MALATYA TURGUT ÖZAL UNIVERSITY

NATURENGS

MTU Journal of Engineering and Natural Sciences

Volume: 3 Issue: 1 - June 2022

https://dergipark.org.tr/tr/pub/naturengs

www.naturengs.com

MALATYA TURGUT ÖZAL UNIVERSITY

NATURENGS

MTU Journal of Engineering and Natural Sciences

Volume: 3 / Issue: 1 / June - 2022

We are delighted to present the second volume of the journal NATURENGS owned by Malatya Turgut Özal University. MTU Journal of Engineering and Natural Sciences – NATURENGS is a double- blind peer-reviewed, open-access international journal which will publish electronically two times in a year by the Malatya Turgut Özal University from June 2020.

We set out with the desire to create an environment where scientific and/or technological studies carried out in universities, industry and other research institutions will be shared. We aim to advance by giving priority to studies involving scientific and / or technological originality.

Manuscripts submitted for publication are analyzed in terms of scientific quality, ethics and research methods in terms of its compliance by the Editorial Board representatives of the relevant areas. Then, the abstracts of the appropriate articles are sent to at least two different referees with a well-known in scientific area. If the referees agree to review the article, full text in the framework of the privacy protocol is sent. By the decisions of referees, either directly or corrected article is published or rejected. Confidential reports of the referees in the journal archive will be retained for ten years. All post-evaluation process is done electronically on the internet.

In the journal's publication policy, we would like to state that we will not compromise on quality. In this process, we know that we have undertaken important tasks, especially the selection of referees and monitoring of evaluations. Our journal is indexed in *Index Copernicus, CiteFactor, Google Scholar, Scientific Indexing Services, ASOS and ESJI* international databases. We will work with the devotion to get our journal into the TR-Index and then the Science Citation Index database as soon as possible.

We would like to thanks our Rector, Prof. Dr. Aysun Bay KARABULUT, who encouraged and supported the establishment of our journal. In addition, we would like to thank all the Authors and Referees who contributed to this issue.

Assist. Prof. Aydan AKSOĞAN KORKMAZ
On behalf of the Editorial Board

MALATYA TURGUT ÖZAL UNIVERSITY

NATURENGS

MTU Journal of Engineering and Natural Sciences

Volume: 3 / Issue: 1 / June- 2022

ISSN: 2717-8013

Contact Information

MTU Journal of Engineering and Natural Sciences – NATURENGS,

Malatya Turgut Özal University, 44210, Battalgazi/Malatya, TURKEY

Phone: +90-422 846 12 55, Fax: +90-422 846 12 25,

e-mail: naturengs@ozal.edu.tr

web: https://dergipark.org.tr/tr/pub/naturengs

CONTENTS

NATURENGS
MTU Journal of Engineering and Natural Sciences
https://dergipark.org.tr/tr/pub/naturengs
DOI: 10.46572/naturengs.1082785

Research Article

Pluriharmonic Conformal Bi-Slant Riemannian Maps

Şener YANAN[1*]

[1]Department of Mathematics, Faculty of Arts and Science, Adıyaman University, Adıyaman, Türkiye.

(**Received:** 04.03.2022; **Accepted:** 20.05.2022)

ABSTRACT: In this study, the notion of pluriharmonic map was applied to conformal bi-slant Riemannian maps from a Kaehler manifold to a Riemannian manifold to examine its geometric properties. Such relations between pluriharmonic map, horizontally homothetic map, and totally geodesic map were obtained.

Keywords: Riemannian map, Conformal Riemannian map, Conformal bi-slant Riemannian map, Pluriharmonic map.

1. INTRODUCTION

The notion of submersion was introduced by O'Neill [1] and Gray [2]. Submersion theory between almost Hermitian manifolds was studied by Watson [3]. Then, Fischer studied the theory of submersion in various types and generalized it to Riemannian maps [4]. Riemannian maps between Riemannian manifolds generalize isometric immersions and Riemannian submersions. Let $\Phi: (\mathrm{M}_1, \mathrm{g}_1) \to (\mathrm{M}_2, \mathrm{g}_2)$ be a smooth map between Riemannian manifolds such that $0 < rank\Phi < min\{dim(M_1), dim(M_2)\}$. Then, the tangent bundle of TM_1 of M_1 has the following decomposition:

$$TM_1 = ker\Phi_* \oplus (ker\Phi_*)^{\perp}.$$

Since $rank\Phi < min\{dim(M_1), dim(M_2)\}$, we have $(range\Phi_*)^{\perp}$. Hence, the tangent bundle of TM_2 of M_2 has the following decomposition:

$$TM_2 = range\Phi_* \oplus (range\Phi_*)^{\perp}.$$

A smooth map $\Phi: (\mathrm{M}_1^m, \mathrm{g}_1) \to (\mathrm{M}_2^m, \mathrm{g}_2)$ is called Riemannian map at $p_1 \in M_1$ if the horizontal restriction $\Phi^h_{*p_1}: \left(ker\Phi_{*p_1}\right)^{\perp} \longrightarrow (range\Phi_*)$ is a linear isometry. Therefore, the Riemannian map satisfies the equation

$$g_1(X,Y) = g_2\left(\Phi_*(X), \Phi_*(Y)\right)$$

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$. Hence, isometric immersions and Riemannian submersions are particular Riemannian maps, respectively, with $\ker\Phi_* = \{0\}$ and $(\mathrm{range}\Phi_*)^{\perp} = \{0\}$ [4]. Moreover, Şahin and Yanan examined conformal Riemannian maps [5-8], see also [9]. We say

*Corresponding Author: syanan@adiyaman.edu.tr
ORCID number of author: 0000-0003-1600-6522

that $\Phi: (\mathrm{M}^{\mathrm{m}}, \mathrm{g}_{\mathrm{M}}) \longrightarrow (\mathrm{N}^{\mathrm{n}}, \mathrm{g}_{\mathrm{N}})$ is a conformal Riemannian map at $p \in M$ if $0 < rank\Phi_{*p} \leq min\{m, n\}$ and Φ_* maps the horizontal space $\left(\ker(\Phi_{*\mathrm{p}})^{\perp}\right)$ conformally onto $range(\Phi_{*p})$, i.e., there exists a number $\lambda^2(\mathrm{p}) \neq 0$ such that

$$g_N\left(\Phi_{*p}(X), \Phi_{*p}(Y)\right) = \lambda^2(p) g_M(X, Y)$$

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$. Also, Φ is called conformal Riemannian if Φ is conformal Riemannian at each $p \in M$. Here, λ is the dilation of Φ at a point $p \in M$ and it is a continuous function as $\lambda: M \to [0, \infty)$ [10]. One can see more research on curvature relations for conformal bi-slant submersions and the relation between submersion theory and bi-slant structure, which is studied by Aykurt Sepet [11,12].

An even-dimensional Riemannian manifold (M, g_M, J) is called an almost Hermitian manifold if there exists a tensor field J of type $(1,1)$ on M such that $J^2 = -I$ where I denotes the identity transformation of TM and

$$g_M(X, Y) = g_M(JX, JY), \forall X, Y \in \Gamma(TM).$$

Let (M, g_M, J) be an almost Hermitian manifold and its Levi-Civita connection ∇ concerning g_M. If J is parallel concerning ∇, i.e.

$$(\nabla_X J)Y = 0,$$

we say M is a Kaehler manifold [13].

Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(X, Y) + (\nabla\Phi_*)(JX, JY) = 0$$

for $X, Y \in \Gamma(TM)$ [14].

Here, we recall some basic definitions of conformal Riemannian maps from a Kaehler manifold to a Riemannian manifold.

Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a conformal Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N).

1. If the map Φ satisfies the following condition:

   $$J(ker\Phi_*) \subset (ker\Phi_*)^{\perp},$$

   Then Φ is called a conformal anti-invariant Riemannian map [6].

2. If the following conditions are satisfying:
   i. There exists a subbundle of $ker\Phi_*$ such that $J(D_1) = D_1$,
   ii. There exists a complementary subbundle D_2 to D_1 in $ker\Phi_*$ such that $J(D_2) \subset (ker\Phi_*)^{\perp}$,

   We say that Φ is a conformal semi-invariant Riemannian map [7].

3. If for any non-zero vector $X \in \Gamma(ker\Phi_*)$ at a point $p \in M$; the angle $\theta(X)$ between the space $ker\Phi_*$ and JX is a constant, i.e. it is independent of the choice of the tangent vector $X \in \Gamma(ker\Phi_*)$ and the choice of the point $p \in M$, we say that Φ is a conformal slant Riemannian map. In this case, the angle θ is called the slant angle of the conformal slant Riemannian map [8].
4. If the vertical distribution $ker\Phi_*$ of Φ admits two orthogonal complementary distributions D_θ and $D_\perp$ such that D_θ is slant and $D_\perp$ is anti-invariant, i.e., we have

$$ker\Phi_* = D_\theta \oplus D_\perp.$$

Hence, Φ is called a conformal hemi-slant Riemannian map and the angel θ is called hemi-slant angle of the conformal Riemannian map [15].

5. At last, Φ is called a conformal semi-slant Riemannian map if there is a distribution $D_1 \subset ker\Phi_*$ such that

$$ker\Phi_* = D_1 \oplus D_2, J(D_1) = D_1$$

and the angle $\theta = \theta(X)$ between JX and the space $(D_2)_p$ is constant for nonzero $X \in (D_2)_p$ and $p \in M$, where D_2 is the orthogonal complement distribution of D_1 in $ker\Phi_*$. The angel θ is called semi-slant angle of the map [16].

Therefore, we define conformal bi-slant Riemannian maps from a Kaehler manifold to a Riemannian manifold. Some geometric properties of conformal bi-slant Riemannian maps are examined via pluriharmonic map.

2. MATERIAL AND METHODS

This section gives several definitions and results for the study for conformal bi-slant Riemannian maps. Let $\Phi: (M, g_M) \longrightarrow (N, g_N)$ be a smooth map between Riemannian manifolds. The second fundamental form of Φ is defined by

$$(\nabla\Phi_*)(X, Y) = \nabla^{\Phi}_X \Phi_*(Y) - \Phi_*(\nabla_X Y)$$

for $X, Y \in \Gamma(TM)$. The second fundamental form $(\nabla\Phi_*)$ is symmetric. Note that Φ is said to be totally geodesic map if $(\nabla F_*)(X, Y) = 0$ for all $X, Y \in \Gamma(TM)$ [17]. Here, we define O'Neill's tensor fields $\mathcal{T}$ and $\mathcal{A}$ as

$$\mathcal{A}_X Y = h\nabla_{hX} vY + v\nabla_{hX} hY,$$

$$\mathcal{T}_X Y = h\nabla_{vX} vY + v\nabla_{vX} hY$$

for $X, Y \in \Gamma(TM)$ with the Levi-Civita connection ∇ of g_M. Here, we denote by v and h the projections on the vertical distribution $ker\Phi_*$ and the horizontal distribution $(ker\Phi_*)^\perp$, respectively. For any $X \in \Gamma(TM)$, $\mathcal{T}_X$ and $\mathcal{A}_X$ are skew-symmetric operators on $(\Gamma(TM), g)$ reversing the horizontal and the vertical distributions. Also, $\mathcal{T}$ is vertical, $\mathcal{T}_X = \mathcal{T}_{vX}$ and $\mathcal{A}$ is horizontal, $\mathcal{A}_X = \mathcal{A}_{hX}$. Note that the tensor field $\mathcal{T}$ is symmetric on the vertical distribution [1]. In addition, by definitions of O'Neill's tensor fields, we have

$$\nabla_U V = \mathcal{T}_U V + v\nabla_U V,$$

$$\nabla_U X = h\nabla_U X + \mathcal{T}_U X,$$

$$\nabla_X V = \mathcal{A}_X V + v\nabla_X V,$$

$$\nabla_X Y = h\nabla_X Y + \mathcal{A}_X Y$$

for $X, Y \in \Gamma((ker\Phi_*)^\perp)$ and $U, V \in \Gamma(ker\Phi_*)$ [18].

If a vector field X on M is related to a vector field X' on N, we say X is a projectable vector field. If X is both a horizontal and a projectable vector field, we say X is a basic vector field on M [19]. When we mention a horizontal vector field throughout this study, we always consider a basic vector field. On the other hand, let $\Phi: (M^m, g_M) \longrightarrow (N^n, g_N)$ be a conformal Riemannian map between Riemannian manifolds. Then, we have

$$(\nabla\Phi_*)(X,Y)\,|_{range\Phi_*} = X(\ln\lambda)\Phi_*(Y) + Y(\ln\lambda)\Phi_*(X) - g_M(X,Y)\Phi_*\big(grad(\ln\lambda)\big)$$

where $X, Y \in \Gamma((ker\Phi_*)^\perp)$ [10]. Hence, we obtain $\nabla^\Phi_X \Phi_*(Y)$ as

$$\nabla^\Phi_X \Phi_*(Y) = \Phi_*(h\nabla_X Y) + X(\ln\lambda)\Phi_*(Y) + Y(\ln\lambda)\Phi_*(X) - g_M(X,Y)\Phi_*\big(grad(\ln\lambda)\big) + (\nabla\Phi_*)^\perp(X,Y)$$

where $(\nabla\Phi_*)^\perp(X,Y)$ is the component of $(\nabla\Phi_*)(X,Y)$ on $(range\Phi_*)^\perp$ for $X, Y \in \Gamma((ker\Phi_*)^\perp)$ [6].

3. RESULTS AND DISCUSSION

In this section, we define conformal bi-slant Riemannian maps, give their decomposition and study some theorems for conformal bi-slant Riemannian maps by applying the notion of pluriharmonic map on certain distributions. Therefore, we want to obtain relations among geometric structures.

Definition 3.1. Let (M, g_M, J) be a Kaehler manifold and (N, g_N) be a Riemannian manifold. Then, a conformal Riemannian map $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ is called a conformal bi-slant Riemannian map if and only if D_1 and D_2 are slant distributions with their slant angles θ_1 and θ_2, respectively, such that

$$ker\Phi_* = D_1 \oplus D_2.$$

Here, if the slant angles satisfy that $\theta_1, \theta_2 \neq 0, \frac{\pi}{2}$, Φ is called a proper conformal bi-slant Riemannian map [20].

We explain decompositions of distributions for the conformal bi-slant Riemannian map Φ.

Suppose that Φ is a conformal bi-slant Riemannian map from a Kaehler manifold (M, g_M, J) to a Riemannian manifold (N, g_N). For any $U \in \Gamma(ker\Phi_*)$, we have

$$U = PU + QU,$$

where $PU \in \Gamma(D_1)$ and $QU \in \Gamma(D_2)$. On the other hand, we have

$$JU = \psi U + \phi U,$$

for $U \in \Gamma(ker\Phi_*)$ where $\psi U \in \Gamma(ker\Phi_*)$ and $\phi U \in \Gamma((ker\Phi_*)^{\perp})$. Also, for any $X \in \Gamma((ker\Phi_*)^{\perp})$, we write

$$JX = BX + CX,$$

where $BX \in \Gamma(ker\Phi_*)$ and $CX \in \Gamma((ker\Phi_*)^{\perp})$. Therefore, the horizontal distribution $(ker\Phi_*)^{\perp}$ can be decomposed as

$$(ker\Phi_*)^{\perp} = \phi D_1 \oplus \phi D_2 \oplus \mu,$$

where μ is the orthogonal complementary distribution of $\phi D_1 \oplus \phi D_2$ in $(ker\Phi_*)^{\perp}$ [20].

We have the following theorem same as conformal bi-slant submersions.

Theorem 3.2. Let Φ be a conformal bi-slant Riemannian map from an almost Hermitian manifold (M, g_M, J) to a Riemannian manifold (N, g_N) with slant angles θ_1 and θ_2. Then, we have

$$\psi^2 U_i = -(cos^2\,\theta_i)U_i$$

for $U_i \in \Gamma(D_i)$, $i = 1,2$ [12].

Recall that, Φ is said to be a horizontally homothetic map if $h\big(grad(ln\,\lambda)\big) = 0$. It means that horizontal part of the gradient vector field of the dilation λ is equal to zero [19]. On the other hand, Φ is said to be totally geodesic map if $(\nabla\Phi_*)(E,F) = 0$ for $E, F \in \Gamma(TM)$ [5].

Firstly, we derive new notions by using pluriharmonic map, see [7,8]. Hence, let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a D_1- pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(U_1, V_1) + (\nabla\Phi_*)(JU_1, JV_1) = 0$$

for $U_1, V_1 \in \Gamma(D_1)$.

Theorem 3.3. Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a conformal bi-slant Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N). If Φ is a D_1- pluriharmonic map, then the following two assertions imply the third assertion,

i. D_1 defines a totally geodesic foliation on M,
ii. The map Φ is a horizontally homothetic map and $(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) = 0$,
iii. $h\nabla_{U_1}\phi\psi V_1 + \phi\mathcal{T}_{U_1}\phi V_1 + Ch\nabla_{U_1}\phi V_1 = \mathcal{T}_{\psi U_1}\psi V_1 + \mathcal{A}_{\phi U_1}\psi V_1 + \mathcal{A}_{\phi V_1}\psi U_1$

for $U_1, V_1 \in \Gamma(D_1)$.

Proof. Firstly, we know that Φ is a D_1- pluriharmonic map, then we have

$$(\nabla\Phi_*)(U_1, V_1) + (\nabla\Phi_*)(JU_1, JV_1) = 0 \tag{1}$$

for $U_1, V_1 \in \Gamma(D_1)$. Since M is a Kaehler manifold by using the notion of second fundamental form of a map and its restriction to the horizontal distribution, we get

$$0 = (\nabla\Phi_*)(U_1, V_1) + (\nabla\Phi_*)(JU_1, JV_1) \tag{2}$$

$$\begin{aligned} 0 = &-\Phi_*\big(\nabla_{U_1}V_1\big) - \Phi_*\big(\nabla_{\psi U_1}\psi V_1 + \nabla_{\phi U_1}\psi V_1 + \nabla_{\phi V_1}\psi U_1\big) \\ &+(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) + \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) \\ &+\phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big) \end{aligned} \tag{3}$$

for $U_1, V_1 \in \Gamma(D_1)$. Then, from O'Neill's tensor fields by using Eq. (3), we get

$$\begin{aligned} 0 = &\Phi_*\big(J\nabla_{U_1}\psi V_1 + \nabla_{U_1}\phi\psi V_1\big) - \Phi_*\big(\mathcal{T}_{\psi U_1}\psi V_1 + \mathcal{A}_{\phi U_1}\psi V_1 + \mathcal{A}_{\phi V_1}\psi U_1\big) \\ &+(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) + \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) \\ &+\phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big) \end{aligned} \tag{4}$$

$$\begin{aligned} 0 = &\Phi_*\big(\nabla_{U_1}\psi^2 V_1 + \nabla_{U_1}\phi\psi V_1\big) + \Phi_*\big(J\mathcal{T}_{U_1}\phi V_1 + Jh\nabla_{U_1}\phi V_1\big) \\ &-\Phi_*\big(\mathcal{T}_{\psi U_1}\psi V_1 + \mathcal{A}_{\phi U_1}\psi V_1 + \mathcal{A}_{\phi V_1}\psi U_1\big) \\ &+(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) + \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) \\ &+\phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big). \end{aligned} \tag{5}$$

From Theorem 3.2. in Eq. (5), we obtain

$$\begin{aligned} cos^2\,\theta_1\,\Phi_*\big(\nabla_{U_1}V_1\big) = &\Phi_*\big(h\nabla_{U_1}\phi\psi V_1\big) + \Phi_*\big(\phi\mathcal{T}_{U_1}\phi V_1 + Ch\nabla_{U_1}\phi V_1\big) \\ &-\Phi_*\big(\mathcal{T}_{\psi U_1}\psi V_1 + \mathcal{A}_{\phi U_1}\psi V_1 + \mathcal{A}_{\phi V_1}\psi U_1\big) \\ &+(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) + \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) \\ &+\phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big). \end{aligned} \tag{6}$$

Now, consider that i. and ii. are satisfied in Eq. (6). Since D_1 defines a totally geodesic foliation on M and Φ is a horizontally homothetic map, we have $\Phi_*\big(\nabla_{U_1}V_1\big) = 0$, $(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) = 0$ and $\phi U_1(ln\,\lambda)\Phi_*(\phi V_1) + \phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big) = 0$ for $U_1, V_1 \in \Gamma(D_1)$, respectively. Hence, one can clearly see the proof of iii. from Eq. (6). If ii. and iii. are satisfied in Eq. (6), we get

$$cos^2\,\theta_1\,\Phi_*\big(\nabla_{U_1}V_1\big) = 0. \tag{7}$$

So, easily we say that D_1 defines a totally geodesic foliation on M for $U_1, V_1 \in \Gamma(D_1)$. The proof of i. is completed. Suppose that i. and iii. are satisfied in Eq. (6), we obtain

$$\begin{aligned} 0 = &(\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1) + \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) + \phi V_1(ln\,\lambda)\Phi_*(\phi U_1) \\ &-g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big). \end{aligned} \tag{8}$$

In Eq. (8), if we separate components as to which one belongs to $range\Phi_*$ or its orthogonal complement distribution $(range\Phi_*)^{\perp}$, we obtain $0 = (\nabla\Phi_*)^{\perp}(\phi U_1, \phi V_1)$. Hence, we get

$$0 = \phi U_1(ln\,\lambda)\Phi_*(\phi V_1) + \phi V_1(ln\,\lambda)\Phi_*(\phi U_1) - g_M(\phi U_1, \phi V_1)\Phi_*\big(grad(ln\,\lambda)\big). \quad (9)$$

For $\phi U_1 \in \Gamma(\phi D_1)$, since Φ is a conformal map, we get from Eq. (9),

$$\begin{aligned} 0 = {} & \phi U_1(ln\,\lambda) g_N\big(\Phi_*(\phi V_1), \Phi_*(\phi U_1)\big) \\ & + \phi V_1(ln\,\lambda) g_N\big(\Phi_*(\phi U_1), \Phi_*(\phi U_1)\big) \\ & - g_M(\phi U_1, \phi V_1) g_N\Big(\Phi_*\big(grad(ln\,\lambda)\big), \Phi_*(\phi U_1)\Big) \end{aligned} \quad (10)$$

$$\begin{aligned} 0 = {} & \lambda^2 \phi U_1(ln\,\lambda) g_M(\phi V_1, \phi U_1) \\ & + \lambda^2 \phi V_1(ln\,\lambda) g_M(\phi U_1, \phi U_1) \\ & - \lambda^2\, g_M(\phi U_1, \phi V_1)\phi U_1(ln\,\lambda) \end{aligned} \quad (11)$$

$$0 = \lambda^2 \phi V_1(ln\,\lambda) g_M(\phi U_1, \phi U_1). \quad (12)$$

In Eq. (12), we have $\phi V_1(ln\,\lambda) = 0$. This means, the dilation λ is a constant on ϕD_1. On the other hand, if we take $U_1 = V_1$, $\phi U_2 \in \Gamma(\phi D_2)$ and $U_3 \in \Gamma(\mu)$ from Eq. (9), we get

$$0 = -\lambda^2 \phi U_2(ln\,\lambda) g_M(\phi U_1, \phi U_1), \quad (13)$$

$$0 = -\lambda^2 U_3(ln\,\lambda) g_M(\phi U_1, \phi U_1), \quad (14)$$

respectively. From Eq. (13) and Eq. (14), we get $\phi U_2(ln\,\lambda) = 0$ and $U_3(ln\,\lambda) = 0$, respectively. Hence, the dilation λ is a constant on ϕD_2 and μ. Therefore, the map Φ is a horizontally homothetic map. iii. is satisfied. The proof is completed.

Similarly, we have the following notion and theorem.

Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a D_2- pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(U_2, V_2) + (\nabla\Phi_*)(JU_2, JV_2) = 0$$

for $U_2, V_2 \in \Gamma(D_2)$.

Theorem 3.4. Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a conformal bi-slant Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N). If Φ is a D_2- pluriharmonic map, then the following two assertions imply the third assertion,

i. D_2 defines a totally geodesic foliation on M,
ii. The map Φ is a horizontally homothetic map and $(\nabla\Phi_*)^{\perp}(\phi U_2, \phi V_2) = 0$,
iii. $h\nabla_{U_2}\phi\psi V_2 + \phi\mathcal{T}_{U_2}\phi V_2 + Ch\nabla_{U_2}\phi V_2 = \mathcal{T}_{\psi U_2}\psi V_2 + \mathcal{A}_{\phi U_2}\psi V_2 + \mathcal{A}_{\phi V_2}\psi U_2$

for $U_2, V_2 \in \Gamma(D_2)$.

Proof. The proof of the Theorem 3.4. can get similarly with Theorem 3.3.

Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a $(ker\Phi_*)^{\perp}$- pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(X,Y) + (\nabla\Phi_*)(JX,JY) = 0$$

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$.

Theorem 3.5. Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a conformal bi-slant Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N). If Φ is a $(ker\Phi_*)^{\perp}$-pluriharmonic map, then one of the following assertions implies the other assertion,

i. $\mathcal{T}_{\mathcal{B}X}BY + \mathcal{A}_{\mathcal{C}X}BY + \mathcal{A}_{\mathcal{C}Y}BX = 0$,
ii. The map Φ is a horizontally homothetic map and $(\nabla\Phi_*)^{\perp}(X,Y) + (\nabla\Phi_*)^{\perp}(CX,CY) = 0$,

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$.

Proof. If Φ is a $(ker\Phi_*)^{\perp}$- pluriharmonic map, we have

$$(\nabla\Phi_*)(X,Y) + (\nabla\Phi_*)(JX,JY) = 0 \tag{15}$$

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$. Then, by using definition of second fundamental form of a map and its decomposition onto $range\Phi_*$ and $(range\Phi_*)^{\perp}$ in Eq. (15), we obtain

$$\begin{aligned} 0 = {} & (\nabla\Phi_*)^{\perp}(X,Y) + (\nabla\Phi_*)^{\perp}(CX,CY) - \Phi_*(\mathcal{T}_{\mathcal{B}X}BY + \mathcal{A}_{\mathcal{C}X}BY + \mathcal{A}_{\mathcal{C}Y}BX) \\ & + X(ln\,\lambda)\Phi_*(Y) + Y(ln\,\lambda)\Phi_*(X) - g_M(X,Y)\Phi_*\big(grad(ln\,\lambda)\big) \\ & + CX(ln\,\lambda)\Phi_*(CY) + CY(ln\,\lambda)\Phi_*(CX) - g_M(CX,CY)\Phi_*\big(grad(ln\,\lambda)\big) \end{aligned} \tag{16}$$

for $X, Y \in \Gamma((ker\Phi_*)^{\perp})$. If i. is satisfied in Eq. (16), we have $\mathcal{T}_{\mathcal{B}X}BY + \mathcal{A}_{\mathcal{C}X}BY + \mathcal{A}_{\mathcal{C}Y}BX = 0$. So, we get from Eq. (16),

$$\begin{aligned} 0 = {} & (\nabla\Phi_*)^{\perp}(X,Y) + (\nabla\Phi_*)^{\perp}(CX,CY) \\ & + X(ln\,\lambda)\Phi_*(Y) + Y(ln\,\lambda)\Phi_*(X) \\ & - g_M(X,Y)\Phi_*\big(grad(ln\,\lambda)\big) \\ & + CX(ln\,\lambda)\Phi_*(CY) + CY(ln\,\lambda)\Phi_*(CX) \\ & - g_M(CX,CY)\Phi_*\big(grad(ln\,\lambda)\big). \end{aligned} \tag{17}$$

In Eq. (17), we know that $0 = (\nabla\Phi_*)^{\perp}(X,Y) + (\nabla\Phi_*)^{\perp}(CX,CY)$ since they belong to $(range\Phi_*)^{\perp}$. On the other hand, from elements of $range\Phi_*$ we examine horizontally homotheticness of the map. Hence, from Eq. (17) by using conformality of the map we have

$$\begin{aligned} 0 = {} & 2X(ln\,\lambda)g_N\big(\Phi_*(Y),\Phi_*(X)\big) + 2Y(ln\,\lambda)g_N\big(\Phi_*(X),\Phi_*(X)\big) \\ & - 2g_M(X,Y)g_N\Big(\Phi_*\big(grad(ln\,\lambda)\big),\Phi_*(X)\Big) \end{aligned} \tag{18}$$

$$0 = 2\lambda^2 Y(ln\,\lambda) g_M(X,X) \tag{19}$$

for $X = CX$ and $Y = CY$. Here, since $\lambda^2 \neq 0$ and $g_M(X,X) \neq 0$, we get $Y(ln\,\lambda) = 0$. It means that λ is a constant on horizontal distribution $(ker\Phi_*)^{\perp}$. Hence, the map Φ is a horizontally homothetic map. ii. is satisfied. If ii. is satisfied in Eq. (16), we have

$$0 = X(ln\,\lambda)\Phi_*(Y) + Y(ln\,\lambda)\Phi_*(X) - g_M(X,Y)\Phi_*\big(grad(ln\,\lambda)\big)$$
$$+CX(ln\,\lambda)\Phi_*(CY) + CY(ln\,\lambda)\Phi_*(CX) - g_M(CX,CY)\Phi_*\big(grad(ln\,\lambda)\big)$$

and

$$0 = (\nabla\Phi_*)^{\perp}(X,Y) + (\nabla\Phi_*)^{\perp}(CX,CY).$$

So, from Eq. (16), we obtain

$$0 = -\Phi_*(\mathcal{T}_{BX}BY + \mathcal{A}_{CX}BY + \mathcal{A}_{CY}BX). \tag{20}$$

Hence, Eq. (20) shows us that i. is satisfied. The proof is completed.

Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a $ker\Phi_*$ - pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(U,V)+(\nabla\Phi_*)(JU,JV) = 0$$

for $U, V \in \Gamma(ker\Phi_*)$.

Theorem 3.6. Let $\Phi: (M, g_M, J) \longrightarrow (N, g_N)$ be a conformal bi-slant Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N). If Φ is a $ker\Phi_*$-pluriharmonic map, then the following two assertions imply the third assertion,

i. $ker\Phi_*$ defines a totally geodesic foliation on M,
ii. $\mathcal{A}_{\phi V}\psi PU + \mathcal{A}_{\phi U}\psi V + \mathcal{T}_{\psi U}\psi V + h\nabla_{\psi QU}\phi V = \phi\mathcal{T}_U\phi V + Ch\nabla_U\phi V + h\nabla_U\phi\psi V$,
iii. The map Φ is a horizontally homothetic map and $(\nabla\Phi_*)^{\perp}(\phi U, \phi V) = 0$

for $U, V \in \Gamma(ker\Phi_*)$.

Proof. If Φ is a $ker\Phi_*$- pluriharmonic map, we have

$$(\nabla\Phi_*)(U,V)+(\nabla\Phi_*)(JU,JV) = 0 \tag{21}$$

for $U, V \in \Gamma(ker\Phi_*)$. By using decomposition theorems for conformal bi-slant Riemannian maps in Eq. (21), we get

$$0 = (\nabla\Phi_*)(U,V) + (\nabla\Phi_*)(J(PU + QU), \phi V + \psi V)$$
$$0 = (\nabla\Phi_*)(\psi PU, \psi V) + (\nabla\Phi_*)(\phi V, \psi PU)$$
$$+(\nabla\Phi_*)(\phi U, \psi V) + (\nabla\Phi_*)(\psi QU, \phi V + \psi V)$$
$$+(\nabla\Phi_*)(\phi U, \phi V) + (\nabla\Phi_*)(U,V). \tag{22}$$

Then, from definition of the second fundamental form of a map in Eq. (22), we get

$$0 = -\Phi_*\big(\mathcal{T}_{\psi\mathcal{U}}\psi V + \mathcal{A}_{\phi\mathcal{V}}\psi PU + \mathcal{A}_{\phi\mathcal{U}}\psi V + h\nabla_{\psi QU}\phi V\big)$$
$$+(\nabla\Phi_*)^{\perp}(\phi U,\phi V) + (\nabla\Phi_*)^{\top}(\phi U,\phi V)$$
$$+\Phi_*\big(J(\mathcal{T}_{\mathcal{U}}\phi V + h\nabla_U\phi V)\big) + \Phi_*(\nabla_U J\psi V). \quad (23)$$

Now, by using Theorem 3.2. and horizontal restriction of the second fundamental form of a map in Eq. (23), we obtain

$$0 = -\Phi_*\big(\mathcal{T}_{\psi\mathcal{U}}\psi V + \mathcal{A}_{\phi\mathcal{V}}\psi PU + \mathcal{A}_{\phi\mathcal{U}}\psi V + h\nabla_{\psi QU}\phi V\big)$$
$$+(\nabla\Phi_*)^{\perp}(\phi U,\phi V) + \phi U(ln\,\lambda)\Phi_*(\phi V)$$
$$+\phi V(ln\,\lambda)\Phi_*(\phi U) - g_M(\phi U,\phi V)\Phi_*\big(grad(ln\,\lambda)\big)$$
$$+\Phi_*(\phi\mathcal{T}_{\mathcal{U}}\phi V + Ch\nabla_U\phi V) + \Phi_*(-cos^2\,\theta\,\nabla_U V + h\nabla_U\phi\psi V)$$
$$cos^2\,\theta\,\Phi_*(\nabla_U V) = -\Phi_*\big(\mathcal{T}_{\psi\mathcal{U}}\psi V + \mathcal{A}_{\phi\mathcal{V}}\psi PU + \mathcal{A}_{\phi\mathcal{U}}\psi V + h\nabla_{\psi QU}\phi V\big)$$
$$+(\nabla\Phi_*)^{\perp}(\phi U,\phi V) + \phi U(ln\,\lambda)\Phi_*(\phi V)$$
$$+\phi V(ln\,\lambda)\Phi_*(\phi U) - g_M(\phi U,\phi V)\Phi_*\big(grad(ln\,\lambda)\big)$$
$$+\Phi_*(\phi\mathcal{T}_{\mathcal{U}}\phi V + Ch\nabla_U\phi V + h\nabla_U\phi\psi V). \quad (24)$$

In Eq. (24), if i. and ii. are satisfied we have $\nabla_U V = 0$ and $\mathcal{A}_{\phi\mathcal{V}}\psi PU + \mathcal{A}_{\phi\mathcal{U}}\psi V + \mathcal{T}_{\psi\mathcal{U}}\psi V + h\nabla_{\psi QU}\phi V = \phi\mathcal{T}_{\mathcal{U}}\phi V + Ch\nabla_U\phi V + h\nabla_U\phi\psi V$, respectively. So, we get

$$0 = (\nabla\Phi_*)^{\perp}(\phi U,\phi V) + \phi U(ln\,\lambda)\Phi_*(\phi V)$$
$$+\phi V(ln\,\lambda)\Phi_*(\phi U) - g_M(\phi U,\phi V)\Phi_*\big(grad(ln\,\lambda)\big). \quad (25)$$

Similarly, we obtain $(\nabla\Phi_*)^{\perp}(\phi U,\phi V) = 0$, clearly. Hence, Eq. (25) turns into

$$0 = \phi U(ln\,\lambda)\Phi_*(\phi V) + \phi V(ln\,\lambda)\Phi_*(\phi U) - g_M(\phi U,\phi V)\Phi_*\big(grad(ln\,\lambda)\big). \quad (26)$$

For $\phi V \in \Gamma((ker\Phi_*)^{\perp})$, from the conformality of the map we get

$$0 = \phi U(ln\,\lambda)g_N\big(\Phi_*(\phi V),\Phi_*(\phi V)\big) + \Phi V(ln\,\lambda)g_N\big(\Phi_*(\phi U),\Phi_*(\phi V)\big)$$
$$-g_M(\phi U,\phi V)g_N\Big(\Phi_*\big(grad(ln\,\lambda)\big),\Phi_*(\phi V)\Big)$$
$$0 = \lambda^2\phi U(ln\,\lambda)g_M(\phi V,\phi V). \quad (27)$$

In Eq. (27), since $\lambda^2 \neq 0$ and $g_M(\phi V,\phi V) \neq 0$, we get $\phi U(ln\,\lambda) = 0$. It means that λ is a constant on $(ker\Phi_*)^{\perp}$. Hence, the map Φ is a horizontally homothetic map. iii. is proved. The other cases of the proof could be seen clearly. The proof is completed.

Let $\Phi\colon (M, g_M, J) \longrightarrow (N, g_N)$ be a map from a complex manifold (M, g_M, J) to a Riemannian manifold (N, g_N). Then Φ is called a $mixed$ - pluriharmonic map if Φ satisfies the following equation:

$$(\nabla\Phi_*)(X,U)+(\nabla\Phi_*)(JX,JU) = 0$$

for $X \in \Gamma((ker\Phi_*)^{\perp})\ and\ \ U \in \Gamma(ker\Phi_*)$.

Theorem 3.7. Let Φ: $(M, g_M, J) \longrightarrow (N, g_N)$ be a conformal bi-slant Riemannian map between a Kaehler manifold (M, g_M, J) and a Riemannian manifold (N, g_N). If Φ is a $mixed$ - pluriharmonic map, then one of the following assertions imply the other assertion,

i. $\mathcal{A}_X U = -\mathcal{T}_{BX}\psi U - \mathcal{A}_{CX}\psi U - \mathcal{A}_{\phi U}BX$,

ii. The map Φ is a horizontally homothetic map and $(\nabla\Phi_*)^{\perp}(CX, \phi U) = 0$

for $X \in \Gamma((ker\Phi_*)^{\perp})\ and\ \ U \in \Gamma(ker\Phi_*)$.

Proof. If Φ is a $mixed$ - pluriharmonic map, by direct calculations we obtain

$$0 = (\nabla\Phi_*)(X, U)+(\nabla\Phi_*)(JX, JU)$$

$$0 = -\Phi_*(\mathcal{A}_X U) - \Phi_*\big(\mathcal{T}_{BX}\psi U + \mathcal{A}_{CX}\psi U + \mathcal{A}_{\phi U}BX\big)$$
$$+(\nabla\Phi_*)^{\perp}(CX, \phi U) + CX(ln\,\lambda)\Phi_*(\phi U) + \phi U(ln\,\lambda)\Phi_*(CX)$$
$$-g_M(CX, \phi U)\Phi_*\big(grad(ln\,\lambda)\big) \tag{28}$$

for $X \in \Gamma((ker\Phi_*)^{\perp})\ and\ \ U \in \Gamma(ker\Phi_*)$. In Eq. (28), if i. is satisfied we get

$$0 = (\nabla\Phi_*)^{\perp}(CX, \phi U) + CX(ln\,\lambda)\Phi_*(\phi U) + \phi U(ln\,\lambda)\Phi_*(CX)$$
$$-g_M(CX, \phi U)\Phi_*\big(grad(ln\,\lambda)\big). \tag{29}$$

In Eq. (29), we get easily $(\nabla\Phi_*)^{\perp}(CX, \phi U) = 0$. Hence, from Eq. (29) we get

$$0 = CX(ln\,\lambda)\Phi_*(\phi U) + \phi U(ln\,\lambda)\Phi_*(CX) - g_M(CX, \phi U)\Phi_*\big(grad(ln\,\lambda)\big). \tag{30}$$

For $CX, \phi U \in \Gamma((ker\Phi_*)^{\perp})$, from Eq. (30) we obtain

$$0 = \lambda^2\phi U(ln\,\lambda)g_M(CX, CX) \tag{31}$$

and

$$0 = \lambda^2 CX(ln\,\lambda)g_M(\phi U, \phi U), \tag{32}$$

respectively. From Eq. (31) and Eq. (32), we say λ is a constant on horizontal distribution. Hence, the map Φ is a horizontally homothetic map. ii. is proved. The converse of this situation is clear. The proof is completed.

4. CONCLUSIONS

Throughout this study, we obtained geometric relations by using derivations of the notion of pluriharmonic map as $D_1, D_2, (ker\Phi_*)^{\perp}, ker\Phi_*$ and mixed - pluriharmonic map onto conformal bi-slant Riemannian maps.

Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

Author Contribution

Şener Yanan contributed 100% at every stage of the article.

REFERENCES

[1] O'Neill, B. (1966). The fundamental equations of a submersion. *Michigan Math. J.*, 13: 458-469.

[2] Gray, A. (1967). Pseudo - Riemannian almost product manifolds and submersions. *J. Math. Mech.*, 16(7): 715- 737.

[3] Watson, B. (1976). Almost Hermitian submersions. *J. Differ. Geom.*, 11(1): 147-165.

[4] Fischer, A.E. (1992). Riemannian maps between Riemannian manifolds. *Contemp. Math.*, 132: 331-366.

[5] Şahin, B. (2017). *Riemannian Submersions, Riemannian Maps in Hermitian Geometry, and Their Applications*, Academic Press, Elsevier.

[6] Şahin, B., Yanan, Ş. (2018). Conformal Riemannian maps from almost Hermitian manifolds. *Turk. J. Math.*, 42(5): 2436-2451.

[7] Şahin, B., Yanan, Ş. (2019). Conformal semi-invariant Riemannian maps from almost Hermitian manifolds. *Filomat*, 33(4): 1125-1134.

[8] Yanan, Ş., Şahin, B. (2022). Conformal slant Riemannian maps. *Int. J. Maps Math.*, 5(1): 78-100.

[9] Yanan, Ş. (2021). Conformal generic Riemannian maps from almost Hermitian manifolds. *Turkish Journal of Science*, 6(2): 76-88.

[10] Şahin, B. (2010). Conformal Riemannian maps between Riemannian manifolds, their harmonicity and decomposition theorems. *Acta Appl. Math.*, 109(3): 829-847.

[11] Aykurt Sepet, S. (2020). Pointwise bi-slant submersions. *Celal Bayar University Journal of Science*, 16(3): 339-343.

[12] Aykurt Sepet, S. (2021). Conformal bi-slant submersions. *Turk. J. Math.*, 45: 1705-1723.

[13] Yano, K., Kon, M. (1984). *Structures on Manifolds*, World Scientific.

[14] Ohnita, Y. (1987). On pluriharmonicity of stable harmonic maps, *Jour. London Math. Soc., s2-35(3): 563-587.*

[15] Yanan, Ş. (2022). Conformal hemi-slant Riemannian maps. *Fundamentals of Contemporary Mathematical Sciences*, 3(1): 57-74.

[16] Yanan, Ş. (2022). Conformal semi-slant Riemannian maps from almost Hermitian manifolds onto Riemannian manifolds. *Filomat*, 36(5): 1719-1732.

[17] Nore, T. (1986). Second fundamental form of a map. *Ann. di Mat. Pura ed Appl., 146*: 281-310.

[18] Falcitelli, M., Ianus S., Pastore, A.M. (2004). *Riemannian Submersions and Related Topics*, World Scientific.

[19] Baird, P., Wood, J.C. (2003). *Harmonic Morphisms between Riemannian Manifolds*, Clarendon Press.

[20] Yanan, Ş. (2022). Conformal bi-slant Riemannian maps. *Eastern Anatolian Journal of Science,* article in press.

NATURENGS
MTU Journal of Engineering and Natural Sciences
https://dergipark.org.tr/tr/pub/naturengs
DOI: 10.46572/naturengs.1033182

Research Article

Classification of Single and Combined Power Quality Disturbances Using Stockwell Transform, Relieff Feature Selection Method and Multilayer Perceptron Algorithm

Düzgün AKMAZ[1*]

[1] Department of Electrical and Electronics Engineering, Engineering Faculty, Munzur University, Tunceli, Turkey.

(**Received:** 06.12.2021; **Accepted:** 18.05.2022)

ABSTRACT: In this study, a method based on Stockwell transform (ST), ReliefF feature selection method and Multilayer Perceptron Algorithm (MPA) algorithm was developed for classification of Power Quality (PQ) disturbance signals. First of all, ST was applied to different PQ signals to obtain classification features in the method. Then, total of 30 different classification features were obtained by taking different entropy values of the matrix obtained after ST and different entropy values of the PQ signals. The use of all of the classification features obtained causes the method to be complicated and the training/testing times to be prolonged. Therefore, so as to determine the effective ones among the classification features and to ensure high classification success with less classification features, ReliefF feature selection method was used in this study. PQ disturbances were classified by using 8 different classification features determined by ReliefF feature selection method and MPA. The simulation results show that the method provides a high classification success in a shorter training/testing time. At the same time, simulation results have shown that the method was successful on testing data with noise levels of 35 dB and above after only one training.

Keywords: Classification, Multilayer perceptron algorithm, Power quality, Relief feature selection, S-transform.

1. INTRODUCTION

In some cases, different faults can occur in power systems due to technical or environmental factors. It is necessary that these signals must be continuously monitored and classified in order to respond more effectively and faster to faults that may occur in power systems. When the methods used in these studies were examined, it is seen that firstly, classification features were obtained with a certain signal processing method, and then Power Quality (PQ) disturbances were classified using these classification features and a classification algorithm [1-4]. Many different signal processing tools such as Discrete Wavelet transform [5-8], Wavelet packet transform [9,10], Gabor–Wigner transform [11], Hilbert–Huang transform [12], Variational Mode decomposition [13] and Stockwell transform (ST) [14-19] were used to analyze PQ disturbances. After obtaining the classification features, different classification algorithms such as neural networks [5,14,15], probabilistic neural network [12,19], modular neural network

*Corresponding Author: dakmaz@munzur.edu.tr
ORCID number of authors: [1] 0000-0002-4183-6424.

[16], support vector machines [6,10], fuzzy logic [7], fuzzy expert system [17], fuzzy k-nearest neighbor [9] and decision tree [13] were used to classify PQ disturbances.

In addition to these, in some studies, different feature selection methods were also used for classification. The main purpose of feature selection methods is to determine the effective ones among the classification features. In this way, it is aimed to get high classification success with fewer features. Artificial bee colony feature selection method [8], genetic algorithm [9], particle swarm optimization [17] and statistical approach [19] were used as a feature selection method. Detailed analyzes of the methods used in the analysis of PQ signals can be found in [1-4].

ST signal processing method was used in many studies in view of the fact that it can examine PQ signals in both the time and frequency domains [14-19]. Also, different neural network structures were used in many different studies because of their being effective for classification [5,12,14-16]. Therefore, ST was used as signal processing method, while Multilayer Perceptron Algorithm (MPA) was used as classification algorithm in this study. At the same time, ReliefF feature selection method was used to determine effective classification features in this study.

In the literature, Relief feature selection method with DWT/ Hyperbolic S-transform [20] and ReliefF feature selection method with Fourier transform were used [21]. Differently, in this study, ReliefF feature selection method with ST-MPA were used.

The simulation results showed that the proposed method was successful. At the same time, simulation results show that the proposed method can classify PQ disturbances at different noise levels with high accuracy after only one training.

2. STOCWELL TRANSFORM (ST)

ST is an effective signal processing tool that enables the analysis of signals in the time-frequency domain.

For a continuous-time signal, its ST is given as in Eq.(1) [22].

$$S(\tau,f)=\int_{-\infty}^{\infty}h(t)\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(\tau-t)^2f^2}{2}}e^{-j2\pi ft}dt \quad (1)$$

The discrete Fourier transform of continuous time *h(t)* is given by Eq.(2).

$$H\left[\frac{n}{NT}\right]=\frac{1}{N}\sum_{k=1}^{N-1}h(kt).e^{j2\pi nk} \quad (2)$$

In this Eq.(2), *n* takes *n= 0, 1, … N-1* values. By replace *n/NT* for *f* and *lT* for τ in Eq. (1), the generalized S-transform is obtained as in Eq. (3). In Eq. (3), the Gaussian function is represented as in Eq.(4), with *l*, *m* and *n = 0, 1, …, N-1* [22].

$$S\left[lT,\frac{n}{NT}\right]=\sum_{m=0}^{N-1}\mathbf{H}\left[\frac{m+n}{NT}\right]G(m,n)e^{j2\pi ml/N} \quad (3)$$

$$G(m,n)=e^{-2\pi^2m^2\alpha^2/n^2} \quad (4)$$

Two-dimensional TF Time-Frequency contour is obtained with the S-transform of the signal in Eq.(3). The output of S-transform is an $N \times M$ matrix called S-matrix whose rows pertain to frequency and columns to time.

3. RELIEFF FEATURE SELECTION METHOD

The logic of the method is similar to neighborhood algorithms. Also, it works by weighting the closest samples in the classes to which the sample with the feature belongs or not. The general ReliefF algorithm is shown in Figure 1 [23,24].

1.initialize vector **W**

2.for $i=1$ **to** m **do**

3. randomly select an instance R_i

4. find k nearest hits H_j

5. for all classes $C \neq cl(R_i)$ do

6. from class C find k nearest misses M_j(C)

7. **end for**

8. **end for**

9. **for** A=1 **to** a **do**

10. $AH = -\sum_{j=1}^{k} diff \frac{A,R_i H_j}{m.k}$

11. $AM = \sum_{C \neq cl(R_i)} \left[\left(\frac{P(C)}{1-P(cl(R_i))} \right) \sum_{j=1}^{k} diff\left(A, R_i, M_j(C)\right) \right] / (m.k)$

12. $W[A] = W[A] + (AH + AM)$

13. **end for**

14. **end for**

Figure 1. Pseudo code of ReliefF feature selection algorithm

As seen here, m is the number of iteration. At the same time, m corresponds to the number of samples from data to perform the estimation. Each selected sample R_i equally contributes to the a-size weights vector W. The number of features in the dataset is indicated by a. The algorithm by random selection picks an instance R_i then it searches for k nearest neighbours from the same class (H_j). The algorithm also searches for k nearest neighbours from each of class (nearest misses M_j(C)). The algorithm updates the vector $W[A]$ of estimations of the qualities of attributes depending on R_i, H_j, and M_j(C). The whole process is repeated m times. A detailed explanation of the method can be found in [23,24].

4. MULTILAYER PERCEPTRON ALGORITHM

The MPA is basically a neural network-based artificial intelligence algorithm. This algorithm consists of the input layer, hidden layers, and output layer. One or more hidden layers can be used according to the methods [25,26].

Eq.(5) and Eq.(6) show the connection between input and output functions.

$$T = \sum_{i=1}^{n} x_i w_i + w_0 \tag{5}$$

$$Y=f(T) \tag{6}$$

As can be seen from the Eq.(5) and Eq.(6), the inputs are multiplied by weights and then summed up with a threshold. Finally, an activation function is used to calculate the output. Sigmoid activation function is shown Eq.(7) [25,26].

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{7}$$

5. SIMULATION RESULTS

In this study, sine, sag, swell, harmonic, transient, interruption, sag with harmonic, swell with harmonic and flicker signals were investigated. The above-mentioned signals were produced synthetically and the parameters of the signals were randomly selected within the ranges specified in [27]. The sampling frequency of the signals was chosen as 3.2 kHz.

The MATLAB program was used for the signal generation and feature extraction. The WEKA program was used for the ReliefF and MPA algorithms.

WEKA program is a currently-used program in data mining and different PQ studies [28,29]. The sigmoid activation function was used in the MPA algorithm in WEKA and the number of neurons was determined automatically. While 19 neurons were determined for 30 classification features, 8 neurons were determined for 8 classification features. The network structure was single layer and backpropagation method was used. The general steps of the classification method used in the study are shown in Figure 2.

So as to test the success of the method, training and testing data were generated under different conditions.100 different simulations were performed for each PQ disturbance in the training data. In total, 900 different faults occurred for 9 different PQ disturbances.40 dB noise was added to the generated training data.

Similarly, 100 different simulations were performed for each PQ event in the testing data. In total, 900 different faults occurred for 9 different PQ disturbances. 40 dB noise was added to the generated testing data.

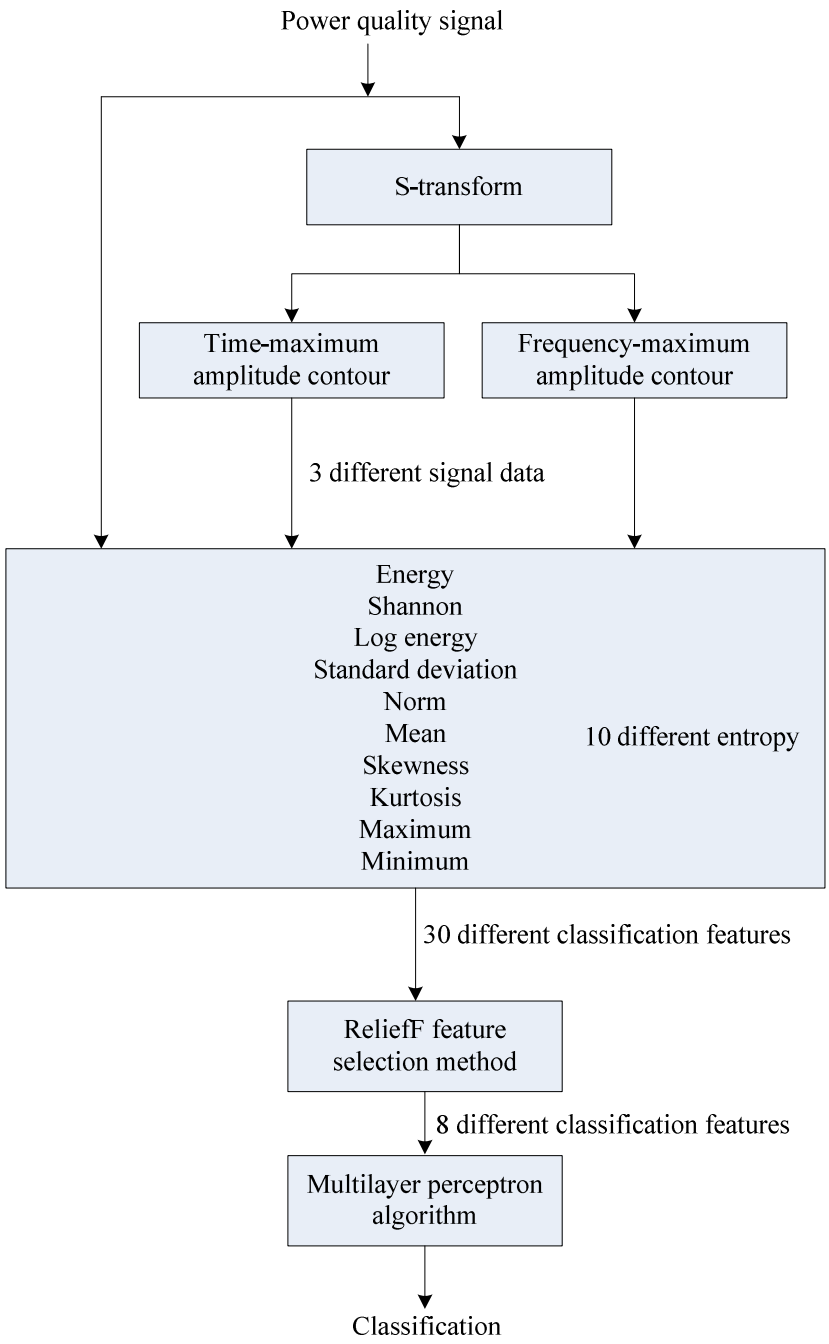

Figure 2. Classification method

Then, in order to obtain classification features, ST was applied to PQ disturbance signals. Two-dimensional (TF Time-Frequency) matrix of each signal was obtained after ST. This is an *NxM* matrix containing complex numbers. The columns of the matrix relate to time and the rows of the matrix relate to frequency. The time - the largest amplitude sequences (TmA- Time-maximum amplitudes) were obtained by searching for the largest values in the column values. The frequency - largest amplitude sequences (FmA- Frequency-maximum amplitude) were obtained by searching for the largest values in the row values.

Thirty (30) different classification features were obtained in total by taking 10 different entropy values of the PQ signal, the time-maximum amplitude signal and the frequency-maximum amplitude signal. These used entropies are Energy, Shannon, Log energy, Standard deviation, Norm, Mean, Skewness, Kurtosis, Maximum, and Minimum.

Then, PQ disturbances were classified by using 30 different classification features and MPA algorithm. Table 1 presents the simulation results obtained with 30 different classification features and MPA algorithm. There is 40 dB noise in the PQ signals in both the training and testing data.

Table 1. Simulation results obtained with 30 different classification features and MPA method (Training and testing data with 40dB noise)

	Training	**Testing**
Classification success %	99.88	98.88
Time taken to build model (s)	4.74	5.2
Time taken to test model (s)	0.01	0.02

As can be seen in the table, the success in the testing data was 98.88%. The WEKA program results showed that time taken to test model on the training data was 0.01 s and time taken to test model on the testing data was 0.02 s.

In addition, the results of the WEKA program showed that the time taken to create the model in the training data was 4.74 seconds. As can be seen in this table, a high classification success was achieved with 30 different classification features.

However, using the feature selection method, similarly high classification success can be achieved with less classification features. In this study, ReliefF feature selection method and MPA algorithm were used to provide a high classification success with less classification features. Using fewer classification features may shorten the training/testing times of the method.

First, the most effective features for classification were determined by the ReliefF method. This method was applied to the training data. Then, these features were added sequentially and the test success in the MPA method was examined. Here, the highest classification success was tried to be achieved by using the least classification feature. Table 2 shows the features selected by the ReliefF algorithm and the classification success of the MPA method.
As seen in this table, the highest classification success was achieved in the testing data with at least 8 selected features.

Table 2. Features determined by ReliefF method and classification success of MPA algorithm.

Number of different data	**Test classification success %**	Determined features
1	33.66	14
2	64.66	14,23
3	80.55	14,23,20
4	91.55	14,23,20,10
5	92.55	14,23,20,10,9
6	96.33	14,23,20,10,9,28
7	97.44	14,23,20,10,9,28,17
8	**98.22**	**14,23,20,10,9,28,17,26**
9	97.55	14,23,20,10,9,28,17,26,27
10	97.77	14,23,20,10,9,28,17,26,27,25
11	96.88	14,23,20,10,9,28,17,26,27,25,19

Table 3 shows the definitions of the determined features. According to Table 3, it can be seen that different entropy values of different signals were selected for classification.

Table 3. Description of the classification features determined by the ReliefF algorithm.

Dimension	**Feature**	**Signal** **Time-maximum amplitude** **Frequency-maximum amplitude**	**Entropy**
1	14	Time-maximum amplitude	Standard deviation
2	23	Frequency-maximum amplitude	Log energy
3	20	Time-maximum amplitude	Minimum
4	10	Signal	Minimum
5	9	Signal	Maximum
6	28	Frequency-maximum amplitude	Kurtosis
7	17	Time-maximum amplitude	Skewness
8	26	Frequency-maximum amplitude	Mean

Training and testing were carried out according to the 8 selected classification features in Table 3. Other classification features were removed from the database. The WEKA program was used again and the simulation results were examined. In Table 4, the simulation results obtained with 8 different classification features were shown.

Table 4. Simulation results obtained with 8 different classification features and MPA method (Training and testing data with 40dB noise).

	Training	**Testing**
Classification success %	98.88	98.22
Time taken to build model (s)	1.24	1.22
Time taken to test model (s)	0	0.01

As can be seen in Table 4, a classification success of 98.22% was obtained in the testing data with 8 selected classification features. In Table 1, a classification success of 98.88% was obtained in the testing data with 30 features. Table 1 and Table 4 showed that successful

classification results were obtained by using 30 classification features with MPA and 8 determined classification features with MPA.

However, as seen in Table 1 and Table 4, the method (training/testing) was realized in a much shorter time with 8 classification features. For example, as seen in Table 1, when 30 classification features were used, the time required for the time taken to build the model was 4.74 s. As can be seen from Table 4, it takes 1.24 s with 8 classification features. Thus, a high classification success was achieved in a shorter time with 8 selected features. Therefore, in this study, it is proposed to use 8 classification features and MPA to classify PQ disturbances.
The MPA structure for the 8 determined features was shown in Figure 3.

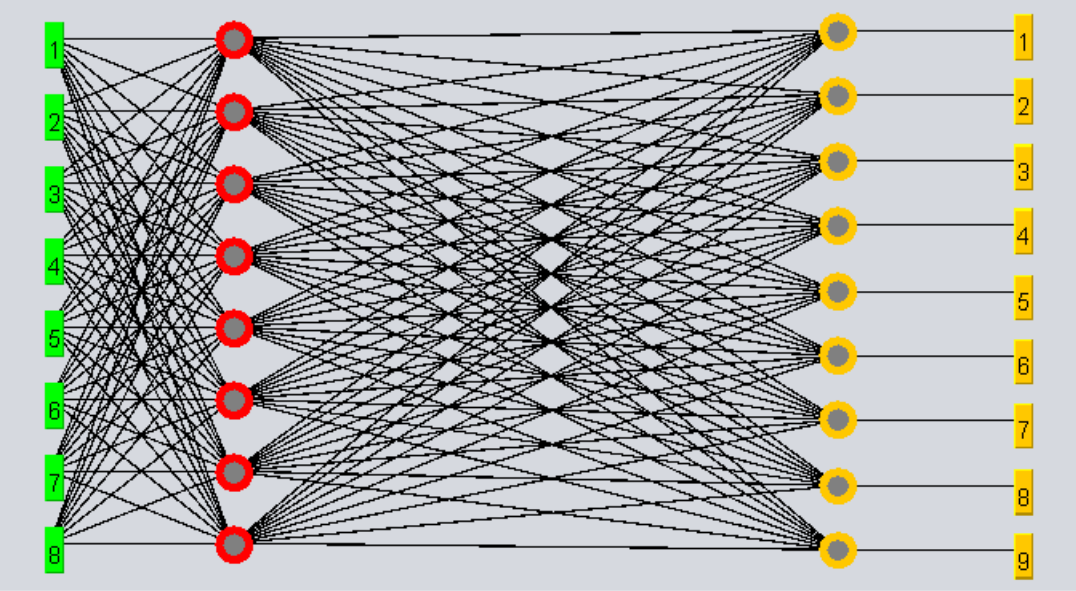

Figure 3. The MPA structure for the 8 determined features

At the same time, the proposed method should be tested on data with different noise levels. To do this, testing data with different noise were used. 30,35,40,45 and 50dB noise were added to the testing data. In the training data, 40 dB noise was used. Table 5 shows the classification success of the method for different noise levels.

Table 5. Classification success of 8 different classification features and MPA method in testing data with different noise levels (Training data with 40dB noise)

Test data noise level (dB)	**30**	**35**	**40**	**45**	**50**
Success (%)	79.22	94.55	98.22	98.66	98.66

Table 5 shows that the implemented method was effective in classifying PQ disturbances at different noise levels once trained. After the training was implemented, it was determined that the method was successful in noises of 35 dB and above.

In Table 6, the classification results of different methods in PQ signals with 40db noise were compared. As can be seen from this table, the proposed method has achieved an acceptable classification success.

Table 6. Comparison of the different methods (40dB noise results)

Ref	Signal processing	Classification algorithm	Feature selection method	Number of disturbances	Feature number	Success (40dB) %
[9]	Wavelet packet transform	Fuzzy k-nearest neighbour algorithm	GA	10	16	95.96
[8]	DWT	Probabilistic neural network	Artificial bee colony	16	9	98.6
Pro.	ST	MPA	ReliefF	9	8	98.22

6. CONCLUSIONS

Nowadays, many different methods were developed for classifying power quality signals. It is requested that these methods are as effective and simple as possible. In this study, a method based on Stockwell-transform, ReliefF feature selection method and Multilayer Perceptron Algorithm was applied for classification. As a result of the application of signal processing approach, 30 different classification features were obtained. However, at the present time, it is expected to achieve high classification success with less classification features. Therefore, ReliefF feature selection method was applied in the study. By using ReliefF feature selection method, 8 different classification features were determined. When the simulation results were examined, it shows that using 8 determined features instead of 30 features turns the classification method into a simpler structure and can make classification faster. The simulation results showed that the method was successful for classification. At the same time, all the obtained results showed that the proposed method provided an acceptable classification success.

Declaration of Competing Interest

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

Author Contribution

Düzgün AKMAZ contributed 100% at every stage of the article.

REFERENCES

[1] Ahsan, M. K., Pan, T., Li, Z. (2018). A three decades of marvellous significant review of power quality events regarding detection & classification. *Journal of Power and Energy Engineering*, *6*(8), 1-37.

[2] Choong, F., Reaz, M. B. I., Mohd-Yasin, F. (2005). Advances in signal processing and artificial intelligence technologies in the classification of power quality events: a survey. *Electric Power Components and Systems*, *33*(12), 1333-1349.

[3] Khokhar, S., Zin, A. A. B. M., Mokhtar, A. S. B., Pesaran, M. (2015). A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances. *Renewable and Sustainable Energy Reviews*, *51*, 1650-1663.

[4] Mishra, M. (2019). Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review. *International transactions on electrical energy systems*, *29*(8), e12008.

[5] Uyar, M., Yildirim, S., Gencoglu, M. T. (2008). An effective wavelet-based feature extraction method for classification of power quality disturbance signals. *Electric power systems Research*, *78*(10), 1747-1755.

[6] Erişti, H., Demir, Y. (2010). A new algorithm for automatic classification of power quality events based on wavelet transform and SVM. *Expert systems with applications*, *37*(6), 4094-4102.

[7] Meher, S. K., Pradhan, A. K. (2010). Fuzzy classifiers for power quality events analysis. *Electric power systems Research*, *80*(1), 71-76.

[8] Khokhar, S., Zin, A. A. M., Memon, A. P., Mokhtar, A. S. (2017). A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement*, *95*, 246-259.

[9] Panigrahi, B. K., Pandi, V. R. (2009). Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm. *IET generation, transmission & distribution*, *3*(3), 296-306.

[10] Manimala, K., Selvi, K., Ahila, R. (2012). Optimization techniques for improving power quality data mining using wavelet packet based support vector machine. *Neurocomputing*, *77*(1), 36-47.

[11] Cho, S. H., Jang, G., Kwon, S. H. (2009). Time-frequency analysis of power-quality disturbances via the Gabor–Wigner transform. *IEEE transactions on power delivery*, *25*(1), 494-499.

[12] Kumar, R., Singh, B., Shahani, D. T. (2015). Recognition of single-stage and multiple power quality events using Hilbert–Huang transform and probabilistic neural network. *Electric Power Components and Systems*, *43*(6), 607-619.

[13] Achlerkar, P. D., Samantaray, S. R., Manikandan, M. S. (2016). Variational mode decomposition and decision tree based detection and classification of power quality disturbances in grid-connected distributed generation system. *IEEE Transactions on Smart Grid*, *9*(4), 3122-3132.

[14] Lee, I. W., Dash, P. K. (2003). S-transform-based intelligent system for classification of power quality disturbance signals. *IEEE Transactions on Industrial Electronics*, *50*(4), 800-805.

[15] Uyar, M., Yildirim, S., Gencoglu, M. T. (2009). An expert system based on S-transform and neural network for automatic classification of power quality disturbances. *Expert Systems with Applications*, *36*(3), 5962-5975.

[16] Bhende, C. N., S. Mishra, B. K. Panigrahi. "Detection and classification of power quality disturbances using S-transform and modular neural network." *Electric power systems research* 78.1 (2008): 122-128.

[17] Behera, H. S., Dash, P. K., Biswal, B. J. A. S. C. (2010). Power quality time series data mining using S-transform and fuzzy expert system. *Applied Soft Computing*, *10*(3), 945-955.

[18] Dash, P. K., Panigrahi, B. K., Panda, G. (2003). Power quality analysis using S-transform. *IEEE transactions on power delivery*, *18*(2), 406-411.

[19] Huang, N., Xu, D., Liu, X., Lin, L. (2012). Power quality disturbances classification based on S-transform and probabilistic neural network. *Neurocomputing*, *98*, 12-23.

[20] Hajian, M., Foroud, A. A., Abdoos, A. A. (2014). New automated power quality recognition system for online/offline monitoring. *Neurocomputing*, *128*, 389-406.

[21] Valencia-Duque, A. F., Meza, A. Á., Orozco-Gutiérrez, A. A. (2019). Automatic identification of power quality events using a machine learning approach. *Scientia Et Technica*, *24*(2), 183-189.

[22] Karasu, S., Saraç, Z. (2017, May). Classification of power quality disturbances with S-transform and artificial neural networks method. In 2017 25th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[23] Palma-Mendoza, R. J., Rodriguez, D., De-Marcos, L. (2018). Distributed ReliefF-based feature selection in Spark. *Knowledge and Information Systems*, *57*(1), 1-20.

[24] Parlar, T. (2021). A heuristic approach with artificial neural network for Parkinson's disease. *International Journal of Applied Mathematics Electronics and Computers*, *9*(1), 1-6.

[25] Morariu, D., Crețulescu, R., Breazu, M., The weka multilayer perceptron classifier, International Journal of Advanced Statistics and IT&C for Economics and Life Sciences, 7(1). 2018.

[26] Area, S., Mesra, R., Analysis of Bayes, neural network and tree classifier of classification technique in data mining using WEKA, 2012.

[27] Moravej, Z., Abdoos, A. A., Pazoki, M. J. E. P. C. (2009). Detection and classification of power quality disturbances using wavelet transform and support vector machines. Electric Power Components and Systems, 38(2), 182-196.

[28] Vinayagam, A., Veerasamy, V., Radhakrishnan, P., Sepperumal, M., Ramaiyan, K. (2021). An ensemble approach of classification model for detection and classification of power quality disturbances in PV integrated microgrid network. Applied Soft Computing, 106, 107294.

[29] Kiranmai, S. A., Laxmi, A. J. (2018). Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy. Protection and Control of Modern Power Systems, 3(1), 1-12.

NATURENGS
MTU Journal of Engineering and Natural Sciences
https://dergipark.org.tr/tr/pub/naturengs
DOI: 10.46572/naturengs.1096613

Research Article

The Thermal and Mechanical Properties of Building Stones from the Afyon, İzmir, Muğla and Denizli Region

Ayşe Biçer[1*]

Department of Bio Engineering, Faculty of Engineering and Natural Sciences Malatya Turgut Ozal University, Malatya-Turkey

(**Received:** 31.03.2022; **Accepted:** 13.05.2022)

ABSTRACT: This study investigated the thermal and mechanical properties of Iscehisar stone, Alacati stone, Milas marble, and Denizli schist. Samples were taken from quarries for each stone. Chemical analyses were carried out. Afterward, thermal conductivity, compressive stress, water absorption, water vapor permeability, and wear experiments were conducted. Alacati stone had the highest thermal conductivity (0.381 W/mK). Denizli schist had the highest compressive strength (95 MPa). All samples had water absorption rates lower than 30%. The results were compared with energy-saving, strength, and comfort with other building materials.

Keywords: Iscehisar stone, Alacati stone, Milas marble, Denizli schist, Building material.

1. INTRODUCTION

Natural stones are prevalent in all parts of the world. Most historic buildings are made of natural stones, such as Pyramids, Greek Acropolises, Roman Amphitheaters, and Ottoman fountains and mosques. Natural Stones are commonly used because they are stronger and more durable than other traditional building materials. Stones were brought from hundreds of miles away to build those structures. However, sometimes natural stones are used because there are rich stone reserves around the construction area.

We need to know some physico mechanical properties to determine where to use natural stones. For example, while density and porosity negatively affect some physical and mechanical properties, they positively affect heat and sound insulation. Those properties are even more critical when it comes to the construction of outdoor spaces and large structures under different climatic conditions.

Having a considerable and quite diverse mineral base, Turkey has one of the world's largest natural stone reserves. Natural stones build roads, patios, walkways, walls, and dams. Natural stone production has become an important sector with developments in the construction sector in recent years [1]. Natural stones have become more popular because the prices of construction materials have risen, and the demand for housing has exceeded its supply.

*Corresponding Author: ayse.bicer@ozal.edu.tr
ORCID number of authors: [1] 0000-0003-4514-5644

There is a large body of research on building stones. Some of those studies focus on natural stones' geological formation, properties, and geographical distribution [1- 9]. Other studies address building stones in the construction industry [10 - 18].

We need to know the physicomechanical properties of natural stones to use them as building materials in the best way. This study focused on the thermal and mechanical properties of Iscehisar stone (Afyon), Alacati stone (İzmir), Milas White Marble (Muğla), and Denizli schist. Local people have used those stones throughout history. This study investigated why they were preferred.

2. MATERIALS AND METHODS

2.1. Materials

Iscehisar Stone (Afyon)
Samples were obtained from the andesite quarries of Ağın Mountain, north of the Iscehisar sub-district of Afyonkarahisar to determine the technical characteristics of the Iscehisar stone. The quarries in question are up and running, and andesites produced by them are used as building blocks in the region. Those andesites are grayish, pinkish, or red-purple. Iscehisar andesites are volcanic rocks used as parquet, flooring, and covering stones in buildings, walkways, parks, and gardens in different regions of Turkey (Fig 1). In addition, Iscehisar stone is also used to restore historical buildings and as sitting groups and flower beds [19]. In recent years, Iscehisar andesite has become popular among domestic and foreign natural stone users because it is homogeneous and unfading with unpolished, wiped, hammered, or rough-hewn surface forms. It has a hardness of 3-4 Mohs.

a) b)

Figure 1. Iscehisar stone examples a) Iscehisar (Koca) bridge, [20], b) Building surfacing [21]

Alacati stone (Izmir)
Alacati stone (Izmir) is called so because of the stone quarries in the Alacati region of the Cesme district of Izmir province [22]. It is also known as “Alapietra.” It is a light-colored, slightly porous type of sedimentary stone composed of volcanic ash, sand, and lava particles. It is white, off-white, or cream. It is easy to chip and shape because it is soft when quarried. It hardens over time and turns into a natural building material. There is no need for plastering, whitewashing, or similar applications on the interior and exterior facades of buildings made of Alacati stone. It is also a natural thermal insulator. It is used as a decoration element in indoor and outdoor spaces, providing a visual richness in walls, arches, columns, and gardens (Fig 2). It has been used to build stone houses in the region throughout history. It has a hardness of 3 Mohs.

a) b)

Figure 2. Alacati stone examples a) Pazaryeri mosque [23], b) Stone houses [24]

Milas White Marble

Samples were obtained from the marble quarries located in the Milas district of Muğla province. The Milas White Marble has homogeneous lines with very light gray tones in white color. It has a hardness of 3-4 Mohs. It is used in all applications in interior and exterior areas and special projects (baths, fireplaces, etc.) and home decorations. It is also widely used in buildings, floors, counters, stairs, steps, exteriors, tombs, and sculptures [25] (Fig 3).

a) b)

Figure 3. Milas White Marble a) floor and stair covering [26], b) processed state [27]

Denizli schist

There are rich schist reserves around Çal and Bekilli districts in the northeast of Denizli. Denizli schist, also called a slate stone, is used as decorative and accessory coating stones in domestic and foreign markets [28]. Slate stone is introduced to domestic markets, especially in Istanbul, Izmir, Bursa, and Antalya [29]. Slate stone is used in walkways, park and garden arrangements, wall covering, and pool edges (Fig 4). It has a hardness of 5-6 Mohs.

a) b)

Figure 4. Denizli schist applications a) Wall covering [30], b) Flooring [31]

2.2. Methods

Thermal experiments were performed on 150x60x20 mm samples, while compressive strength and wear experiments were performed on 100x100x100 mm samples.

A Shotherm-QTM unit based on DIN 51046 hot wire method was used to measure the thermal conductivity values of the samples [32**,** 33]. Measurements were made from three different points of each sample, and the average of the three values was taken.

The compressive strength and wear tests were carried out according to TS 699 standard [34]. Table 2 shows the results.

Water absorption rates should be below the critical value of 30 % because of the risk of cracking, fragmentation or dispersion in case building stones come into contact with water below 0 °C [35]. Water absorption rates were calculated using Equation (1). Fig 5 shows the variation of the weights of the samples with time. Fig 6 shows the drying rates.

$$\text{Water absorption percent} = \{[W_d - W_k] / W_k\} x100 \quad (1)$$

where,

W_k and W_d are the dry and water absorbed weights, respectively.

Table 2 shows the measurement and calculation results collectively.

3. RESULTS AND DISCUSSIONS

This study investigated some stones' thermal and mechanical properties from Afyon, İzmir, Muğla, and Denizli in the Aegean Region, Turkey. Below are the results.

Table 1 shows the components determined as a result of chemical analysis collectively.

Table 1. The chemical composition of the samples (%)

component / Material	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	K_2O	Loss of ignition	Undefined
İşcehisar stone:	58.40	14.7	4.88	4.71	2.83	2.92	8.25	3.85
Alacatı Stone	57.88	8.34	0.83	10.54	0.76	4.01	10.22	7.45
Milas Matble (white)	0.231	0.085	0.09	53.7	0.62	0.06	43.7	1.53
Denizli schist	70.18	4.81	0.06	0.21	0.64	3.25	16.11	4.75

Iscehisar andesites are volcanic rocks with a volume abrasion and water absorption of 7.43 % and 5.39 %, respectively. They are used in many regions in buildings, walkways, and paving stones. Iscehisar andesites have a thermal conductivity coefficient of 0.531 W/mK, which is better than concrete, granite, marble, limestone, and sandstone in Table 3. This result indicates that Iscehisar andesites can provide heat and sound insulation if used as a filling material on bricks, briquettes, and walls. Iscehisar andesites have a compressive strength of 70 MPa, close to sandstone and better than limestone, marble, and common brick. In addition to these features, Iscehisar andesites are commonly used due to abundant stone reserves and low costs.

Alacati stone is a lightweight building material that is easy to sculpt and shape because it is soft when first quarried. Therefore, it is widely used in buildings. There is an increased demand for Alacati stone because it is used on interior and exterior facades without the need for plaster, whitewash, and similar applications. It is also used as an ornamental element. This is indicated by the fact that it has been widely used to build stone houses in the region for a very long time. It has a thermal conductivity of 0.381 W/mK, suggesting that it is better than the other construction materials (Table 3) and can provide heat and sound insulation in buildings. It has a water absorption rate of smaller than 30 %, indicating that it can be used in humid environments. It has a compressive strength of 16.29 MPa, which is smaller than artificial materials such as concrete, briquette, brick, and aerated concrete (Table 3). Therefore, it should not be used as a bearing element. However, it has a density value of 1.370 kg/m^3, suggesting that it can reduce building loads if used as aggregate in concrete. It has a volume abrasion rate of 25.48 %, indicating that it should not be used as a flooring material.

Milas Marble has a thermal conductivity of 2.720 W/mK, which is not very promising. However, it has a compressive strength of 64.7 MPa and a volume abrasion of 1.10 %, indicating that it can be used as a flooring material. It has a water absorption of 0.18 %, suggesting that it can be used to make kitchen counters as it is suitable for wet floors.

Denizli Schist has compressive strength, volume abrasion, and water absorption of 95 MPa, 0.12 %, and 0.33 %, respectively. These values show that it can be used to construct walkways, park and garden arrangements, wall coverings, and poolsides. It has a heat transfer coefficient of 1.163 W/mK, suggesting that it can be used as a wallcovering. There is an increased demand for Denizli schist because it looks nice and provides insulation. It has a higher thermal conductivity than granite, sandstone, and marble. Its thermal conductivity is similar to that of limestone. However, it looks better than all materials, except for granite (Table 3).

For five reasons, there has been an increased demand for Iscehisar stone, Alacati stone, Milas marble, and Denizli schist. First, they have good thermal and mechanical properties. Second, Turkey is rich in reserves. Third, they are easy to access. Fourth, they are low cost. Fifth, they are easy to process.

Table 2. Thermal and mechanical properties of stones

Materials	Density (kg/m^3)	Thermal conductivity (W/mK)	Compressive strength (MPa)	Water absorption (%)	Volume abrasion (%)
İşcehisar stone	2.335	0.531	70	5.39	7.43
Alacatı stone	1.370	0.381	16.29	24.9	25.48
Milas matble	2.720	2.461	64.7	0.18	1.10
Denizli schist	2.150	1.163	95	0.33	0.12

The change in weights concerning time in the water absorption test is shown in Figure 5, and the drying rates are shown in Figure 6. In the case of examining the drying rates, it can be said that the four local stones examined have breathing ability, albeit slightly. In Figure 7, the thermal and mechanical properties of the examined stones are shown collectively to evaluate them together.

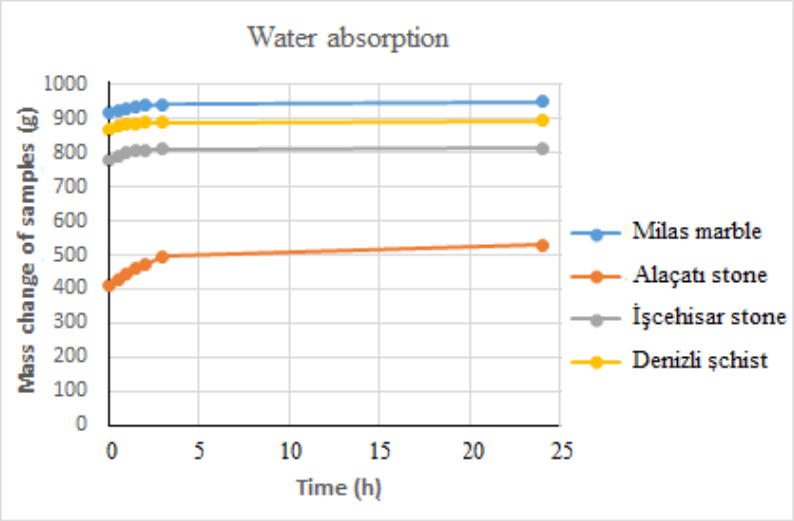

Figure 5. Mass change of stones according to time in water absorption test

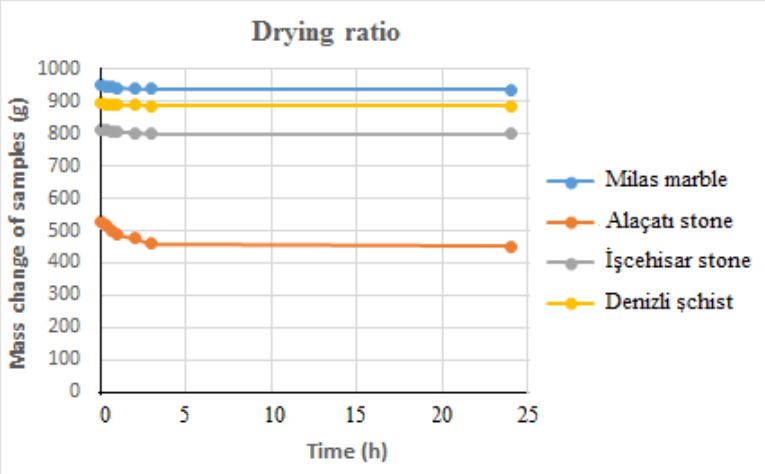

Figure 6. Mass change of stones according to time in drying test

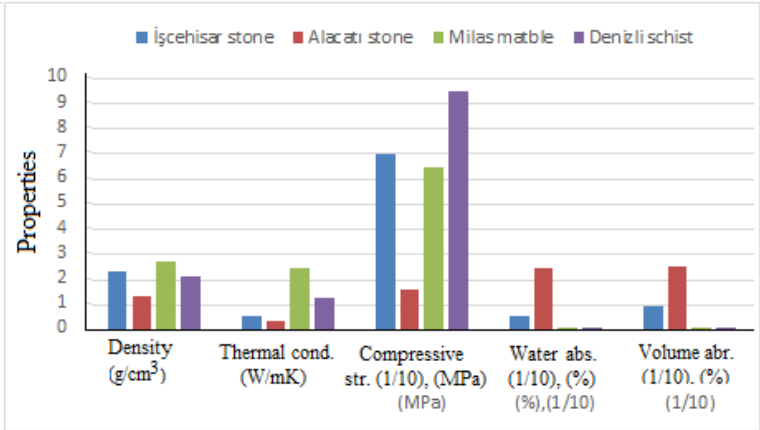

Figure 7. View of the physical properties of the stones together

Table 3. The physical properties of some building materials [36].

Materials	Density (kg/m^3)	Thermal conductivity (W/mK)	Compressive strength (MPa)
Concrete	1906	0.814	20
Granite	2643	1.73	120
Limestone	2483	1.16	35
Sandstone	2235	1.85	80
Marble	2603	2.77	50
Common brick	1602	0.692	16

4. CONCLUSIONS

This study investigated some stones' thermal and mechanical properties from Afyon, İzmir, Muğla, and Denizli in the Aegean Region, Turkey. The results are as follows:

✓ Iscehisar stone has a low thermal conductivity (0.531 W/mK) and compressive strength (70 MPa) value. Therefore, it can provide an energy economy advantage if used on load-bearing walls or as a coating material.

✓ İzmir Alacati stone can be used as an aggregate and wall covering material in low-density concretes, briquettes, and bricks because it is a lightweight material (1370 kg/m^3) with a low heat transfer coefficient (0.381 W/mK). If it is used for that purpose, it can reduce heating costs. It has a high volume abrasion (25.48 %), which cannot be used as a floor covering material.

✓ Milas Marble has a thermal conductivity of 2.461 W/mK, which is not promising. However, it can be used as a floor covering for wet floors.

✓ Denizli schist has compressive strength, volume abrasion, and water absorption of 95 MPa, 0.12 %, and 0.33 %. Therefore, it can be used in applications, such as floors, wall cladding, and poolsides.

✓ Iscehisar, Alacati, Milas marble, and Denizli schist have good thermal and mechanical properties. Moreover, those stones are easy to access and process at low costs. In addition, Turkey is rich in reserves.

✓ While the rocks with small density have lower thermal conductivity values, stones with bigger density have higher compressive strength values.

Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

Author Contribution

Ayşe Biçer contributed 100% at every stage of the article.

REFERENCES

[1] Taşlıgil N,, Şahin G. (2016). Yapı malzemesi olarak kullanılan Türkiye doğal taşlarının iktisadi coğrafya odağında analizi, *Marmara Coğrafya Dergisi,* 33, 607-640.

[2] Pivko D. (2003). Natural stones in earth's history, *Acta Geologica,* 58, 73-86.

[3] Gevrek A.İ., Kazancı N. (1991). İgnimbrit: oluşumu ve özellikleri, *Jeoloji Mühendisliği Dergisi*, Ankara.38, 39-42.

[4] Kazancı N., Gürbüz A. (2014), Jeolojik miras nitelikli Türkiye doğal taşları, *Türkiye Jeoloji Bülteni*. 57: 1.

[5] Dinçer İ., Özvan A., Akın M., Tapan M., Oyan V. (2012). İgnimbiritlerin kapiler water absorption potansiyellerinin değerlendirilmesi: Ahlat Taşı örneği. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 17 (2), 64-71.

[6] Duran F. (2009). Erciyes volkanizmasının oluşumu, Koçcağız Köyü (Kayseri) dolayının stratigrafisi ve tüflerin yapı-kaplama taşı olarak kullanılabilirliği. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü*, Yüksek Lisans Tezi*, Adana.

[7] Kaygısız H. (2010). Kayseri yöresindeki yapıtaşlarının fiziko-mekanik özelliklerinin belirlenmesi. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, *Yüksek Lisans Tezi*, Adana.

[8] Türkdönmez O., Bozcu, M. (2012). The geological, petrographical and engineering properties of rhyolitic tuffs (Çan Stone) in Çan-Etili Area (Çanakkale), NW Turkey: Their usage as building and covering stones, *Open Journal of Geology,* 2, 25-33.

[9] Devecioglu A.G. (2001). An investigation on the heat conduction parameters of porous building stones, *Master Thesis*, Firat University.

[10] Bakış A., Işık E., Hattatoğlu F., Akıllı A. (2014). Jeolojik miras nitelikli Ahlat Taşı'nın inşaat sektöründe kullanımı, *III. Uluslararası Ahlat-Avrasya Bilim, Kültür ve Sanat Sempozyumu Bildiriler Kitabı* (Editörler Doğru M. ve Aksoy E.), 46-59, 22-24 Eylül Ahlat- Bitlis.

[11] Bicer A. (2019-a). Ahlat ve Malazgirt yapı taşlarının bazı fiziksel özellikleri, *Fırat Üniversitesi Müh. Bil. Dergisi*, 31(2), 301-307.

[12] Bicer A. (2019-b). Some physical properties of the building stones from southeastern Anatolia region, *Bartın University International Journal of Natural and Applied Sciences,* 2(1), 9-15.

[13] Bicer A. (2019-c). Some physical properties of the building stones from Elazıg-Nevsehir region, *Nevşehir Bilim ve Teknoloji Dergisi,* 8(2), 96-102.

[14] Demir I. (2005). The usage properties of Kırşehir regional rocks as crushed stone aggregate, *Journal of Polytechnic*, 8(1), 111-121.

[15] Çavumirza M., Kılıç Ö., Anıl M. (2003). Mucur (Kırşehir) yöresi kireçtaşı mermerleri ve travertenlerinin fiziko-mekanik özellikleri, *Türkiye ıv mermer sempozyumu (mersem'2003) bildiriler kitabı* 18-19 Aralık.

[16] Daloğlu G., Emir E. (2010). The assessment of tuffs located at Eskisehir-Derbent region as the natural building stone , *Journal of Engineering and Architecture Faculty of Eskişehir Osmangazi University*, Vol: XXIII, No:1, 2010.

[17] Kılıç G (2009). Edirne (Keşan) Bölgesi kumtaşlarının yapı taşı olarak kullanılabilirliği, *Doktora Tezi*, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Mimarlık Anabilim Dalı, Edirne, 2009.

[18] Kılıç G., Gültekin A.H. (2010).Sürdürülebilir bir yapı malzemesi olarak kumtaşı*, Uluslararası Sürdürülebilir Yapılar Sempozyumu (ISBS 2010*), Gazi Üniversitesi, 51-54, Ankara, 26-28 Mayıs.

[19] Çelik M.Y. (2019). Iscehisar andezitinin donma çözülme sürecinde bazı fiziksel parametrelerdeki değişiminin incelenmesi, Afyon Kocatepe Üniversitesi, Afyon Meslek Yüksek Okulu, DOI:10.21205/deufmd.2019216229.

[20] URL-1. https://afyoncitymap.com/listing/citymap/Iscehisar-koca-koprusu/

[21] URL-2. https://www.reklamzamani.net/firma/33960-Iscehisar-tas-duvar-dekorasyon--bekar-yapi.

[22] Gündüz L., Kalkan Ş.O., Aydoğdu N.K. (2016). İzmir-Alacati taşının kuru karışım hafif beton agregası olarak kullanılabilirliği üzerine teknik bir analiz, *8.Uluslararası Kırmataş Sempozyumu*, 13-14 Ekim, Kütahya, 390-399.

[23] Avsaroglu N. (2020). Anadolu'nun binlerce yıllık doğal taşları, *MTA Genel Müd*. Ankara.

[24] URL-3. https://www.emlakhaberi.com/sektorel-haberler/koy-alacati-tas-evleri-hayallerinizin-otesinde-h50518.html

[25] Bahadır A., Türk Y.N., Koca M.Y. (2002). The mıneralogıcal, chemical, physical and mechanical properties of Muğla Region Marbies, *Jeoloji Mühendisliği 28 (1)* 2002.

[26] URL-4. https://www.elitrestorasyon.com/Content/Arayuz/Files/dogaltas/840x460/5a521ecd-0e0f-49f0-a7cf-164564c7c33f.jpg

[27] URL-5. https://mermermarble.com/mugla

[28] Yağız S. (2011). Properties of schist extracted in the city of Denizli surroundings as construction material, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi,* 17(3), 157-163.

[29] Özkul M., Yağız S. (2007). Çal bölgesinin jeolojisi ve doğal taş kaynakları, *Çal Sempozyumu Bildirileri*, 94-104. Denizli

[30] URL-6. http://www.dogal-tas.org/urun/310/denizli-kayrak-tasi

[31] URL-7. http://www.dogal-tas.org/urun/284/ebatli-denizli-kayrak-tasi

[32] Vysnıauskas V.V, Zıkas A.A. (1988), Determination of the thermal conductivity of ceramics by the Hot-Wire Technique. *Heat Transfer Soviet Research*, 20 (1): 137-142.

[33] Denko S. (1990), Shotherm Operation Manual No 125-2. K.K. Instrument products department, 13-9, Shiba Daimon, Tokyo, 105, Japan.

[34] TS 699, (2009), The test and experiment methods of natural building stones, *TSE*, Ankara.

[35] BS 812-109 Standards, (1990). Testing aggregates-part 109: methods for determination of moisture content. British Standards Institution.

[36] Toksoy M. (1988), Thermal conductivity coefficients of industrial materials, *Journal of Engineers and Machinery,* 347, 12-15.

NATURENGS
MTU Journal of Engineering and Natural Sciences
https://dergipark.org.tr/tr/pub/naturengs
DOI: 10.46572/naturengs.1097970

Research Article

Using and Comparing Machine Learning Techniques for Automatic Detection of Spam URLs

Muhammed YILDIRIM[1*]

[1]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, Malatya, Turkey.

(**Received:** 03.04.2022; **Accepted:** 18.05.2022)

ABSTRACT: With the developing technology, the issue of cyber security has become one of the most common and current issues in recent years. Spam URLs are one of the most common and dangerous issues for cybersecurity. Spam URLs are one of the most widely used attacks to defraud users. These attacks cause users to suffer monetary losses, steal private information, and install malicious software on their devices. It is very important to detect such threats promptly and take precautions against them. Detection of spam URLs is mainly done by using blacklists. However, these lists are insufficient to detect newly created URLs. Machine learning techniques have been developed to overcome this deficiency in recent years. In this study, URL classification was made using different machine learning techniques. In the study, 9 different classifiers were preferred for URL classification. The performances of the classifiers were compared in the URL classification process. In addition, similar studies in the literature have been comprehensively examined and these studies have been discussed. In addition, since the preparation of datasets in the natural language processing process greatly affects the training of models, these steps are discussed in detail.

Keywords: Cyber Security, Machine Learning, NLP, URL Detection, Classifiers.

1. INTRODUCTION

Developing technology has led to the emergence of different fields. Natural language processing is one of these current areas. Recently, it is a new field in that researchers have extensively researched and developed various applications. Naturel Language Processing (NLP) is a sub-branch of artificial intelligence that aims to understand, analyze, interpret and produce the natural language humans use with the developed systems. NLP brings together the steps of linguistics, artificial intelligence, computer technologies, statistics, and data processing [1, 2]. Evolving technology has led to the emergence of different fields. Natural language processing is one of these current areas. Recently, it is a new field in that researchers have extensively researched and developed various applications. NLP is a sub-branch of artificial intelligence that aims to understand, analyze, interpret and produce the natural language humans use with the developed systems. NLP brings together the steps of linguistics, artificial intelligence, computer technologies, statistics, and data processing [3].

Natural language processing is used in different fields. It is used in many areas, such as text summarization [4], sentiment analysis [5], correction of typos [6], translation systems [7], information extraction [8], and natural language production. NLP was also used in this study to

[*]Corresponding Author: muhammed.yildirim@ozal.edu.tr
ORCID number of authors: [1] 0000-0003-1866-4721

detect Spam URLs. Spam URLs appear as unwanted pop-ups and links that we encounter on the internet every day. Spam URLs can cause users to experience financial losses and steal their private information. In addition, spam URLs can change the ranking of searched pages and negatively affect network traffic. There are improved methods for detecting these URLs. The most commonly used method is the blacklist method, which keeps URL records. The biggest disadvantage of the blacklist method is that it cannot detect newly created spam URLs. Studies have been carried out in recent years to cope with this problem by using artificial intelligence and machine learning methods [9, 10]. In the machine learning approach, models are first trained using training data. These models, which are then introduced, classify newly emerged URLs as spam or normal. This approach is preferred to eliminate the problem of not detecting new sites in the black list method.

The number of words in natural languages is relatively high and words can have more than one meaning. It is a complicated process for machines to understand the different meanings of words. Therefore, machine understanding of natural languages is a difficult process. The increasing importance of NLP with each passing day lies in the acceleration of artificial intelligence studies with the developing technology.

NLP is an up-to-date field that enables the communication between humans and machines. NLP is a popular sub-branch of artificial intelligence that aims to understand natural languages by machines, analyze these languages and make inferences from them [11]. Machines' understanding of people's language will solve many real-life problems and allow people to find a more comfortable space.

In this study, for the machines' URL data to be processed, it must first be converted into a format that the machines can understand. In the study, firstly, text preprocessing steps were applied to the data in the dataset. This step has a great impact on the performance of the models. Then, feature maps were obtained using the bag of words method. The feature maps obtained in the last step were classified into different classifiers.

1.1.Related Works

Spam URLs are attacks that put both individual users and companies in a difficult position. There are various studies in the literature to minimize these attacks.
Do Xuan et al. tested machine learning methods on two different datasets to classify URL addresses in their study. Two different classifiers were used in this study. The researchers also ran the models in 10 and 100 iterations in this study and compared the results. Accuracy values of 93.39% and 90.70%, respectively, were obtained in each dataset in 100 iterations of the SVM classifier [12]. The researchers stated that they prioritized time and accuracy in this study.

Patgiri et al. used machine learning methods to detect malicious URLs in their study. In the study, URLs were classified as good and bad. Two different classifiers were used in the study and the dataset was divided into train and test at different rates. The results obtained by separating the dataset as train and test at different ratios were compared. The researchers stated that their accuracy value in the Random Forest classifier was higher than in the SVM classifier [13].

Jain et al. used URL addresses for phishing detection. Researchers stated that they have developed a new system to prevent phishing in their studies and that the system they developed

works with 14 features. It has been seen that SVM and Naive Bayes are used in the proposed system, and the SVM classifier is more successful. The success rate of this proposed system in detecting phishing has been 90% [14].

Joshi and colleagues explained that most of today's cyberattacks and scams originate from malicious websites. They stated that malicious URLs are delivered to users in different ways and that these URLs cause different harm to users. This study observed that machine learning methods were used to detect malicious URLs and an average of 92% accuracy value was obtained from 5 different data used for testing [15].

In this study, Goh et al. used 2 different datasets to detect spam URLs. In this study, the researchers obtained accuracy values by using different classifiers. In this study, the most successful results were obtained in the RF classifier and these accuracy values were 93.7% and 85.2% in each dataset, respectively [16].

Sun et al. used three different datasets for URL classification in their study. In this study, they used different machine learning techniques for spam detection. These nine machine learning techniques they use are frequently used in the literature. They obtained the highest F-measure value in the RF classifier in the first dataset and the C5.0 classifier in the second and third datasets. These values are 82.19%, 87.48% and 91.90%, respectively [17].

1.2.Contributions and Innovation

NLP has become one of the most popular topics in information technology in recent years. Because the developing technology has brought large amounts of digital data with it, it is of great importance to process these data and draw meaningful conclusions from them. It is difficult for machines to process data, especially in natural languages. In this study, URLs that put users and companies in a difficult situation regarding cyber security have been identified. In this study, for the classification of URLs, the data was first prepared in a way that the classifiers could understand. At this stage, data cleaning and editing steps are available. This is a step that closely concerns the performance of the models. After this step, the bag of words matrix was obtained. 9 different machine learning methods were used to classify the URLs in the dataset. Performance metrics obtained in 9 different classifiers are discussed in detail.

1.3.Flow of Paper

Organization of the paper; In the first part, general information and related studies are given, and in the second part, the background part is provided. This section examines the dataset, data cleaning methods, and classifiers used in the study. In the third section, the experimental results and the last section, the results are discussed.

2. BACKGROUND

In this section, the dataset used in the study was examined, the data cleaning and data preparation stages were detailed and the techniques used in the study were discussed. A summary representation of the proposed model is given in Figure 1.

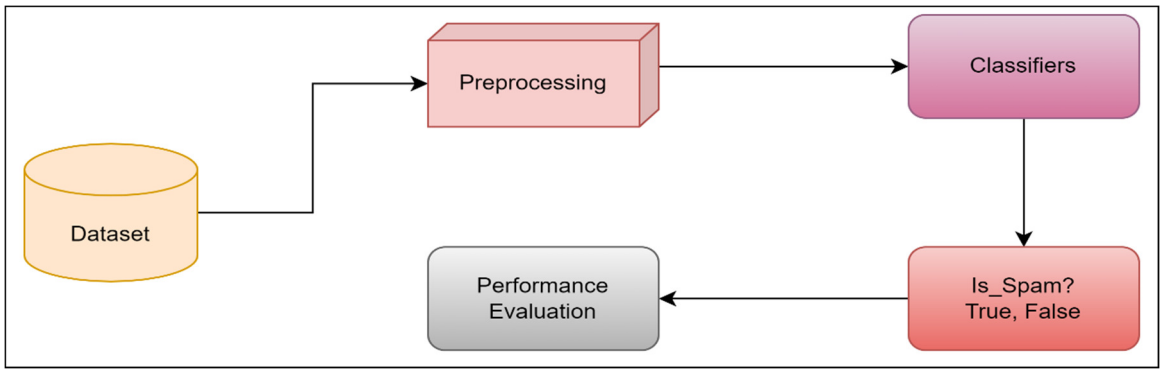

Figure 1. Summary of the proposed model

2.1. Dataset

The dataset used in the study was taken from Kaggle. This dataset consists of 148303 data in total. 101021 of these data are not malicious spam URLs [18]. The remaining 47282 are spam URLs. Sample data from the dataset are given in Figure 2.

index	url	is_spam
0	https://briefingday.us8.list-manage.com/unsubscribe	true
1	https://www.hvper.com/	true
2	https://briefingday.com/m/v4n3i4f3	true
3	https://briefingday.com/n/20200618/m#commentform	false
4	https://briefingday.com/fan	true

Figure 2. Example URLs from the dataset

Machine learning methods cannot do URL classification directly on the text. This dataset, which consists of URLs, must first be prepared in a format that machine learning methods can understand. In the study, the steps in Figure 3 were performed before the data were classified, and the data were prepared in a format that the models could understand. This data cleaning and data preparation process greatly impacts the performance of the models.

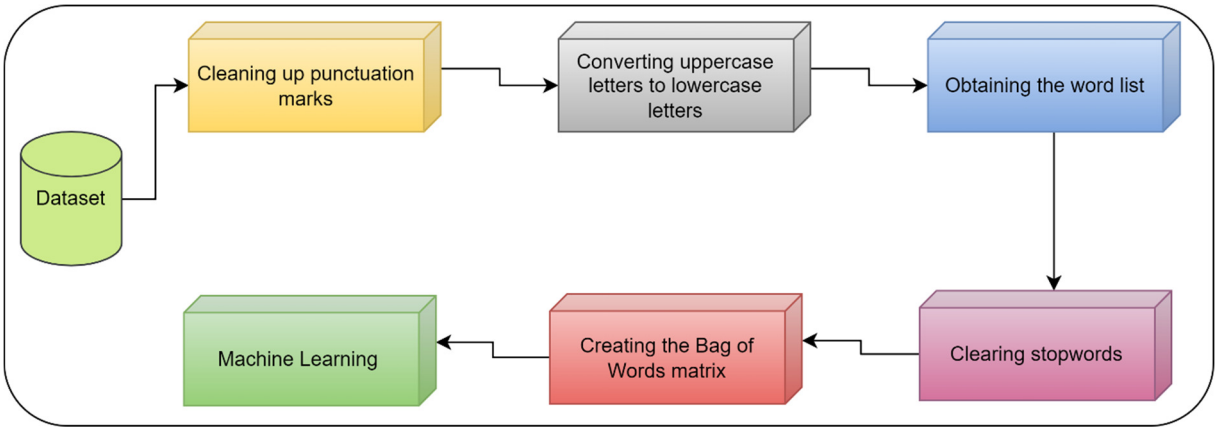

Figure 3. Data cleaning and preparation steps

In this study, before determining whether the URLs in the dataset are spam, the unstructured data in the dataset should be provided in a structured form that can be understood by machine

learning classifiers [19]. Data mining basic preprocessing should be applied to the URLs in the dataset at this stage. In these pre-processing steps, punctuation marks, Html tags, numeric expressions, and stop words are extracted from the data, and operations such as upper and lower case conversion are applied to the data. After these stages, keywords are obtained and included in the word bag called the bag of words. In this way, the most repetitive 1000 words among the words in the whole dataset are determined as keywords and a 1x1000 matrix is formed for each data in the dataset. When the entire dataset is considered complete, a 143000x1000 matrix is formed called the document term matrix (DTM). It was created by giving 1 if the keywords are included in the matrix and 0 if they are not. Thus, the dataset will be transformed into a structured form that classical machine learning classifiers can understand.

2.2. Machine Learning Techniques

Parallel to the rapid development of technology, the amount of data kept in databases also increases. For these data stored in datasets to make sense, they must be processed. These processed data can be used in different ways in different places. Health, economy, finance, agriculture, and cyber security are just a few areas. Processing and analyzing the large amount of raw data stored in databases is quite difficult with traditional database systems. Machine learning is the set of methods and algorithms necessary for processing and analyzing data. It is possible to develop a problem-specific model in machine learning.
In this study, 9 different methods accepted in the literature were used while determining the URL. In these methods, the models are first trained with the training data. Thanks to this training data, the learning process is realized. Then, when new inputs come to the trained network, the network is asked to produce the result closest to the desired value. In this way, it is aimed to place the new entry in the correct class. After the models used in the study were trained with the training data, the models were tested with the test data. Classifier and methods used in the study k-nearest neighbors(KNN) [20], Random Forest(RF) [21], Naive-Bayes (NB) [22], Gradient Boosting (GB) [23], Discriminant Analysis [24], LightGBM [25], Logistic Regression [26], XgBoost [27], Support Vector Machine (SVM) [28].

3. EXPERIMENTAL RESULTS

This study for URL classification was carried out in a Python environment. In the study, confusion matrices obtained in different models were given separately and compared. A confusion matrix is a table often used to describe the performance of a classification model on a set of test data for which the actual values are known [29]. An example confusion matrix is given in Figure 4.

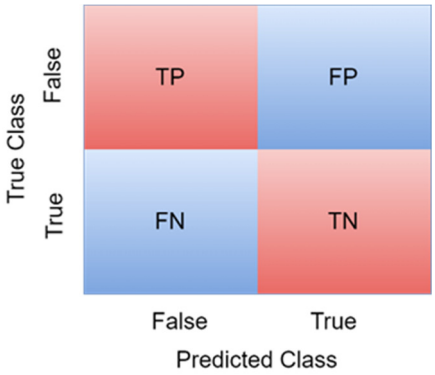

Figure 4. Confusion matrix example

In the study, 9 different parameters were used to compare the performance of the models [30]. These parameters and their formulas are given in Table 1.

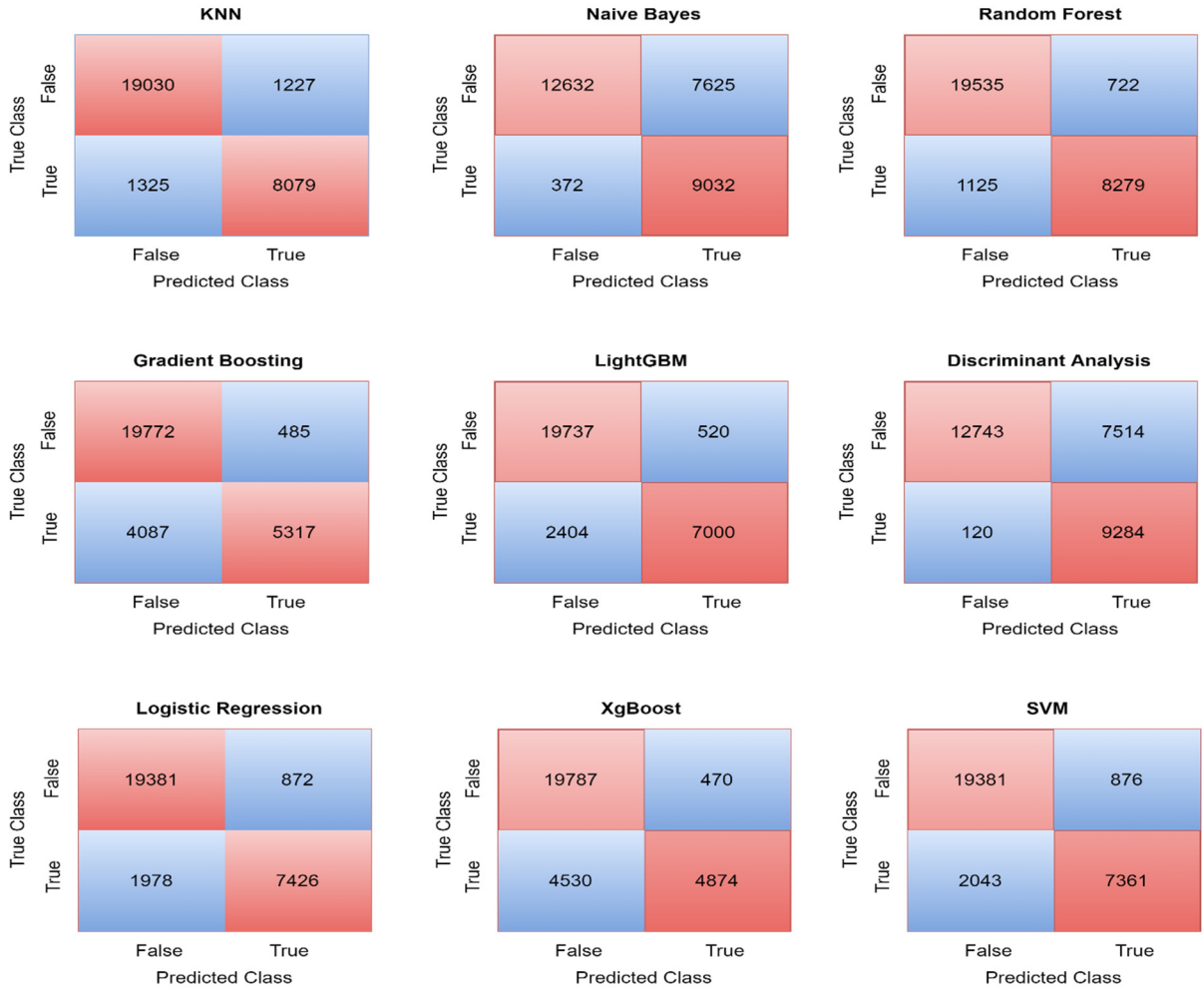

Figure 5. Confusion matrices obtained from the models

The accuracy values obtained in the machine learning methods while determining the URL are given in Table 2.

Table 1. Performance Measurement Parameters

F1-Score	**Accuracy**	**Specificity**	**Sensitivity**
F1=2TP/(2TP+FP+FN)	Acc = (TP+TN) / (Total)	Spc= TN / (FP+TN)	Sens = TP / (TP+FN)
Precision	**FPR**	**FNR**	**FDR**
PPV = TP / (TP+FP)	FPR = FP / (FP + TN)	FNR = FN / (FN+TP)	FDR= FP / (FP + TP)

In the study, to be able to classify URLs, first of all, the data in the dataset was cleaned and converted into a format that the models could understand. The Bag of Words matrices obtained in the last step of this process became the input values for the models. Confusion matrices obtained in 9 different models used in the study are given in Figure 5.

Table 2. Accuracy values(%)

KNN: 91.39	Naive Bayes: 73.03	Random Forest: **93.77**
Gradient Boosting: 84.58	LightGBM 90.14	Discriminant Analysis: 74.26
Logistic Regression: 90.39	XgBoost: 83.14	SVM: 90.15

When Table 2 is examined, it is seen that the highest accuracy value was obtained with 93.77% in the Random Forest classifier and the lowest accuracy value was obtained with the Naive Bayes classifier with 73.03%. When the confusion matrix in Figure X obtained from the Random Forest classifier is examined, it is seen that the Random Forest classifier classified 27814 of the 29661 test URLs correctly and misclassified the 1847 URLs. Other performance metrics obtained in the Random Forest classifier are given in Table 3.

Table 3. Performance metrics of Random Forest classifier(%)

F1-Score	**Accuracy**	**Specificity**	**Sensitivity**
F1= 95.49	Acc = 93.77	Spc = 91.98	Sens = 94.55
Precision	**FPR**	**FNR**	**FDR**
PPV = 96.44	FPR = 8.02	FNR = 5.45	FDR= 3.56

The proposed model is compared with similar studies in the literature in Table 4.

Table 4. Similar studies on URL classification

Study	**Year**	**Methods**	**Accuracy**
Do Xuan [12]	2020	Classifiers (SVM)	90.70%
Patgiri [13]	2019	Classifiers (SVM)	90.14%
Jain [14]	2018	SVM	90%
Joshi [15]	2019	Machine Learning	92%
Goh [16]	2015	Classifiers	85.2%-93.7%
Sun [17]	2020	Classifiers	82.9%,87.48%,91.90%
This Study	2022	Machine Learning Classifiers (RF)	93.77%

When Table 4 is examined, it is seen that the proposed model has either better or similar results than similar studies in the literature. Therefore, it is seen that the proposed model can be used in spam URL detection.

4. CONCLUSION

URL classification was made using the Bag of Word matrix in the study. Spam URLs leave users and companies in a complicated situation. These spam URLs are one of the most dangerous issues in cybersecurity. Also, spam URLs are used in fraud. To detect spam URLs, 9 different machine learning methods were used in the study. Among these methods, it has been seen that the most successful method is Random Forest. This study guides both machine learning researchers in academia and professionals and practitioners in the cyber security industry. It is among our aims to train the study with CNN, LSTM style models by using

different document matrix extraction methods. In addition, this is one of the limitations of our study.

Acknowledgments
Thank you to the researchers for sharing their datasets.

Declaration of Competing Interest
There is no conflict of interest.

Author Contribution
Muhammed Yıldırım contributed 100% at every stage of the article.

REFERENCES

[1] Adam, E.E.B., Deep learning based NLP techniques in text to speech synthesis for communication recognition. Journal of Soft Computing Paradigm (JSCP), 2020. 2(04): p. 209-215.

[2] Rajput, A., Natural language processing, sentiment analysis, and clinical analytics, in Innovation in Health Informatics. 2020, Elsevier. p. 79-97.

[3] Arthur, M.P., Automatic source code documentation using code summarization technique of NLP. Procedia Computer Science, 2020. 171: p. 2522-2531.

[4] Widyassari, A.P., et al., Review of automatic text summarization techniques & methods. Journal of King Saud University-Computer and Information Sciences, 2020.

[5] Nemes, L., A. Kiss, Social media sentiment analysis based on COVID-19. Journal of Information and Telecommunication, 2021. 5(1): p. 1-15.

[6] Neysiani, B.S., S.M. Babamir. Effect of Typos Correction on the validation performance of Duplicate Bug Reports Detection. in 10th International Conference on Information and Knowledge Technology (IKT), Tehran, Iran. 2020.

[7] Rivera-Trigueros, I., Machine translation systems and quality assessment: a systematic review. Language Resources and Evaluation, 2021: p. 1-27.

[8] Popovski, G., B.K. Seljak,, T. Eftimov, A survey of named-entity recognition methods for food information extraction. IEEE Access, 2020. 8: p. 31586-31594.

[9] Lai, C.-M., H. Shiu Jr, J. Chapman, Quantifiable Interactivity of Malicious URLs and the Social Media Ecosystem. Electronics, 2020. 9(12).

[10] Chen, Q., et al. Detecting filter list evasion with event-loop-turn granularity javascript signatures. in 2021 IEEE Symposium on Security and Privacy (SP). 2021. IEEE.

[11] Thanaki, J., Python natural language processing. 2017: Packt Publishing Ltd.

[12] Do Xuan, C., H.D. Nguyen, and T.V. Nikolaevich, Malicious URL detection based on machine learning. International Journal of Advanced Computer Science and Applications, 2020. 11(1).

[13] Patgiri, R., et al. Empirical study on malicious URL detection using machine learning. in International Conference on Distributed Computing and Internet Technology. 2019. Springer.

[14] Jain, A.K., B. Gupta, PHISH-SAFE: URL features-based phishing detection system using machine learning, in Cyber Security. 2018, Springer. p. 467-474.

[15] Joshi, A., et al., Using lexical features for malicious URL detection--a machine learning approach. arXiv preprint arXiv:1910.06277, 2019.

[16] Goh, K.L., A.K. Singh, Comprehensive literature review on machine learning structures for web spam classification. Procedia Computer Science, 2015. 70: p. 434-441.

[17] Sun, N., et al., Near real-time twitter spam detection with machine learning techniques. International Journal of Computers and Applications, 2020: p. 1-11.

[18] URL-1,https://www.kaggle.com/shivamb/spam-url-prediction, Last Accessed Date: 01.01.2022.

[19] Bingol, H., B. Alatas. Rumor Detection in Social Media using machine learning methods. in 2019 1st International Informatics and Software Engineering Conference (UBMYK). 2019. IEEE.

[20] Zhang, M.-L., Z.-H. Zhou, ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 2007. 40(7): p. 2038-2048.

[21] Pal, M., Random forest classifier for remote sensing classification. International journal of remote sensing, 2005. 26(1): p. 217-222.

[22] Rish, I. An empirical study of the naive Bayes classifier. in IJCAI 2001 workshop on empirical methods in artificial intelligence. 2001.

[23] Friedman, J.H., Stochastic gradient boosting. Computational statistics & data analysis, 2002. 38(4): p. 367-378.

[24] Klecka, W.R., G.R. Iversen, and W.R. Klecka, Discriminant analysis. Vol. 19. 1980: Sage.

[25] Ke, G., et al., Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 2017. 30.

[26] Wasserman, S., P. Pattison, Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. Psychometrika, 1996. 61(3): p. 401-425.

[27] Chen, T., C. Guestrin. Xgboost: A scalable tree boosting system. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.

[28] Suykens, J.A., J. Vandewalle, Least squares support vector machine classifiers. Neural processing letters, 1999. 9(3): p. 293-300.

[29] Eroglu, Y., et al., Diagnosis and grading of vesicoureteral reflux on voiding cystourethrography images in children using a deep hybrid model. Computer Methods and Programs in Biomedicine, 2021. 210: p. 106369.

[30] Yildirim, M., A. Çinar,, E. Cengİl. Classification of flower species using CNN models, Subspace Discriminant, and NCA. in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). 2021. IEEE.

NATURENGS
MTU Journal of Engineering and Natural Sciences
https://dergipark.org.tr/tr/pub/naturengs
DOI: 10.46572/naturengs.1100358

Research Article

Sarcasm Detection in Online Social Networks Using Machine Learning Methods

Harun BİNGÖL[1*], Muhammed YILDIRIM[2]

[1]Department of Software Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, Malatya, Turkey.
[2]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, Malatya, Turkey.

(**Received:** 08.04.2022; **Accepted:** 13.05.2020)

ABSTRACT: Our lives have completely changed since the internet came into our lives. Role models for people are the people around them and people all over the world. Although there are positive aspects to this situation, we will deal with the negative aspects in this study. One of these negative aspects is that people share their ideas on social networks without supervision. In this way, people who use social networks are told offensive words by people they do not know in real life. Sometimes these words are not directly insulting, but they are expressed sarcastically and annoy the interlocutor. In this study, detecting sarcastic words in social networks is considered a classification problem. Since the data type used in the proposed method is text-based, both text mining and machine learning methods are used together. In this study, the sarcastic word classification process was carried out using a dataset obtained from the Twitter social network, which includes two public classes. The performance of the proposed method was obtained with the Random Forest algorithm with an accuracy of 94.9%.

Keywords: Social networks, Sarcasm detection, Text mining, Classification.

1. INTRODUCTION

Personal computers began to enter our lives in the 1980s. About 10 years later, in the 1990s, the use of the internet started to become widespread. The development of the internet is currently divided into three phases. Web 1.0, Web 2.0 and finally, Web 3.0 technologies. When Web 1.0 technology first entered our lives, people only existing accessed content, in other words, it is the most primitive internet technology. With the start of Web 2.0 technology in the 2000s, many applications were developed, from personal blog pages to social networks. The most popular social networks are Facebook, Twitter, Instagram and TikTok. Thanks to these technologies, people can comment on any photo, create their content and upload it to the internet, allowing all people to access this content.

In addition to facilitating access to information, this event also caused a parabolic increase in the amount of data on the internet. It also allowed people to increase their social interaction. Thanks to the internet, it has become effortless and ordinary to buy any product, share the negative aspects of this product, and influence other people. Even the fact that people who are far between continents and cannot see each other physically meet and marry thanks to this technology does not surprise anyone. Web 3.0 technology, on the other hand, can be described

*Corresponding Author: harun.bingol@ozal.edu.tr
ORCID number of authors: [1]0000-0001-5071-4616, [2] 0000-0003-1866-4721

as the interpretation of data produced by Web 2.0 technology by computer systems. In this technology, also known as logical web technologies, it is possible to personalize people according to their frequency and internet use habits, thanks to algorithms developed using machine learning and artificial intelligence methods. The primary motivation for the progress of all these technologies is the need for access to information. Social networks that have entered our lives with Web 2.0 have made accessing information much easier. However, this situation also brought with it some undesirable negative aspects.

The first of these is the attack on personal rights. This is a crime. But people are not aware that they are committing a crime because they commit this crime on virtual platforms. Thinking that these crimes committed on online social networks will go unpunished, they continue their insults and humiliating innuendo without any boundaries. At the beginning of these crimes are insults, humiliation, swearing, phishing, fake news, mocking and sarcastic remarks. It is essential to prevent these undesirable situations [1]. The scientific world is constantly working to prevent crimes committed on the internet. If an intelligent system can be created, such attacks on personal rights can be prevented before they are transmitted to other people, thanks to artificial intelligence working in the background. Thus, a crime is prevented at the source before it is committed. Detecting the sarcastic word is a crucial step in sentiment analysis, considering the prevalence of sarcasm in emotional texts and the difficulties of detecting it [2]. This study used a dataset obtained from Kaggle [3], which was created using the online social networking platform Twitter data, to detect the sarcastic word. This dataset was first preprocessed by text mining and then classified by machine learning methods.

Many scientific studies continue to solve the problem of sarcastic word detection in social networks. The performances of the methods they have proposed in these scientific studies have been measured by many performance evaluation metrics [4]. The performance of the method we proposed in this study was evaluated using accuracy, precision, recall, and f_measure metrics.

The organization of this study is as follows: Chapter 2, scientific studies on the sarcastic word detection problem were examined. In Chapter 3, the proposed model for sarcastic word detection, text classification steps used during the experiments and machine learning methods are mentioned. In Chapter 4, the results of the experimental studies are given, and finally, the results and future work are shown in Chapter 5.

2. RELATED STUDIES

It is a fairly new problem that has attracted the attention of the scientific world. There are many studies on the detection of many sarcastic words in the literature. This section describes some of the research approaches and their results for detecting current sarcasm.

Campel and Katz state that sarcasm occurs in many different dimensions, such as unsuccessful expectations, pragmatic insincerity, negative tension, the presence of the victim, and along with stylistic components such as emotional words [5].

Joshi et al. presented a computational system that makes use of context incompatibility to detect sarcastic words. They classified two types of incompatible features as explicit and implicit. They also stated that they proposed a method that reveals the inconsistency between sentences. They said that a 10-20% F_measure value was obtained for the success of the method they offered in the study [6].

Riloff et al. developed a system to identify sarcastic words in tweets. They stated they developed a new pre-loaded algorithm that automatically learns positive and negative situations from sarcastic tweets [7].

Ghosh et al., by looking at a specific context type they have done, stated that they provided a complementary contribution to the modeling context studies available for the detection of sarcastic words. In their study, they sought answers to two questions. First, does modeling the speech context help with sarcasm detection? Second, can it be determined which part of the speech context triggers the sarcastic response? In response to the first question, they noted that Long Short-Term Memory (LSTM) networks, which can model both the context and the sarcastic response, outperform LSTM networks that only read the response. In response to the second question, they stated that an evaluation of the attention weights produced by the LSTM models was made. They emphasized that attention-based models can describe sarcastic speech characteristics [8].

Mishra et al., in their study, tried to observe the difference in the behavior of the eye when reading Sarcastic and Non-Sarcastic sentences. Starting from this observation, the cognitive features obtained from the eye movement data of the reader and the linguistic and stylistic features for the detection of sarcastic words were examined. They stated that they made statistical classification using the advanced feature set obtained. Augmented cognitive features improved their sarcastic remark detection (in terms of the F_measure metric) by 3.7% compared to the performance of the best-reported system [9].

3. SARCASM DETECTION MODEL

The representation of the data used in systems designed with artificial intelligence algorithms directly affects the stability of the system and the performance achieved. If the studied data is text-based, it must be converted to an appropriate representation. This is why basic text mining operations are so important. Since the tweets used in this study are text-based, text mining operations were carried out as preprocessing methods to extract useful information from the text. The flow chart for the proposed sarcastic word detection model is given in Figure 1.

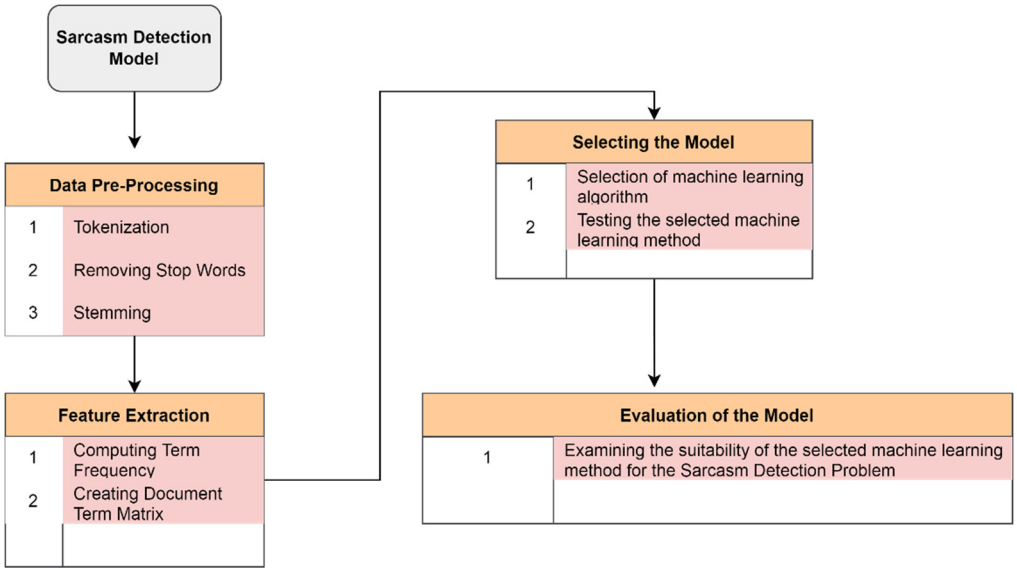

Figure 1. Flowchart of the sarcasm detection model

3.1. Data Preprocessing Steps

Text mining is a sub-branch of natural language processing. This application takes its place in natural language processing when viewed from the top. Like data mining, text mining is based on implementing some basic steps required to access information from raw data. Since the data type used during the experiments is text-based, text mining is done. The preprocessing applied to the raw tweet data during the experiments is shown in Table 1.

Table 1. Basic Preprocessing Steps.

Input: Entering textual data
Output: Preprocessed textual data
1. Extraction of numerical expressions from textual data
2. Removing punctuation marks from textual data
3. Removing texts with less than N characters
4. Applying the case converter to text
5. Removing stop words from the text
6. Rooting of textual data

3.1.1. Removing Numeric Expressions and Punctuation Marks

In this preprocessing step, numerical expressions, spaces and punctuation marks are removed from the text-based data and divided into small pieces called tokens. Here, the term token represents a useful semantic unit for processing data within a document [10]. It is the basic step for teaching words to machines.

3.1.2. N-Character Optimization and Case Conversion

Some words used in English are not keywords that help us solve the sarcastic word detection problem. These words need to be removed from the dataset. The number of N characters can be predetermined. In this study, N=3 was determined. In addition, integrity is important in the data to be analyzed, so all data is converted to uppercase or lowercase letters. In this study, all data has been converted to lowercase.

3.1.3. Removing Stop Words

Stop words are words that don't convey any information. Stop words include conjunctions and pronouns. In English, there are nearly 500 stop words. A the, by, of, while, did, that, on, afterward, once, and before are examples of these terms. [11].

3.1.4. Finding Root

In this process step, words are to reach their roots. It is the step of obtaining root states freed from their inflected state. The goal of rooting is to find the core forms of words with similar meanings but diverse word forms [11].

3.2. Feature Extraction, Selection, and Suggested Model

The common problem in data mining, text mining, image processing, and all other data analysis studies is the excess data size. When the data size is large, processing and storing this data takes a lot of time and requires the extra cost to store it. In addition, more processor and memory elements are required to process this data. For these reasons, it is important to remove residual and unnecessary features from the dataset that will not affect the model's accuracy. In this study,

a feature selection method selects root terms whose frequency is greater than the predetermined threshold value in the datasets. Each document's terms in the dataset are weighted, and the document is turned into a vector of term weights. Vector Space Model is the name for this type of representation (VUM). Each word in VUM is represented with a value representing the word's weight in the document. In this study, Term Frequency (TF) method was used to calculate the weights [12].

Binary Vectors: Data containing text in the dataset are represented as 0's and 1's.

TF: It refers to the number of repetitions of word roots in the data as shown in Equation (1).

$$TF_{ij} = \frac{n_{ij}}{|d_i|} \quad (1)$$

di, i'. is the sum of all terms in the document. *nij* is *i'*. *j'* in the document. is the number of words. After calculating the TF value for each word of the document, a Document Term Matrix (DTM) is created using the weights of the words. This matrix is the m x n matrix. In DTM, each row represents the documents, each column indicates the term, and each cell shows the weight of the terms in the document [11]. The DTM used during the experiments is shown in Table 2.

Table 2. Document Term Matrix (DTM).

	T_1	T_2	…	T_n
D_1	W_{11}	W_{12}	…	W_{1n}
D_2	W_{21}	W_{22}	…	W_{2n}
…	…	…	…	…
D_m	W_{m1}	W_{m2}	…	W_{mn}

3.3. Machine Learning Algorithms Used in Experiments

The first of the machine learning algorithms used in the experiments is the J48 algorithm. This algorithm is the WEKA-edited version of the C4.5 decision tree algorithm. Decision trees are frequently used in regression and classification problems in the literature. A decision tree is one of the most preferred supervised learning methods. In this learning method, a learning set is first created. In this algorithm, rr is the label denoted by the name of an educational status class. Decision trees consist of roots, nodes and leaves [13].

Another of the machine learning methods we use in our experiments is Naive Bayes (NB), a statistical classification method. NB classification is based on Bayes' theorem in statistics. The probability that the available data belong to the determined classes is evaluated. It includes the logic of probabilistic calculation of the effect of each criterion of the data on the result. The NB method is frequently used in the literature due to its simplicity and simplicity compared to other classification algorithms [14].

Another regression analysis used in the study is logistic regression. It is one of the basic fields of statistics. Regression analysis can be defined as predicting the behavior of a random variable using a model. Here, the relationships between dependent and independent variables are examined. Thanks to this relationship between the variables, modeling or estimation are performed. There is more than one type of regression used in statistics. Logistic regression was used in this study [15].

Another machine learning method used in the study is the Random Forest (RF) method. The working logic of the RF method is similar to decision trees. As the name suggests, the random forest consists of a large number of individual decision trees that work as a community. In the RF method, the tree with the highest votes is used. In this way, it is possible to obtain high accuracy values in this method [16].

Yoav Freund and Robert Schapire developed the adaBoost algorithm, another machine learning method, [17]. An AdaBoost algorithm, one of the first applications of the Boosting method, is based on the ensemble learning technique [18]. Unlike other models, this method allows many models to be trained to solve the related problem. The results obtained in these models are then combined for classification or regression. It is possible to get more successful results as more than one model is trained.

4. DATASET, METRICS, AND EXPERIMENTAL RESULTS

This section gives the dataset used in the study, performance measurement metrics, and experimental results.

4.1. Dataset

The dataset used during the experiments is a publicly available two-class text-based dataset. The examples of sarcastic words used in the dataset were taken from www.theonion.com. Samples of non-sarcastic words are taken from www.huffpost.com, www.bbc.com, www.foxnews.com, www.aljazeera.com, and finally www.euronews.com. The original dataset used in the experiments contains 12506 data and 52% of this data is made up of sarcastic words and 48% of it is non-sarcastic [3]. However, 1475 pieces of data were randomly selected during the experiments. Of these chosen data, 722 were determined as sarcastic and 753 as non-sarcastic. A section from the dataset is shown in Figure 2.

Row ID	S Text	S State
1	CNN Triumphs (At Least in Most Demographic Categories)	Non Sarcastic
2	You Did The Best You Could Says Iron Man Action Figur...	Sarcastic
3	New Emails Reveal Warm Relationship Between Kamala ...	Non Sarcastic
4	Donald Trump Jr. Gets Slammed Over Racist Birtherism ...	Non Sarcastic
5	God Urges Rick Perry Not To Run For President	Sarcastic
6	Global Aid Pours into Haiti	Non Sarcastic
7	CNN Anchor Calls Obama Protester 'Rude' And 'Crazy'	Non Sarcastic
8	Federal Prisons Reinstitute Executions By Lethal Inflation	Sarcastic
9	Lou Dobbs Crumbles When Pressed On His 'NAFTA Sup...	Non Sarcastic
10	CNN Still Bent On Debating 'Two Sides' Of The Confeder...	Non Sarcastic
11	'Fox & Friends' Guest Says CNN Partly To Blame For Las...	Non Sarcastic
12	An Open Letter to CNN Regarding Nancy Grace	Non Sarcastic
13	Conservative Pundits: GOP Primary Lacks Foreign Policy...	Non Sarcastic
14	Soldier Back Home From Serving At Mexico Border Still H...	Sarcastic
15	Decades Of Breathing Really Starting To Catch Up With...	Sarcastic
16	U.S.â€"Cuba Relations End After Obama Hit By Foul Bal...	Sarcastic
17	Stanford Students Admit It Was Pretty Obvious Billionai...	Sarcastic
18	LA's Winners Of The Week June 11	Non Sarcastic
19	CNN's Ed Henry: McCain Playing Politics On Bailout	Non Sarcastic
20	Overwhelmed Archaeologists Struggling To Keep Pace ...	Sarcastic
21	Study Finds Majority Of Accidental Heroin Overdoses C...	Sarcastic
22	Deli Worker Searches For Palest Mealiest Tomato To Pu...	Sarcastic
23	Cancer Researchers Develop Highly Promising New Pink...	Sarcastic
24	Kamala Harris Breaks Down Trump's â€"Predatorâ€™ Pl...	Non Sarcastic
25	California's Death-Penalty Regime Ruled Unconstitution...	Non Sarcastic

Figure 2. A snippet of the sarcastic promise dataset.

4.2. Performance Evaluation Metrics

While detecting the proposed method for sarcasm detection in online social networks with machine learning algorithms, some metrics that are frequently used in the literature are used in this application to measure the proposed method's performance [19,20]. The metrics used in the needling word detection problem are shown in Table 3.

Table 3. Metrics

Performance Evaluation Metrics	Formulas
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{FN + TP}$
F-Measure	$\frac{2 * \left(\frac{TP}{TP + FP}\right) * \left(\frac{TP}{TP + FN}\right)}{\left(\frac{TP}{TP + FP}\right) + \left(\frac{TP}{TP + FN}\right)}$

4.3. Experimental Results

In this study, the detection of sarcastic words is considered a classification problem. In this study, the sarcastic promise dataset was tested using different parameters on machine learning algorithms and their performances were compared.

The dataset was tested on seven different machine learning algorithms for sarcastic word detection. The entire dataset is set as training and tests on algorithms. Standard versions of machine learning algorithms were used during the experiments. The results obtained in this application are shown in Table 4. The graph of the values given in Table 4 is given in Figure 3.

Table 4. Performance values of machine learning algorithms (100% training).

	Machine Learning Algorithms						
	J48	Filtered Classifier	Naïve Bayes	Random Forest	Logistic Regression	JRip	AdaBoostM1
Accuracy	0.741	0.761	0.814	**0.949**	0.893	0.752	0.660
F-Measure	0.775	0.795	0.823	**0.947**	0.890	0.788	0.737
Recall	0.659	0.672	0.751	**0.937**	0.871	0.664	0.583
Precision	0.942	0.973	0.912	0.957	0.910	0.969	**0.998**

When Table 4 is examined, it is seen that the Random Forest algorithm gives the highest accuracy value with 94.9%. In addition, it has been determined that this algorithm has higher performance in F_measure and Sensitivity criteria than the other 6 machine learning methods. Only the precision metric is higher in the AdaBoostM1 algorithm than all other machine learning methods.

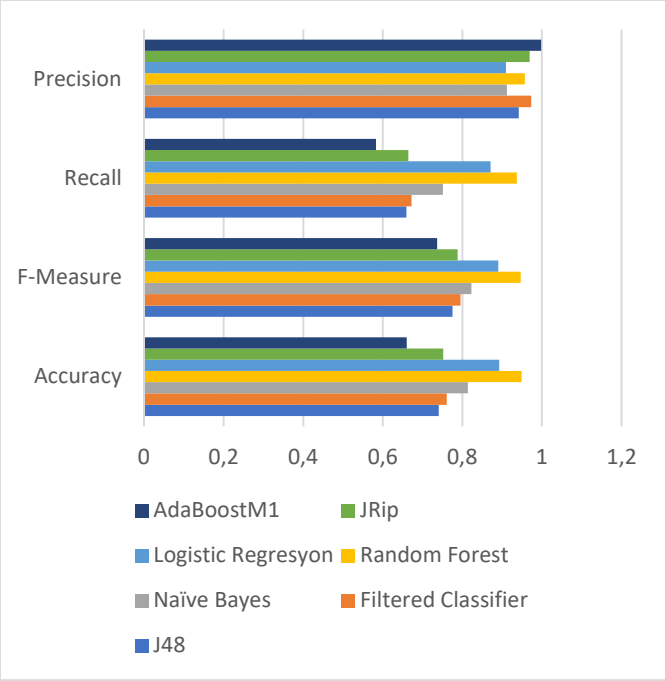

Figure 3. Performance results of algorithms (100% training)

The dataset was tested on seven different machine learning algorithms for sarcastic detection. The dataset is set as 70% training and 30% testing and tests on algorithms. Standard versions of machine learning algorithms were used during the experiments. The results of this application are shown in Table 5. The graph of the values given in Table 5 is given in Figure 4.

Table 5. Performance values of machine learning algorithms (70% training, 30% testing)

	Machine Learning Algorithms						
	J48	Filtered Classifier	Naïve Bayes	Random Forest	Logistic Regression	JRip	AdaBoostM1
Accuracy	0.695	0.727	0.787	**0.830**	0.735	0.690	0.654
F-Measure	0.748	0.772	0.803	**0.817**	0.773	0.749	0.731
Recall	0.613	0.638	0.712	**0.831**	0.650	0.606	0.578
Precision	0.960	0.977	0.920	0.806	0.954	0.983	**0.994**

When Table 5 is examined, it is seen that the Random Forest algorithm gives the highest accuracy value with 83%. In addition, it has been determined that this algorithm has higher performance in F_measure and Sensitivity criteria than the other 6 machine learning methods.

Only the precision metric is higher in the AdaBoostM1 algorithm than all other machine learning methods.

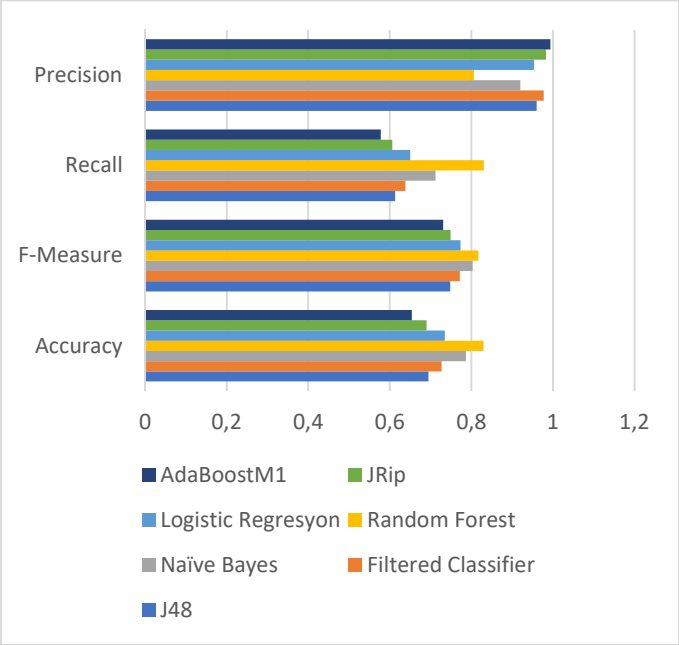

Figure 4. Performance results of algorithms (70% training, 30% testing)

The dataset was tested on seven different machine learning algorithms for sarcastic word detection. The dataset cross-validation is set as 5-fold cross-validation and tests on algorithms. Standard versions of machine learning algorithms were used during the experiments. The results of this application are shown in Table 6. The graph of the values given in Table 6 is given in Figure 5.

Table 6. Performance values of machine learning algorithms (5-fold Cross Validation).

	Machine Learning Algorithms						
	J48	Filtered Classifier	Naïve Bayes	Random Forest	Logistic Regression	JRip	AdaBoostM1
Accuracy	0.707	0.734	0.803	**0.817**	0.740	0.726	0.645
F-Measure	0.757	0.777	**0.815**	0.806	0.778	0.770	0.728
Recall	0.625	0.647	0.738	**0.815**	0.655	0.642	0.573
Precision	0.961	0.973	0.910	0.796	0.959	0.963	**0.998**

When Table 6 is examined, it is seen that the Random Forest algorithm gives the highest accuracy value with 81.7%. The Random Forest algorithm gave better results than other algorithms in terms of sensitivity metrics. In addition, the Naive Bayes algorithm showed the highest performance in terms of the F_measure metric. The precision metric is higher in the AdaBoostM1 algorithm than in all other machine learning methods.

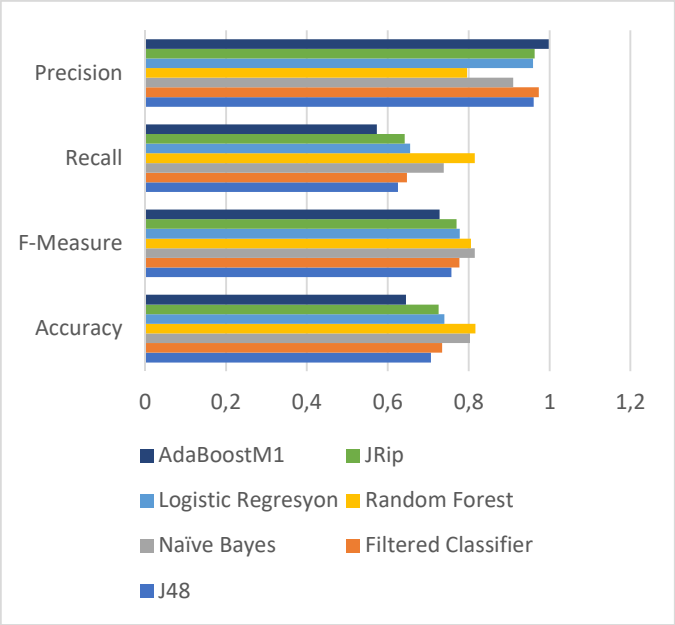

Figure 5. Performance results of algorithms (5-fold Cross Validation)

The dataset was tested on seven different machine learning algorithms for sarcastic word detection. The dataset cross-validation is set as 10-fold cross-validation and tests on algorithms. Standard versions of machine learning algorithms were used during the experiments. The results of this application are shown in Table 7. The graph of the values given in Table 7 is given in Figure 6.

Table 7. Performance values of machine learning algorithms (10-fold Cross Validation).

	Machine Learning Algorithms						
	J48	Filtered Classifier	Naïve Bayes	Random Forest	Logistic Regression	JRip	AdaBoostM1
Accuracy	0.811	0.812	**0.852**	0.822	0.742	0.731	0.642
F-Measure	0.817	0.806	**0.850**	0.814	0.779	0.773	0.726
Recall	0.758	**0.878**	0.822	0.811	0.659	0.646	0.571
Precision	0.886	0.744	0.879	0.816	0.952	0.961	**0.998**

When Table 7 is examined, it is seen that the Naive Bayes algorithm gives the highest accuracy value with 85.2%. The naive Bayes algorithm gave better results than other algorithms in terms of the F_measure metric. In addition, the Filtered Classifiers algorithm showed the highest performance in terms of the Sensitivity metric. The precision metric is higher in the AdaBoostM1 algorithm than in all other machine learning methods.

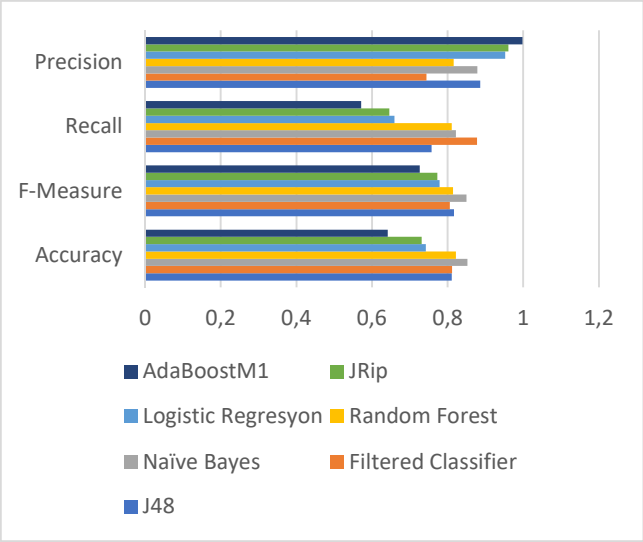

Figure 6. Performance results of algorithms (10-fold Cross Validation)

5. CONCLUSION

Since the internet is an integral part of our lives, people using this technology have increased rapidly. In addition, many positive and negative reflections of the internet on human life appear as an inevitable reality. In this study, using text mining techniques, which is a sub-branch of data mining, a method is proposed to detect sarcastic words in online social networks with machine learning methods on the public dataset. In this study, the detection of sarcastic words is considered a classification problem. The performances of machine learning methods were evaluated using four experiments in terms of accuracy, precision, precision, and F_measure metrics.

Considering the observations of the 4 experiments we have made for the proposed method, the highest accuracy values have been obtained with the Random Forest algorithm. After the Random Forest algorithm, the highest accuracy values were obtained with the Naive Bayes algorithm. The highest percentage values were observed in all experiments where the AdaBoostM1 algorithm was received in precision metrics. It is known that the performance of the algorithms used in the experiments varies according to the problem and dataset to be determined. Regarding future studies, the model's performance can be increased by discovering new algorithms, integrating metaheuristic optimization methods, or producing chaotic, adaptive, or hybrid versions of existing algorithms for more efficient results. Different feature extraction techniques can also be applied to improve the performance of the sarcastic word detection system in terms of many important metrics.

Acknowledgments
Thank you to the researchers for sharing their datasets.

Declaration of Competing Interest
There is no conflict of interest.

Author Contribution
Harun BİNGÖL and Muhammed YILDIRIM contributed equally at every stage of the article.

REFERENCES

[1] Baloglu, U. B., Alatas, B., Bingol, H. (2019, November). Assessment of Supervised Learning Algorithms for Irony Detection in Online Social Media. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-5). IEEE.

[2] Joshi, A., Bhattacharyya, P., Carman, M. J. (2017). Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR), 50(5), 1-22.

[3] Access Date: 06/01/2022, https://www.kaggle.com/theynalzada/news-headlines-for-sarcasm detection ?select =Data.csv

[4] Bingol, H., Alatas, B. (2019, November). Rumor Detection in Social Media using machine learning methods. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). IEEE.

[5] Campbell, J. D., Katz, A. N. (2012). Are there necessary conditions for inducing a sense of sarcastic irony?. Discourse Processes, 49(6), 459-480.

[6] Joshi, A., Sharma, V., Bhattacharyya, P. (2015, July). Harnessing context incongruity for sarcasm detection. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 757-762).

[7] Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., Huang, R. (2013, October). Sarcasm as contrast between a positive sentiment and negative situation. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 704-714).

[8] Ghosh, D., Fabbri, A. R., Muresan, S. (2017). The role of conversation context for sarcasm detection in online interactions. arXiv preprint arXiv:1707.06226.

[9] Mishra, A., Kanojia, D., Nagar, S., Dey, K., Bhattacharyya, P. (2017). Harnessing cognitive features for sarcasm detection. arXiv preprint arXiv:1701.05574.

[10] A. Mullen, L., Benoit, K., Keyes, O., Selivanov, D., Arnold, J. (2018). Fast, consistent tokenization of natural language text. Journal of Open Source Software, 3(23), 655.

[11] Ozbay, F. A., Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 540, 123174.

[12] Azam, N., Yao, J. (2012). Comparison of term frequency and document frequency based feature selection metrics in text categorization. Expert Systems with Applications, 39(5), 4760-4768.

[13] Fitri, S. (2014). Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian, Lazy-Ibk, Zero-R, Dan Decision Tree-J48. Data Manajemen dan Teknologi Informasi (DASI), 15(1), 33.

[14] Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15). Springer, Berlin, Heidelberg.

[15] Menard, S. (2002). Applied logistic regression analysis (Vol. 106). Sage.

[16] Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

[17] Freund, Y., Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.

[18] Schapire, R. E. (2013). Explaining adaboost. In Empirical inference (pp. 37-52). Springer, Berlin, Heidelberg.

[19] Baydoğan, V. C., Alataş, B. (2021). Çevrimiçi Sosyal Ağlarda Nefret Söylemi Tespiti için Yapay Zeka Temelli Algoritmaların Performans Değerlendirmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 33(2), 745-754.

[20] Ozbay, F. A., Alatas, B. (2019). A novel approach for detection of fake news on social media using metaheuristic optimization algorithms. Elektronika ir Elektrotechnika, 25(4), 62-67.