be implemented to DSP designs to achieve higher frequency and how much its effect on latency times compared to our previous solution LowLAG [4].

Many works have been carried out in the past years for sound effect applications. Some of these have been carried out only on the software platform [5] and some have only been carried out on the hardware platform [6]. On the DSP side, main concentration of this article is to cover to audio applications, especially in real-time sound effects. Sound effects are latency-intolerant application for real-time applications. For instance, a person speaking to a microphone and listening the voice with headphones can easily notice a latency if the latency is greater than 5 milliseconds [7]. Software applications offer latency values on these boundaries. Designing complex software applications with high CPU loads or running an application that requires high performance on the CPU

can reduce the performance of the software. For such cases, software solution does not provide enough performance. High performance or high energy efficiency in computing can be achieved with compute and communication specialization for a specific task. Reconfigurable computing platforms allow us to tailor our design to perform better in terms of performance and area. In hardware, arithmetic and memory access of an application can be defined in spatially if the application has some parallelism. Spatial execution defined in hardware brings more performance- or energy-efficiency compared to serial execution model compiled into CPU. Because software applications are being developed with a high level language (such as C/C++), and even if the computation is described in one line of code that makes one useful operation at high-level, it may correspond to many lines of assembly language code. Among the instructions compiled from one-line of code, there are many instructions required to feed the data to the execution unit and it takes several cycles. However, the same operation can be described more efficiently in hardware. This gives the main advantage of low latency introduced by specialized hardware designs compared to software designs. As the software part of the SoC design handles the low load required processes such as address definitions, providing parameter inputs, controlling the hardware units etc., while hardware handles high load required processes. This HW/SW co-design methodology brings advantages of both software and hardware together. There are several studies on sound effects applications that are implemented on system on chip platform such as [8] [9] [10].

In this work, we developed our first work, LowLAG [4], while following the same methodology. There are three main contributions compared to first design, which are: obtaining better latency values with pipelined design (1), controlling hardware specification from software side (2), and designing new sound effects on Simulink that are based on a filter created via FDA Filter Designer Tool.

II. HARDWARE/SOFTWARE CO-DESIGN METHODOLOGY FOR LOW-LATENCY AUDIO PROCESSING

2.1. Low-Latency Sound Effect Design with Specialized Hardware

One of the competition points among Today's Tech companies is decreasing the latency and producing real-time products. Music is one of the important area since there are many real-time products available today. All techniques for low-latency and high-performance are utilized in the design stage (producer's side) in order for a person who listens to the music to get the best result. The artists, musicians and producer want to hear whatever they have done

immediately. Because music is all about harmony, and even a single millisecond delay can disturb the whole harmony and cause decreasing the creativity. This is another motivation point that highlights the need for acceleration on sound effects in terms of low-latency.

Software solutions have limited processing capability, which make the specialized hardware design inevitable in low-latency targets. But, designing a specialized hardware is difficult in terms of design cost, long development cycles, and huge verification efforts. We propose a sound effect design with high level system blocks on Simulink, test and simulate on a personal computer, convert the MATLAB & Simulink design to a hardware design in VHDL language generated via MATLAB HDL Coder, integrate the design to a SoC architecture containing ARM processor and implement the whole SoC in ZedBoard programmable SoC platform. Generated RTL-level specialized hardware design can be integrated into a SoC architecture as an IP (Intelligent Property) core, which can be explained briefly as a reusable part of hardware system in an FPGA (Field-Programmable Gate Array) based SoC system [11].

The desired solution for low-latency audio processing in our design requires processing the audio sample by sample which is synchronized with system clock. Generally, 44.1 KHz is used as a sample rate for most of the audio files, and we designed the system for this sample rate. In ideal conditions, an audio file that has two channels (right and left) have 88200 sample for one second. Our methodology brings the necessity to run at least 88.2 MHz frequency for system core clock and supporting this frequency within the IP core. Pipelining in the IP core design provides to reach especially this frequency requirements rather complicated effect designs.
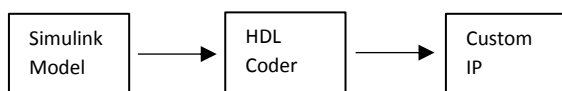
Figure 1. Simulink model to custom IP core

There are two different pipelining methods provided by HDL Coder that are implemented in this work. First method is adding new delay blocks to design stage of the Simulink model and using these blocks as registers. There is a special parameter to determine how many cycles are going to be waited at registers with this method. The second method is adaptive pipelining. This pipelining method is controlled from HDL Coder. There is a requirement that specifies the target platform for using adaptive pipelining. HDL Coder insert registers to block design if adaptive pipelining is enabled.

2.2. Hardware Design: Creating Custom IP with HDL Coder
The sound effect design is described in a Simulink model; whose used blocks must be supported by HDL

Coder [12]. In this work, as a sound effect, a filter design is used additional to previous work. A distortion sound effect with no parameters was designed for previous work. In this work, this distortion effect is updated with pipelining registers and input parameters. The filter design, which is new sound effect, is made on FDA Filter Designer Tool, and this design converted to a Simulink block. The block that is generated uses basic elements such register, gain, addition and subtraction block, which is supported by HDL Coder.

When Simulink block design of sound effect is completed, custom IP block can be generated with that block using HDL Coder. Vivado 2020.2 provides the synthesis tool to HDL Coder for converting the block to IP core design in VHDL. The IP core runs on an FPGA inside of SoC design and uses AXI4-Lite interface in order to communicate with whole system.

2.3. Software Design: Hardware Management and User Interface
The designed custom IP core is controlled over a software application runs on an ARM processor. The application handles the basic processes. It makes the address mapping with BSP (Board Support Package) of the board that is used, routing the audio data between IP cores etc. In order to use custom IP core, driver of the custom IP has to be added to software and associated with application for accessing the IP core. The driver includes the address of required addresses.

Other duty of the software is to provide communication between system and user. The system and user communicate over UART (Universal asynchronous receiver-transmitter) at 115200 baud rates. The interface is managing from a basic console, but it can be improving for advance applications.

2.4. Design Environments
Sound effect is designed on Model Composer and System Generator 2020.2 provided by Xilinx, based on MATLAB & Simulink version 2020b. Model Composer provides the full compatibility for Xilinx boards. Some of the steps for introducing the board, synthesis settings are already defined with Model Composer. It gives setting up the system easier than normal MATLAB environment.

Created custom IP core is implemented to system on Vivado 2020.2. This Vivado version provides new software platform rather than previous work. This new software environment name is Vitis and is used 2020.2 version. Vitis comes up with some change compared to SDK, which was old software environment for Vivado. The application and platform are separated on Vitis, which makes the target platforms configurable.

III. METHODOLOGY AND IMPELEMENTATION

The main idea behind this work has already summarized in Section 2. In this section, all details and project implementation are covered. The path to the solution is described in subsections below.

3.1. Necessity of HW/SW Co-design Methodology

Software based DSP solutions for audio processing may be not enough to handle when desired processing time is real-time. For example, complex sound effects generate more latency while monitoring the input audio to an output. Hardware based solutions are solved this latency problem. But it is hard to implement and results in the loss of functionality advantage of software solutions. When HW/SW co-design methodology is used, the latency results will be so much better than software, and hardware still can be manipulated from the software. Therefore HW/SW co-design methodology offers a solution where is used advantages of both sides.

3.2. System-on-Chip Design Architecture with Sound Effects on ZedBoard

In this work, an SoC design is implemented on ZedBoard, which is provided by Xilinx. ZedBoard includes a Zynq-based programmable SoC platform. This platform works within an ARM processor core. The ZedBoard that we used contains ZC7020 chip.

SoC platform can be explained in two parts, PL (Programmable Logic) that was mentioned as hardware and PS (Processing System) that was mentioned as software. PL side is powered within an FPGA and PS side is run on an ARM processor. PL consists of PS, which can be shown in Figure **2**. The IP cores are the changeable components of the PL which means that IP cores can be used in different designs again.
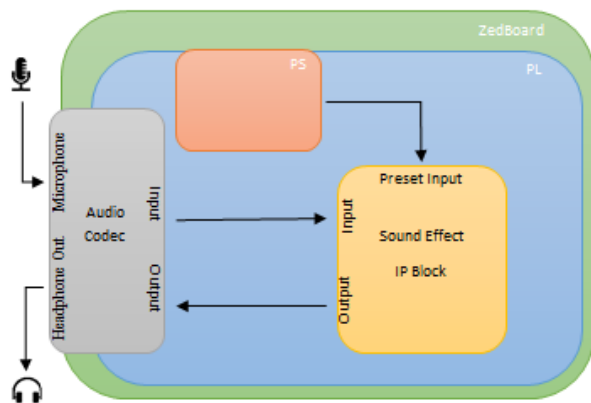
Figure 2. Overview of SoC Architecture

3.3. Sound Effect Design on MATLAB & Simulink

Two different sound effect designs are made on Simulink for this work. First design is related with previous work, which is distortion effect. Addition to previous design, parameter inputs for managing effect from software and pipelining for decreasing latency more were added. Seconds design is about a lowpass filter. Many more filter designs can be implemented like the way is going to explained below.

3.3.1. Distortion Effect Design

Distortion effect distorts the audio signal. It applies a gain to incoming audio signal and saturates that signal in order to keep signal in valid area with a saturation block [13]. In this version, there are 2 parameters for managing distortion effect. First parameter is drive, which provides the adjusting level of distortion level, and second parameter is mix, which provides how much percentage of clean and distorted audio is going to mixed. The design of distortion effect can be shown Figure 3. Components are described in Table 1.

Table 1. Distortion effect block explanations

Block	Explanation
audio_in	Audio signal incoming from MATLAB workspace
ConstantDrive	Input of drive parameter (0 minimum, 1 maximum), that sets intensity of the effect
ConstantMix	Input of mix parameter (0 minimum, 1 maximum), that sets rate of mixing with distorted and clean audio
distortion subsy	Distortion sound effect block
convert & double	Converting blocks between fixed point and double data types.
output.wav	Saves the audio to a multimedia file

As shown in Figure 3, distortion sound effect takes three input parameters and provides an output to obtain effected audio. Audio input is obtained from MATLAB workspace and output audio is saved to a wav file. The incoming data values and saved data values are in double data type. Distortion block performs the computation in fixed point data type with 16 bits length with 14-bit fraction length, which is proposed in [14]. When looking into the subsystem of distortion, there can be seen pipelining methodology is implemented. This methodology uses registers for accelerating the hardware. Delay blocks are used as register. They keep the corresponded data for one clock cycle of the system. It provides to store data in short distances between blocks where they have computation processes. When the pipelined design is implemented, there is no need to wait for completing the process for one sample, it can take more cycles in a one subsystem in sequential line. Implementing pipelined design increases the latency in theorical, but it improves the throughput, processing power and also system clock can be run at higher frequencies, which gives reducing latency values compared to the design without pipelining. Pipelined design of the distortion block is shown in Figure 4. There are several blocks and they are used mainly for multiplication, addition,

saturation, delay and supplying a constant to another block. The delay blocks perform a wait operation for one clock cycle. These delay blocks must be added in right places in the design in order to making processes in parallel. Additional registers can be added for blocks with heavy processing load. z^{-1} block is a delay component and it represents a pipeline register with reconfigurable clock-cycle.

Registers are implemented as with these parameters: input pipeline is 1, output pipeline is 1. These parameters are provided from HDL Coder and, all used blocks are supported from HDL Coder toolbox in Simulink library. All multiplication, subtraction and addition processes are done sample based. In Figure 4, there can be imagined 5 register stages. Each register transmits the data on rising edge of a clock cycle of system to another register stage. Inside of the subsystem signals are used in signed fixed point data type with 16 bit data length and 14 bit fraction length suggested as in [14]. Besides, adaptive pipelining is also used in this block design.
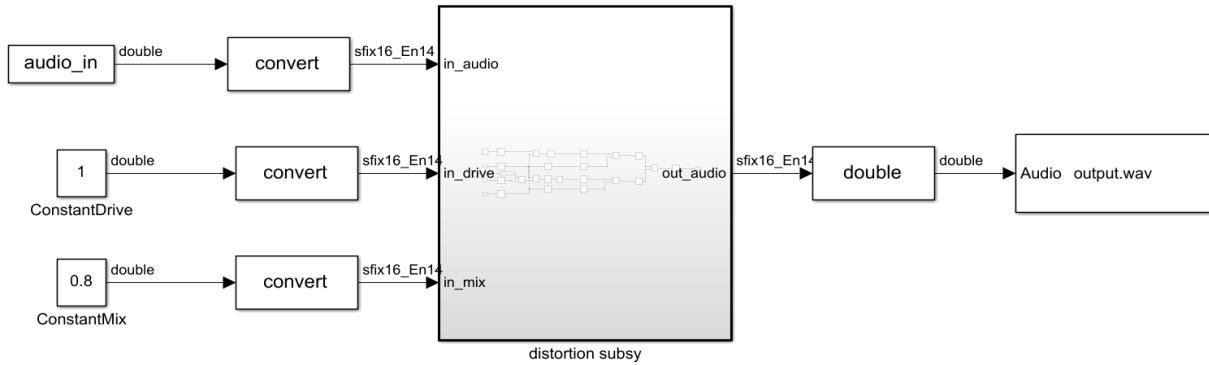
Figure 3. Simulink model of distortion effect
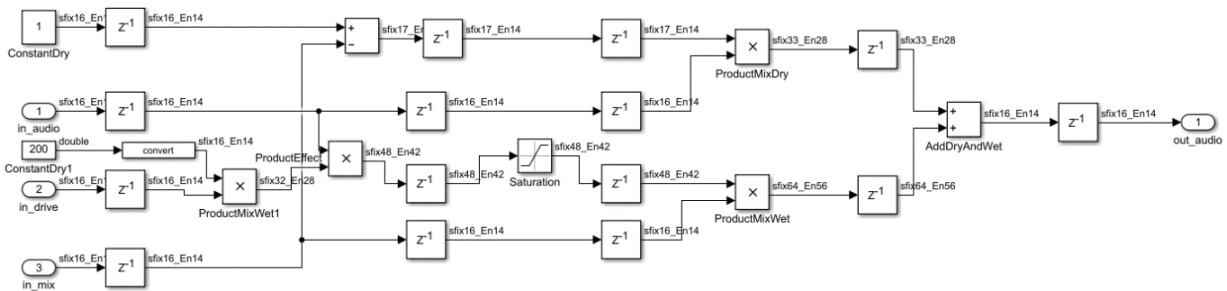
Figure 4. Subsystem of distortion effect in detail

3.3.2. Lowpass Filter Design
A lowpass filter design is created in order to show how Simulink gives ease of design for hardware implementations. This lowpass filter passes lower frequencies than 3 KHz and stops higher frequencies than 4 KHz where can be seen in Figure 5 as magnitude response graph. Filter Designer tool (as known as FDA Tool) is used for creating filter block. Filter designer can be used by entering filterDesigner to MATLAB command window. The tool will be opened after this process. The tool provides some parameters to select and it makes easier to design a filter. In this work, an equiripple FIR filter is used. Equripple filters are suited for specific tolerance for passing and stopping for some frequencies [15]. The order of design may be reduced with other design methods. It may reduce the latency more because reducing order means that less blocks are going to used. Since the methodology is emphasized here, the most efficient design has not been studied. The filter design in the Filter Designer is shown on Figure 5.

When filter design is finished, the design can be converted to a Simulink model from File > Export the Simulink Model selection. Input processing is selected as sample based and all optimization selections are checked before realized model. The block that is generated is added to another Simulink model. This block design is generated automatically, which is shown in Figure 7. In this design, there are 112 delay blocks which is corresponded to design order in the Filter Designer and they are used as register as well with 1 input and 1 output pipeline parameters. The gain blocks in the design enables to making element wise multiplication with a constant [16] defined from Filter Designer. The generated subsystem block is also shown as in Figure 6. The blocks that are shown in figure are described in Table 1 before, except filter_lowpass block. This block is used as lowpass filter sound effect and created from Filter Designer Tool as mentioned above.
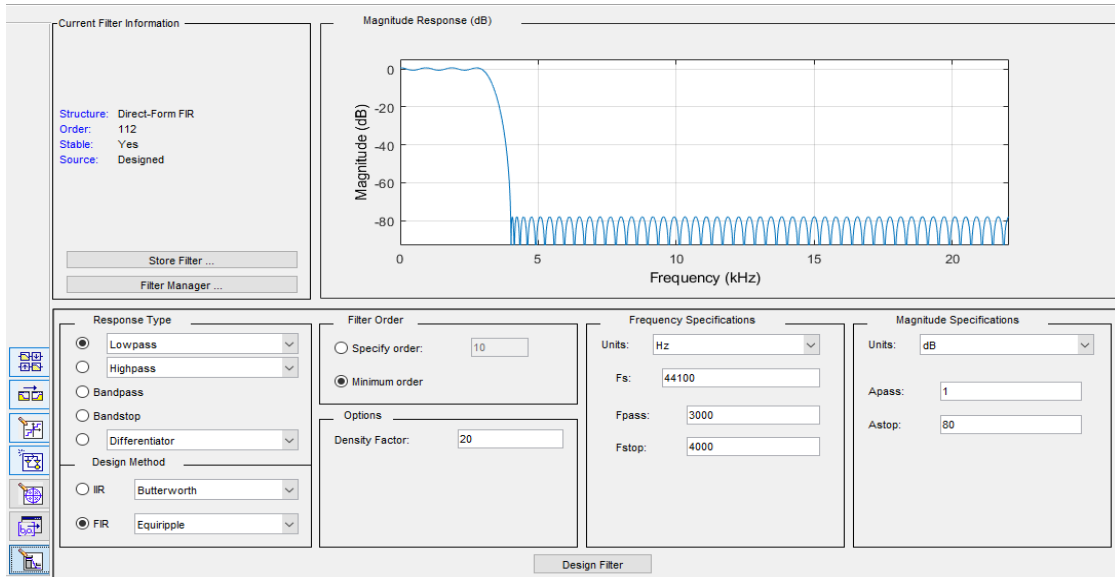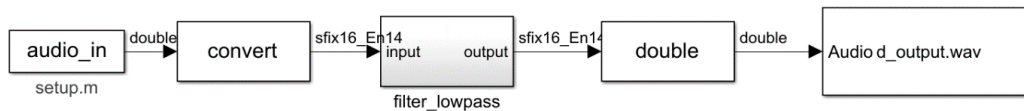
Figure 5. Filter design in the Filter Designer tool

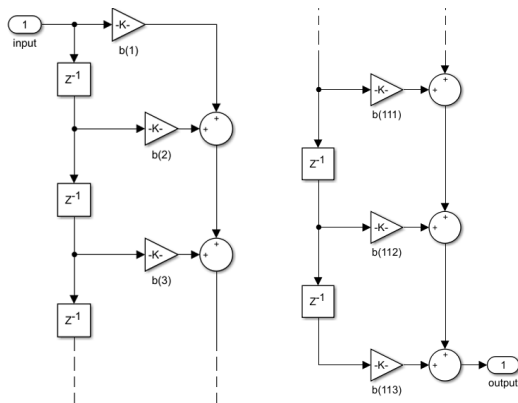
Figure 6. Filter block in the Simulink

Figure 7. Inside of designed filter block in Simulink (dashed lines represent that blocks repeat in the same way from input to output)

The basic blocks in the design are supported from HDL Coder. Gain block parameter K has different values which are defined from Filter Designer tool. Lowpass filter block is selected as use adaptive pipelining from HDL Coder properties. Also, all delay blocks are used as register in order to implement pipelining for better execution times.

3.4. Creating Custom IP Cores via HDL Coder

After design of sound effect has completed, the sound effect block in the Simulink is converted to IP core via HDL Coder. Before converting the filter block to IP core, model and subsystem compatibility is checked with HDL Coder. These steps show if there is any error or warning for better design compatibility. Some of the warnings that is taken in that stage was about reset type of clock settings, changing block names etc.

These checking tools gives facility of seeing any mistake before generating IP core.

HDL Workflow Advisor is a tool that is used for creating custom IP core provided by HDL Coder. This tool provides to selecting target platform, preparing the model for HDL code generation and generating IP core. Since Simulink is opened from Model Composer and System Generator, the board that is going to used is already set up to HDL Coder. It provides pass these processes faster than normal Simulink environment. The settings that are used for ZedBoard are described in Table 2. IP core is generated as used these parameters.

Table 2. HDL Workflow Adviser Parameters

Parameter	Input
Target workflow	IP Core Generation
Target platform	Generic Xilinx Plaform
Synthesis tool	Xilinx Vivado
Tool version	2020.2
Family	Zynq
Device	xc7z020
Package	clg484
Speed	-1
Processor/FPGA synchronization	Coprocessing - blocking
Target platform interfaces	AXI4-Lite
Language	VHDL
Adaptive pipelining	Checked

3.5. Adding Generated IP Core to a Hardware Design in Vivado

The hardware design is synthesized, implemented and routed using Vivado design suite (Vivado v2020.2). First, a block design is created, and this design uses ZYNQ7 Processing System IP core to instantiate ARM processor in the system. An audio codec IP core is used from Zynq Book Tutorials [11] to obtain audio signals from microphone inputs and giving back to the headphone output. All IP cores communicates with each other over AXI interconnect.

Sound effect IP blocks must be added first to the repository. For this purpose, new IP repository is added where is located to generated IP core location in the IP Catalog, it adds IP's automatically from the repository. After adding process is done, IP's can be added to the block design. IP block connections are done automatically thanks to the Vivado. After synthesis, placing and route operations, a bitstream file is generated to program FPGA. The bitstream file must be exported to use this hardware design in composing software. Final block design of the PL part of the SoC architecture is shown in Figure 8.
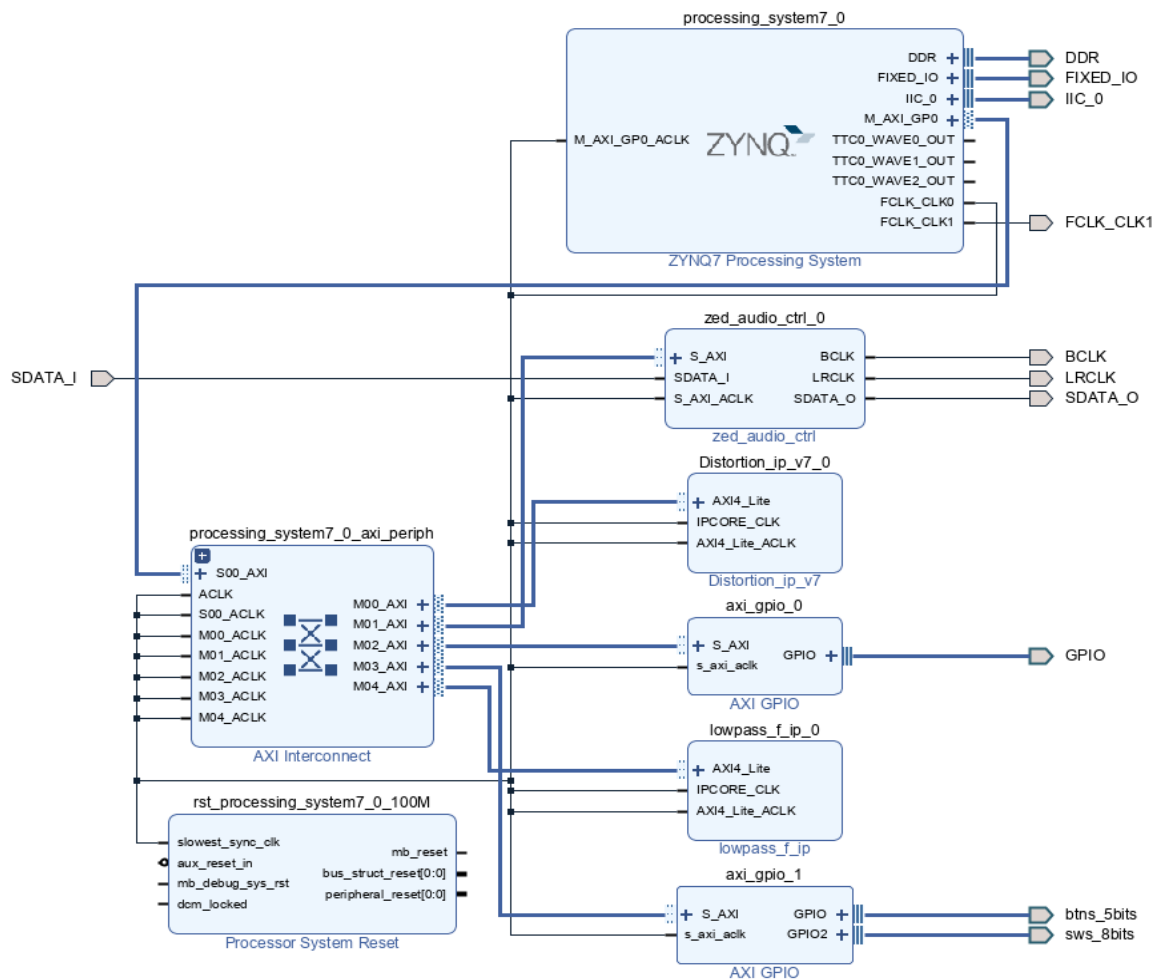
Figure 8. Block design of SoC architecture

3.6. Composing the Software on Vitis

Vitis, which is a recent and unified software development environment provided by Xilinx, is used for software part of the design. New application project and platform project are created first. Platform project is used to define the hardware specification as a platform. It defines hardware specifications such as drivers for hardware design etc. It is used bitstream of hardware, which is generated on Vivado, and it can be updated when hardware design is changed. On the other side, application project is the part of using hardware IP's and using algorithms in order to realize what is desired from software side. The application

project must be created with this platform project instance. IP core drivers must be added to the application project for address definitions of corresponded IP cores.

The application project provides an UART connection in order to communicating with user. According to given input parameters to UART, application transmits the audio signal between the corresponding addresses of sound effect IP core and input & output registers of audio codec IP core. These addresses are determined in HDL Coder IP core generation step and defined in IP core drivers which are created

automatically by HDL Coder. Also, input parameters are managed in here, which was mentioned before changing the behaviour of sound effect.

3.7. Internal Structure of Lowpass Filter Sound Effect IP Core

In the lowpass filter IP core there are 4 different modules. These modules are described in this subsection. Figure 9 illustrates the internal structure of lowpass filter IP core.
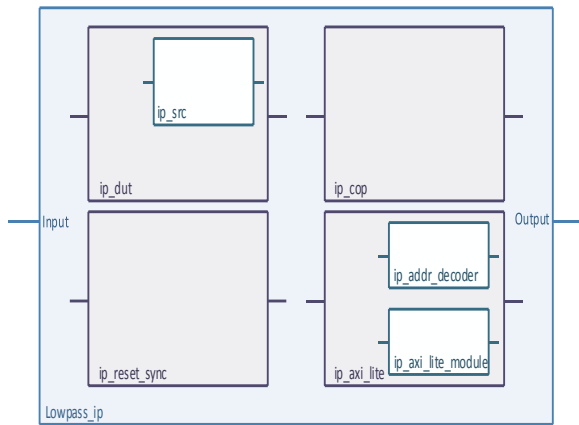

Figure 9. Illustration of Lowpass Filter IP Core in Detail

3.7.1. ip_dut

This submodule is used for processing the audio signal. It has one more component sub-module named ip_src which makes the processing operations and pipelining management. The input and output are fixed point data which was defined in Simulink design. In addition to these, reset, clock and enable signals goes as input to this block.

3.7.2. ip_reset_sync

This submodule manages the reset state of IP. When a reset signal is emitted to this IP core, it resets states for processing and pipelining processes.

3.7.3. ip_cop

This submodule is used as controller of processes. It manages the states of controller such as strobe, enable and ready signals. This module generates a set of signals to control (starting, enabling for write operation, etc.) the other hardware components in the design and also checks the ready signals to determine whether they finished the execution.

3.7.4. ip_axi_lite

This submodule manages the communication of hardware IP core with other IP's in the system through AXI4-Lite interface. It handles the necessary signals with cooperating other submodules that are mentioned above. It has two other submodules inside of itself as component. Input and output signals transfer from this submodule.

IV. EXPERIMENTAL RESULTS

4.1. Latency for Simulink Solutions

Everything described in the methodology section has been carried out on Simulink and ZedBoard. The latency results for Simulink models could not be obtained. Because the delay blocks that are used for pipelining are caused delaying in audio, which makes the simulation results meaningless. But the driver latency is already known from previous work. The latency result can be taken minimum value as 2.447 milliseconds without applying an effect and 2.8 milliseconds with distortion IP block from previous work with ASIO4ALL v2 driver, which can give an idea for comparison the results obtained from ZedBoard. These results were evaluated with a computer which has Intel i7 7700HQ processor, Realtek High Definition Audio driver version 6.0.1.8142, and audio codec ALC269 at 44.1 kHz.

4.2. Latency Calculation for ZedBoard

The latency values are calculated with a TTC timer provided by Zynq device on the ZedBoard. When a sample is processed, it is obtained the time value from physical timer counter register and stored previous value in different variable. The latency calculation is shown in Figure 10 and described below.
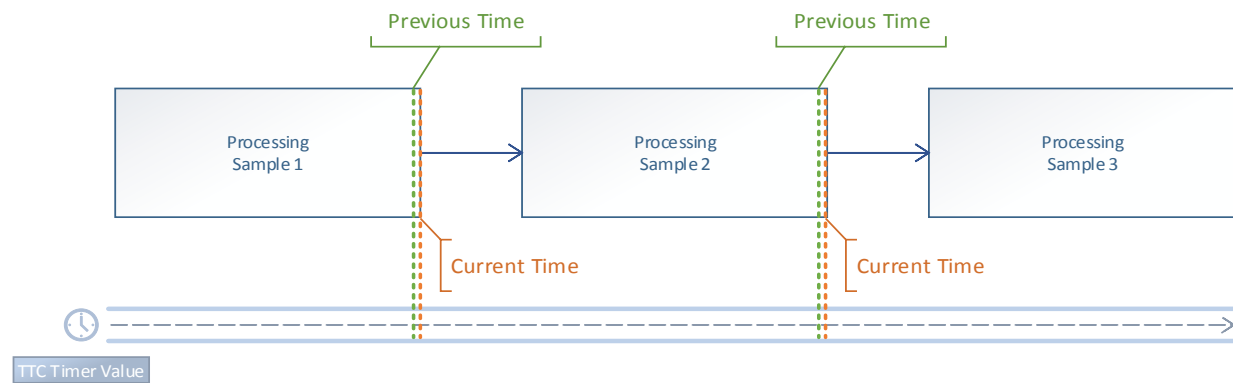

Figure 10. Calculating latency and representation of times

The algorithm of obtaining latency is described with a basic pseudo code below.

```
previous_time = current_time;

current_time = GetTime();

elapsed_time = current_time - previous_time;
```

If relation between the Figure 10 and code above is examined, previous time is the current time in the previous processed sample. Previous time is stored in another variable and elapsed time is calculated as difference of previous and current time. This method provides to obtain a latency value with time elapsed outside the IP core.

4.3. Latency Results for ZedBoard Implementation
Different implementations are tested on ZedBoard. Distortion sound effect and lowpass filter sound effect are implemented together and one by one on hardware. The test specifications and results are described in Table 3. The average values are obtained from 10 test records.

Table 3. Test specifications and results implemented on ZedBoard

Sound Effect(s)	Latency from Pipelining (cycle)	Clock Frequency	Average Clock Cycle Spent	Average Latency (nano-seconds)
Distortion + Lowpass Filter	-	100 MHz	710	1065
Distortion	12	150 MHz	588	882
Lowpass Filter	226	100 MHz	714	1071

Clock frequency shows the frequency that can be reached at maximum value for the designed sound effect IP block. Results show while increasing clock frequency, latency is decreasing. The way of increasing the clock frequency is using pipelined register designs. But if these pipelined registers are increased more than enough, that would be cost of more latency and this puts the results at a disadvantage instead of advantage. Different sound effects have almost same latency values at the same clock frequency.

Unlike similar studies, in this study specific latency values are obtained and compared by using MATLAB & Simulink, which makes much easier to design development and saves time, for IP core design with pipelining technique, including previous study [4]. Significant improvement in latency has been demonstrated. Using pipelining on complex Simulink designs make it synthesizable to an IP Core that is run at desired clock frequencies. This gives the opportunity of processing an audio in a specific time.

As a result, designs with pipelined registers reduce the latency. The best latency values are under the 1 microsecond (882 nanoseconds). There is 12.75% decreasing in latency value for distortion block with pipelined design compared to previous work. Much better Simulink designs can be made that it is opened to development for less latency values. This study also revealed how easy it is and showed have good performance values for designing DSP system by using MATLAB & Simulink.

V. CONCLUSION
This study shows the advantages of HW/SW co-design methodology with a system-on-chip design on audio processing. Very small latency values can be obtained with using specialized hardware while controlling it over a software application for DSP systems without losing any functionality of software. Development time and design productivity significantly improved by making block-based designs in MATLAB & Simulink. High-level DSP systems can be implemented easily with this methodology. For sound-effect applications, the followed methodology can be a viable solution on real time applications. In this context, neural network-based audio synthesis algorithms can also be accelerated as a future work to provide real-time performance via similar design methodology with MATLAB or other high-level synthesis tools.

REFERENCES
[1] Blackledge, J. (2006). In Digital Signal Processing (2nd ed.). Chichester: Horwood Publishing.

[2] Pirkle, W. C. (2019). Designing audio effect plugins in C++: for AAX, AU, and VST3 with DSP theory (2nd ed.). New York, NY, USA: Routledge.

[3] Elhossini, A., Areibi, S., and Dony, R. (2006). An FPGA Implementation of the LMS Adaptive Filter for Audio Processing. *2006 IEEE International Conference on Reconfigurable Computing and FPGA's (ReConFig 2006)*. San Luis Potosi, Mexico: IEEE.

[4] Esen, Y. E., and San, İ. (2020). LowLAG: Low-latency hardware accelerator of a sound effect with system-on-chip design. *ASYU 2020, Innovations in Intelligent Systems and Applications Conference.* İstanbul.

[5] Juillerat, N., Arisona S. M., and Schubiger-Banz, S. (2007). REAL-TIME, LOW LATENCY AUDIO PROCESSING IN JAVA. *Proceeding of the International Computer Music Conference.* Copenhagen.

[6] Meyer-Baese, U. (2007). Digital Signal Processing with Field Programmable Gate Arrays (2nd ed.). Berlin: Springer.

[7]   Audio quailty on networked systems, Yamaha, https://uk.yamaha.com/en/products/contents/proaudio/docs/audio_quality/05_audio_quality.html (June 2020)

[8]   Pfaff, M., Malzner, D., Seifert, J., Traxler, J., Weber, H., and Wiendl, G. (2007). IMPLEMENTING DIGITAL AUDIO EFFECTS USING A HARDWARE/SOFTWARE. *10th Int. Conference on Digital Audio Effects (DAFx-07).* Bordeaux.

[9]   Byun, K., Kwon, Y., Park, S., and Eum N. (2009). Digital Audio Effect System-on-a-Chip Based on Embedded DSP Core. *ETRI Journal.*

[10]   Byun, K., Kwon Y., Koo B., Eum N., Jeong K. and Koo J.. (2009). Implementation of digital audio effect SoC. *2009 IEEE International Conference on Multimedia and Expo.* New York.

[11]   Crockett, L. H., Elliot, R. A., Enderwitz, M. A., and Stewart, R. W. (2014). The Zynq Book: Embedded Processing with the ARM Cortex-A9 on the Xilinx Zynq-7000 All Programmable SoC (First ed.). Glasgow: Strathclyde Academic Media.

[12]   HDL Coder™ Getting Started Guide, The MathWorks, Inc., https://www.mathworks.com/help/pdf_doc/hdlcoder/hdlcoder_gs.pdf (June 2020)

[13]   Limit input signal to the upper and lower saturation values, The MathWorks, Inc., https://www.mathworks.com/help/releases/R2018b/simulink/slref/saturation.html (June 2020)

[14]   Fixed-Point Conversion. In HDL Coder™ User's Guide, The MathWorks, Inc., https://ww2.mathworks.cn/help/pdf_doc/hdlcoder/hdlcoder_ug.pdf (June 2020)

[15]   Signal Processing Toolbox™ User's Guide, The MathWorks Inc., https://www.mathworks.com/help/pdf_doc/signal/signal.pdf (December 2021)

[16]   Multiply input by constant. The MathWorks Inc., https://www.mathworks.com/help/releases/R2018b/simulink/slref/gain.html (June 2020)

Thumbnail Selection with Convolutional Neural Network Based on Emotion Detection

Duygu Algılamaya Dayalı Evrişimli Sinir Ağı ile Küçük Resim Seçimi

Mahmut ÇAKAR[1] (iD) **, Kazım YILDIZ[2]** (iD) **, Önder DEMİR[2]** (iD)

[1]*Marmara University Institute of Pure and Applied Sciences, Computer Engineering, Istanbul, Turkey*
[2]*Marmara University Technology Faculty, Computer Engineering, Istanbul, Turkey*

Abstract

The use of video broadcasting platforms is increasing day by day. The competition for developing platforms for the broadcasting and sharing of movies and TV series is increasing. The purpose of reproducing these platforms is to increase the quality and to trace them on a single platform. Film and TV series platforms use artificial intelligence algorithms for these shares. The aim of this study is to create more attractive cover photos for users by finding suitable frames from a movie or TV series. First, the frames that were transformed into covers/small pictures on the platform were obtained. Unnecessary frames which consist of closed eyes, blurry frames, or faceless images have been removed. Also, deep learning is used to label images with objects and emotions based on the identity of the face. The thumbnails with the most repeating faces were selected by developing a face recognition model at each step. The experimental results showed that the emotion model was successful.

Key words: Thumbnail, Emotion Detection, Video Streaming Platforms, Convolutional Neural Network.

Öz

Video yayın platformlarının kullanımı her geçen gün artmaktadır. Filmlerin ve dizilerin yayınlanması ve paylaşılması için platformlar geliştirme rekabeti artıyor. Bu platformların çoğaltılmasındaki amaç, kaliteyi arttırmak ve tek bir platform üzerinde takip etmektir. Film ve dizi platformları bu paylaşımlar için yapay zeka algoritmaları kullanır. Bu çalışmanın amacı, bir film veya diziden uygun kareler bularak kullanıcılar için daha çekici kapak fotoğrafları oluşturmaktır. Öncelikle platform üzerinde kapak / küçük resim haline getirilen çerçeveler elde edildi. Kapalı gözler, bulanık çerçeveler veya yüzsüz görüntülerden oluşan gereksiz çerçeveler kaldırıldı. Ayrıca derin öğrenme, görüntüleri yüzün kimliğine göre nesneler ve duygularla etiketlemek için kullanılır. En çok tekrar eden yüzlere sahip küçük resimler, her adımda bir yüz tanıma modeli geliştirilerek seçildi. Deneysel sonuçlar duygu modelinin başarılı olduğunu gösterdi.

Anahtar Kelimeler: Küçük resim, Duygu Algılama, Video Akış Platformları, Evrişimsel Sinir Ağı.

I. INTRODUCTION

Developments in technology enable internet content to be accessed as video content on mobile phones anytime and anywhere. For this reason, there is a huge increase in watching video content. YouTube [1] has more than two billion users and a billion hours of video is consumed every day. For this reason, more and more videos are produced every day. Due to the circumstances mentioned, it is very important to choose the title and thumbnail of the video (the thumbnail is a compressed preview of the original version used as a placeholder). The number of online TV series, movies and shows watching platforms is increasing. According to the last quarter report of 2020, Netflix, one of the media services provider, has approximately 204 million paid subscriptions [2], 150 million globally in Amazon Prime [3] and the number of ad-free subscribers in Hulu is 36 million [4] according to 2020 data. has been. Online watching platforms have emerged as competitors over time by aiming for users to spend more time in their application. It is aimed to increase the number of content on the platform and to present the existing content to users better and more attractive.

In the study, it is aimed to produce and label the pictures of the movie in a way that fits the cover art. Thus, it is expected that more original cover photos will be created with labels that may attract users' attention. Users' interests can be famous people, happy or exciting moments. After finding faces on the scene, a convolutional neural network-based algorithm has been developed to discover their identities, emotions, and other objects. In addition, closed eyes on the frame are detected and eliminated, and various parameters of the frame that can provide information are calculated. Based on the information obtained, the frame is tagged to select the most suitable thumbnails for users.

There are many studies on the selection of cover images which proceed by clustering the frames and choosing the most suitable one [5-9]. It is aimed to measure the pictures with parameters such as blur, clarity, colors, scene, composition and the presence of objects in the studies carried out for aesthetically scoring [10-12].

The source of inspiration for another study on the subject is Netflix's AVA system. Netflix also uses its own system and evaluates the AVA system. AVA is a collection of tools and algorithms designed to extract high quality images from videos on Netflix. There are three basic steps in the evaluation phase. Visual metadata such as contrast, color, brightness and motion blur are collected in the first step. Contextual metadata includes elements such as face detection, motion prediction, camera shot identification and object detection in the frame. Finally, photography, cinematography and visual aesthetic design processes are discussed in composition Metadata. Afterwards, sorting by image, face recognition, calculation of different camera angles for visual diversity and filters for maturity are made [13].

An algorithm consisting of five main steps including down-sampling, filtering, feature extraction, sorting and optimization has been developed [14]. In the down-sampling process, there are four steps: removing the first and last ten percent of the promotional video, sampling one frame per second, eliminating similar frames, and sorting the shots by length. In the filtering step, it performs the calculation of saturation, brightness, sharpness and contrast. It also includes removing subtitles embedded in the frame, finding characters, and adjusting the threshold.

Yu and colleagues expanded the video to include the title, description, and audio to define the content. The information obtained was used in selection models. It samples the developed model squares equally over time. Returns the highest aesthetic scores in the subset with a double column convolutional neural network to avoid the computational burden of rendering all frames. Frame properties extracted from VGG16, text properties from ELECTRA and sound properties of TRILL are obtained by the developed model [15].

Huang and friends discussed the task of highlighting the video and thumbnail selection from a different perspective. thumbnails and video clips were selected using the bulleted display, a new feature appearing on online video streaming sites popular in East Asia. The proposed method was compared with a KKStream which is East Asia's popular streaming service provider. Experimental results show that most participants are satisfied with the thumbnails and video clips chosen in their own way. Therefore, the cover screen can be a valuable resource for understanding the video [16].

In another study, automatic thumbnail selection was performed for movies and TV shows. Thumbnail selections are automated using the classifier. The performance of different convolutional neural networks (CNNs), namely VGG-19, Inception-v3 and ResNet-50, has been compared. Hybrid approach is designed which performs best in thumbnail selection. In the hybrid model, CNN feature extraction was used in genetic programming classification. ResNet-50 CNN performed better than other CNN models [17].

The face recognition model was developed by preserving the method in our previous study [23]. In addition, by selecting thumbnails with more repetitive faces, the thumbnail pool is narrowed, and it is aimed to select thumbnails that are more eye-catching and out of context.

Section 2 contains literature information about the study. In Section 3, the steps of the developed methodology are explained and the flow chart of the algorithm is given. In section 4, the results obtained are given and interpreted. Conclusion part is given as finally.

II. MATERIAL AND METHOD

Figure 1 shows the flowchart of the developed model. After taking the frames from the video, there are two stages: selecting the appropriate frames and using the convolutional neural network. Eye aspect ratio is made after detecting the faces in the frame in the down sampling part. The scope of this stage is to eliminate the unnecessary images via face detection to add less pictures to the model. This step is completed with parameter calculations. In order to find faces and objects related to the emotion and character, the convolutional neural network was used as the following stage. In addition, face recognition was performed and the thumbnail of the most repetitive faces was selected. Thus, it was aimed to reduce the number of results as well as identify the characters in the movie.
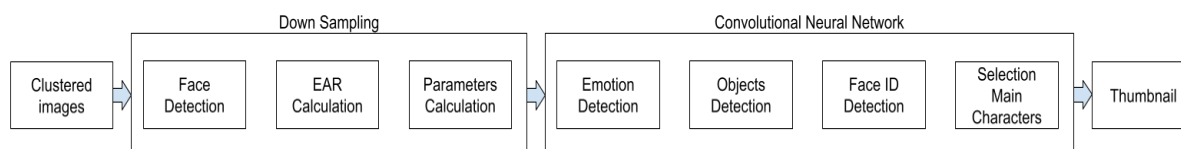

Figure 1. The schematic diagram of proposed method

2.1. Down Sampling

In a movie, the elimination process for unnecessary frames is important. When a two-hour film is analyzed, there are approximately two hundred thousand frames. The use of all these frames together with artificial neural networks is quite costly. In order to eliminate this memory problem, only the frames which have face images were selected using the Haar Cascade.

Comparing the Haar Cascade and Directed Gradients (HOG) in the face detection process, the detection time was measured as 16-50 ms in the haar cascade method and 340-410 ms in the other one. For this reason, Haar Cascade was used in the study for faster and more precise face detection. Histograms of Directed Gradients (HOG) were selected for eye opening detection. Faceless images were eliminated at this stage, and the variance of the Laplacian Filter [18] was used to eliminate blurry frames. The threshold value which may not be suitable for every movie or frame, so it has to be chosen different. For each frame a different threshold value should be calculated.

Figure 2 shows eye aspect ratio samples for several eye images. Face recognition is performed with HOG in the frame of the faces and the eye opening is determined. HOG is used because it allows you to define 68 different points with 6 points for each eye found. Thus, Eye Aspect Ratio (EAR) can be calculated as seen in Figure 2 with 6 points corresponding to one eye [19].

| EYE								
EAR	0.40	0.39	0.34	0.33	0.31	0.29	0.21	0.12

Figure 2. Eye Aspect Ratio samples for several eye images

The 6 dots represented by red dots are p1 (far left), p2 (top left), p3 (top right), p4 (far right), p5 (bottom right), and p6 (bottom-left) respectively. Eye Aspect Ratio is calculated as in equation 1 and eyes are assumed to be closed according to the threshold value.

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{1}$$

Many more parameters can be calculated in the framework that can inform us. These are brightness, dominant color and mist. The Laplacian filter was used to calculate the blur parameter, which is used to find the clearest in similar frames and eliminate the blurry ones. For the brightness value, RGB (Red, Green, Blue) values are converted to HSV (Hue, Saturation, Value / Brightness) format and calculated by taking the average of the value of each square. The K-means algorithm has been used to calculate the dominant color [20]. In K-means, a single cluster is created by choosing k as 1 and the center point is determined as the dominant color. With this value, it is aimed to prevent overlapping of similar colors with the film logo and to calculate the dominant color of the logo.

2.2. Convolutional Neural Network

Convolutional Neural networks require a large set of N-tagged images {x, y} specifying the discrete variable representing the real class as y, as opposed to the input x. It uses a loss function to compare the output of the model with the actual class value (y). It trains the weight matrices in the fully connected layers and parameters of the network by spreading the loss derivative backward according to the parameters in the network using filters in the convolutional layers and updating the parameters by stochastic gradient descent [21]. Convolutional Neural Networks consist of neurons with learnable weights and biases. It is also used effectively in situations such as image recognition, classification and object detection. Using convolutional neural networks, it is ensured that the emotions of the faces in the frame are detected and the objects are identified and associated.

In the study, it was aimed to detect emotions by using the Fer2013 [22] data set in order to determine the emotions on the face. 7 different emotions, including neutral, sadness, surprise, happiness, fear, anger and disgust, are tagged in the data set. The process of perceiving emotions is an important stage in terms of framing and grouping. It is also important that other objects can be detected in the frame. The presence of the pet can be considered as friendship. Finding a weapon or a sword in the frame can be seen as an action. Having a ball in the frame can determine that it is related to sports. Rather than providing information about the entire movie, this information is more important to the user's interest. For example, if someone interested in extreme sports sees a picture of nature on the cover art of the movie, it may increase the probability of choosing that movie.

Fig.3 shows the emotion model architecture which has improved according to our previous study [23]. The new emotion model is based on another study [24] but it's optimizer from ADAM to SGD (Stochastic gradient descent) replaced. Residual network is used with 224x224x3 shaped input. After flatting RESNET output layer, Dense and Dropout with ReLU activation function is applied to model [25]. Output shape is decreased with Dense with Softmax activation function to 7 since there are 7 type of emotion.
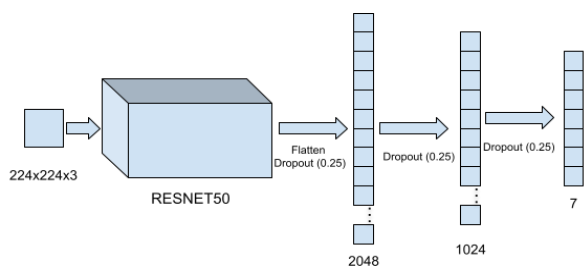
Figure 3. Architecture of Emotion Detection Network

YOLOv3 (You Only Look Once) [26] and Google OpenImages dataset [27] were used for object detection. The Open Images dataset includes about nine million images. It contains complex scenes like object bounding boxes, localized narratives, object segmentations, and more. YOLO algorithm based on prediction grids and anchors. In this study non-maximal suppression approach used to predicted and eliminated duplicates with 13x13 grids and 5 anchors. Although the YOLOv3 algorithm produced fast and accurate results, the OpenImages dataset achieved lower performance than expected. 600 classes in the data set were grouped and reduced.

It is important to identify who owns the faces in the frame to increase the likelihood of the user choosing movies featuring their favorite actor or actresses. Due to this aim, lib's pre-trained face detector [28] is used. In addition to the our previous study [23], a new feature has been added to face recognition. Every face in the frames is added to the face recognition pattern. After the process was finished, the frames with the most repeating faces were selected. As a result, logos are placed on the frames that do not coincide with the face parts. Since this step can also be performed better under human control, it is produced with and without logo.

III. RESULTS

Emotion detection model's accuracy has improved on training from 0.91 to 0.997 and on validation from 0.67 to 0.692. Figure 4 shows train loss and accuracy of the model.
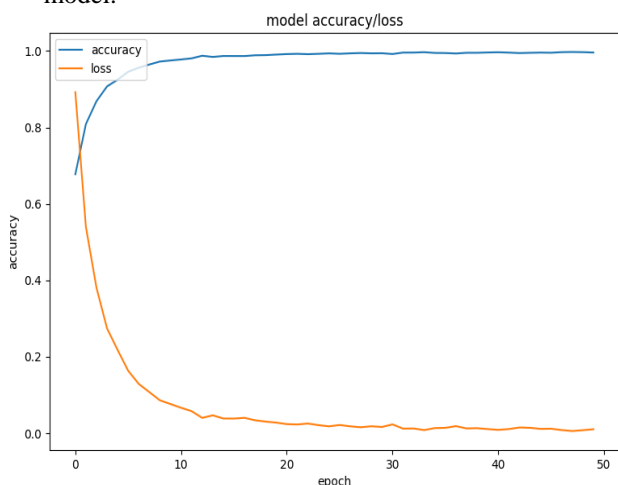

Figure 4. Train Loss and Accuracy Values of Model

Fig.5 and Fig.6 show the loss graph of 17 and 19 classes model loss graph respectively. It can be seen that the 17 classes model more accurate than 19 one. 17 classes model's mAP (mean Average Precision) is 34.28%. Training session was stopped because of 19 grouped classes the model did not reach the expected values. The resulting frames are manually examined so meaningless pictures, incorrectly grouped or tagged are eliminated. Thus, 200,000 frames in the film are reduced to 100 frames In addition, the selection of similar pictures and labels is made at different times and shown to the user.
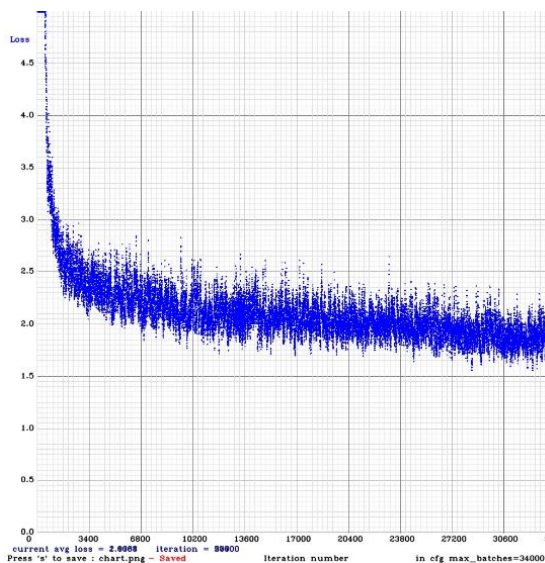

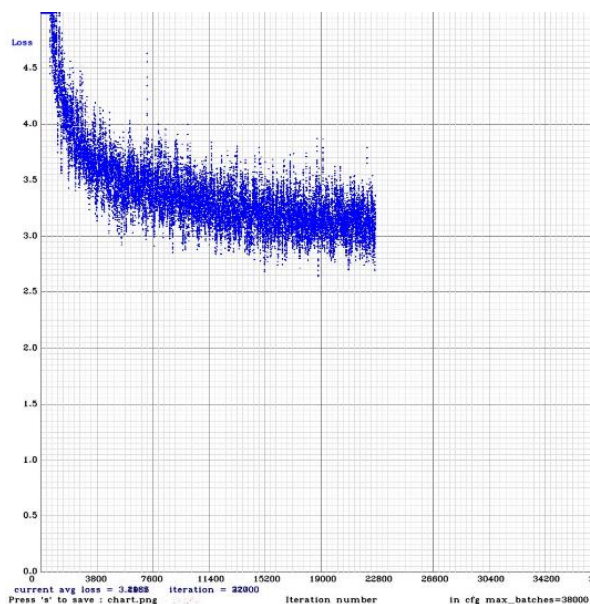**Figure 5.** 17 classes YOLO model Loss Graph

Figure 6. 19 grouped classes YOLO model Loss Graph

Figure. 7 shows sample image output results which suggests a frame with face and the logo of the film is placed dynamically according to faces.

Figure 7. Sample Results from The King of Comedy Movie [29]

IV. CONCULUSION

The aim of this study is to create cover photos of a movie or TV series using image processing and convolutional neural networks. A successful emotion model was used in this study. In addition, in order to reduce the results, a face recognition model was developed at each step and thumbnails with the most repeating faces were selected. With the developed method, dynamic results were obtained for users on online watch platforms. Selected frames were labeled and features that might attract the attention of the user were determined. In subsequent studies, obtaining detailed information about faces, determining the movement at the scene (running, walking, swimming), thus interpreting the photographic composition in the frame. It is aimed to create a different data set to improve the performance of the models. Thus, more accurate results can be obtained by working with different data sets.

REFERENCES

[1] "Youtube for Press.", Youtube. Retrieved February 22, 2020 from www.youtube.com/about/press/

[2] Julia Stoll. (2021). Retrieved February 22, 2021 from https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/#:~:text=Netflix%20had%20203.67%20million%20paid,Netflix's%20total%20global%20subscriber%20base.

[3] Digital Commerce (2021). Retrieved February 15, from https://www.digitalcommerce360.com/article/amazon-prime-membership/#:~:text=Amazon.com%20Inc.%20

0has%20added,Intelligence%20Research%20Partners%20(CIRP).

[4] Hulu (2021), Retrieved January 30, from https://www.businessofapps.com/data/hulu-statistics/

[5] Zeng, X., Li, W., Zhang, X., & Xu, B. (2008, June). Key-frame extraction using dominant-set clustering. In *2008 IEEE international conference on multimedia and expo* (pp. 1285-1288). IEEE.

[6] De Avila, S. E. F., Lopes, A. P. B., da Luz Jr, A., & de Albuquerque Araújo, A. (2011). VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recognition Letters*, *32*(1), 56-68.

[7] Zhuang, Y., Rui, Y., Huang, T. S., & Mehrotra, S. (1998, October). Adaptive key frame extraction using unsupervised clustering. In *Proceedings 1998 international conference on image processing. icip98 (cat. no. 98cb36269)* (Vol. 1, pp. 866-870). IEEE.

[8] Sujatha, C., & Mudenagudi, U. (2011, October). A study on keyframe extraction methods for video summary. In *2011 International Conference on Computational Intelligence and Communication Networks* (pp. 73-77). IEEE.

[9] Gharbi, H., Bahroun, S., & Zagrouba, E. (2019). Key frame extraction for video summarization using local description and repeatability graph clustering. *Signal, Image and Video Processing*, *13*(3), 507-515.

[10] Deng, Y., Loy, C. C., & Tang, X. (2017). Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, *34*(4), 80-106.

[11] Ma, S., Liu, J., & Wen Chen, C. (2017). A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4535-4544).

[12] Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006, May). Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision* (pp. 288-301). Springer, Berlin, Heidelberg.

[13] Riley, M., Machado, L., Roussabrov, B., Branyen, T., Bhawalkar, P., Jin, E., & Kansara, A. (2018). AVA: The Art and Science of Image Discovery at Netflix. *The Netflix Tech Blog*.

[14] Tsao, C. N., Lou, J. K., & Chen, H. H. (2019, March). Thumbnail image selection for VOD services. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 54-59). IEEE.

[15] Yu, Z., & Shi, N. (2020). A Multi-modal Deep Learning Model for Video Thumbnail Selection. arXiv preprint arXiv:2101.00073.

[16] Huang, Y. Y., Kuo, T. Y., & Chen, H. H. (2020, April). Selecting Representative Thumbnail Image and Video Clip from a Video via Bullet Screen. In *Companion Proceedings of the Web Conference 2020* (pp. 48-49).

[17] Pretorious, K., & Pillay, N. (2020, July). A Comparative Study of Classifiers for Thumbnail Selection. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[18] Pech-Pacheco, J. L., Cristóbal, G., Chamorro-Martinez, J., & Fernández-Valdivia, J. (2000, September). Diatom autofocusing in brightfield microscopy: a comparative study. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 3, pp. 314-317). IEEE.

[19] Cech, J., & Soukupova, T. (2016). Real-time eye blink detection using facial landmarks. *Cent. Mach. Perception, Dep.*

Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague, 1-8.

[20] Likas, A., & Vlassis, N. (2003). The global k-means clustering algorithm, Pattern Recognit.

[21] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.

[22] Carrier, P.-L., Courville, A., Goodfellow, I. J., Mirza, M., & Bengio, Y. (2013). FER-2013 Face Database. Technical report, 1365. Universit'e de Montr'eal.

[23] Çakar, M., Yıldız, K., & Demir, Ö. (2020, October). Creating Cover Photos (Thumbnail) for Movies and TV Series with Convolutional Neural Network. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-5). IEEE.

[24] J. Jayalekshmi and T. Mathew, "Facial expression recognition and emotion classification system for sentiment analysis," *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, Thiruvanthapuram, 2017, pp. 1-8, doi: 10.1109/NETACT.2017.8076732.

[25] Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

[26] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[27] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., ... & Ferrari, V. (2020). The open images dataset v4. *International Journal of Computer Vision*, 1-26.

[28] King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, *10*, 1755-1758.

[29] The King of comedy (1983, February 18). Retrieved March 15,2021, from https://www.imdb.com/title/tt0085794/