

# AJUR

American Journal of  
*Undergraduate Research*

Volume 23 | Issue 2 | June 2026

*[www.ajuronline.org](http://www.ajuronline.org)*

Print Edition ISSN 1536-4585  
Online Edition ISSN 2375-8732

# AJUR

American Journal of  
*Undergraduate Research*

---

Volume 23 | Issue 2 | June 2026 | <https://doi.org/10.33697/ajur.2026.171>

---

- 2      *AJUR History and Editorial Board*
- 3      *Exploring the Link Between Mental Health Stigma and Help-Seeking Behavior in Law Enforcement Officers*  
*Taylor Radtke & Cindy Prins*
- 13     *Clustering in High-Dimension – Tools and Challenge*  
*Lucy Liu & Haim Bar*
- 27     *Framing the News: How an Article’s Headline and Content Valence Shape Readers’ Understanding and Impressions*  
*Kyle Smoak, Kenneth Barideaux Jr., & Christa Christ*
- 41     *Physics-Informed Neural Network Frameworks for the Analysis of Engineering and Biological Dynamical Systems Governed by Ordinary Differential Equations*  
*Andrew Particka, Tyrus Whitman, Christopher Diers, Ian Griffin, Charuka Wickramasinghe, & Pradeep Ranaweera*

*American Journal of Undergraduate Research (AJUR)* is a national, independent, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. Each manuscript of AJUR receives a DOI number. AJUR is archived by the US Library of Congress. AJUR was established in 2002, incorporated as a charitable not-for-profit organization in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).

### EDITORIAL TEAM

Dr. Peter Newell, Editor-in-Chief  
Dr. Kestutis Bendinskas, Executive Editor  
Dr. Anthony Contento, Copy Editor

### EDITORIAL BOARD *by subject area*

#### ACCOUNTING

Dr. Dean Crawford,  
[dean.crawford@oswego.edu](mailto:dean.crawford@oswego.edu)

#### ART HISTORY

Dr. Lisa Seppi,  
[lisa.seppi@oswego.edu](mailto:lisa.seppi@oswego.edu)

#### BEHAVIORAL NEUROSCIENCE

Dr. Aileen M. Bailey,  
[ambailey@smcm.edu](mailto:ambailey@smcm.edu)

#### BIOCHEMISTRY

Dr. Kestutis Bendinskas,  
[kestutis.bendinskas@oswego.edu](mailto:kestutis.bendinskas@oswego.edu)

Dr. Nin Dingra,  
[ndingra@alaska.edu](mailto:ndingra@alaska.edu)

#### BIOENGINEERING

Dr. Jorge I. Rodriguez,  
[jorger@uga.edu](mailto:jorger@uga.edu)

Dr. Jessica Amber Jennings,  
[jjennings@memphis.edu](mailto:jjennings@memphis.edu)

#### BIOINFORMATICS

Dr. John R. Jungck,  
[jungck@udel.edu](mailto:jungck@udel.edu)

Dr. Isabelle Bichindaritz,  
[ibichind@oswego.edu](mailto:ibichind@oswego.edu)

#### BIOLOGY, PHYSIOLOGY

Dr. David Dunn,  
[david.dunn@oswego.edu](mailto:david.dunn@oswego.edu)

#### BIOLOGY, DEVELOPMENTAL

Dr. Poongodi Geetha-Loganathan,  
[p.geethaloganathan@oswego.edu](mailto:p.geethaloganathan@oswego.edu)

#### BIOLOGY, MICROBIOLOGY

Dr. Peter Newell,  
[peter.newell@oswego.edu](mailto:peter.newell@oswego.edu)

#### BOTANY

Dr. Julien Bachelier,  
[julien.bachelier@fu-berlin.de](mailto:julien.bachelier@fu-berlin.de)

#### BUSINESS

Dr. Mohammad Tajvarpour,  
[Mohammad.Tajvarpour@oswego.edu](mailto:Mohammad.Tajvarpour@oswego.edu)

Dr. Kathleen Locklear,  
[kathleen.locklear@oswego.edu](mailto:kathleen.locklear@oswego.edu)

#### CHEMISTRY

Dr. Alfredo Castro,  
[castroa@felician.edu](mailto:castroa@felician.edu)

Dr. Charles Kriley,  
[cekriley@gcc.edu](mailto:cekriley@gcc.edu)

Dr. Vadoud Niri,  
[vadoud.niri@oswego.edu](mailto:vadoud.niri@oswego.edu)

#### COMPUTER SCIENCES

Dr. Dele Oluwade,  
[deleoluwade@yahoo.com](mailto:deleoluwade@yahoo.com)

Dr. Mais W Nijim,  
[Mais.Nijim@tamuk.edu](mailto:Mais.Nijim@tamuk.edu)

Dr. Bastian Tenbergen,  
[bastian.tenbergen@oswego.edu](mailto:bastian.tenbergen@oswego.edu)

#### COMPUTATIONAL CHEMISTRY

Dr. Alexander Soudackov,  
[alexander.soudackov@yale.edu](mailto:alexander.soudackov@yale.edu)

#### ECOLOGY

Dr. Chloe Lash,  
[clash@ollusa.edu](mailto:clash@ollusa.edu)

#### ECONOMICS

Dr. Elizabeth Schmitt,  
[elizabeth.schmitt@oswego.edu](mailto:elizabeth.schmitt@oswego.edu)

#### EDUCATION

Dr. Charity Dacey,  
[cdacey@touro.edu](mailto:cdacey@touro.edu)

Dr. Marcia Burrell,  
[marcia.burrell@oswego.edu](mailto:marcia.burrell@oswego.edu)

#### EDUCATION, PHYSICS

Dr. Andrew D. Gavrin,  
[agavrin@iupui.edu](mailto:agavrin@iupui.edu)

#### ENGINEERING, ELECTRICAL

Dr. Michael Omidiora,  
[momidior@bridgeport.edu](mailto:momidior@bridgeport.edu)

#### ENGINEERING, ENVIRONMENTAL

Dr. Félix L Santiago-Collazo,  
[fsantiago@uga.edu](mailto:fsantiago@uga.edu)

#### FILM AND MEDIA STUDIES

Dr. Lauren Steimer,  
[lsteimer@mailbox.sc.edu](mailto:lsteimer@mailbox.sc.edu)

Dr. Ashely Young,  
[AY13@mailbox.sc.edu](mailto:AY13@mailbox.sc.edu)

#### GEOLOGY

Dr. Rachel Lee,  
[rachel.lee@oswego.edu](mailto:rachel.lee@oswego.edu)

#### HISTORY

Dr. Richard Weyhing,  
[richard.weyhing@oswego.edu](mailto:richard.weyhing@oswego.edu)

Dr. Murat Yasar,  
[murat.yasar@oswego.edu](mailto:murat.yasar@oswego.edu)

#### HONORARY EDITORIAL

#### BOARD MEMBER

Dr. Lorrie Clemo,  
[lorrie.a.clemo@gmail.com](mailto:lorrie.a.clemo@gmail.com)

#### JURISPRUDENCE

Bill Wickard, Esq.,  
[William.Wickard@KLGates.com](mailto:William.Wickard@KLGates.com)

#### KINESIOLOGY

Dr. David Senchina,  
[david.senchina@drake.edu](mailto:david.senchina@drake.edu)

#### LINGUISTICS

Dr. Taylor Miller,  
[taylor.miller@oswego.edu](mailto:taylor.miller@oswego.edu)

#### LITERARY STUDIES

Dr. Melissa Ames,  
[mames@ein.edu](mailto:mames@ein.edu)

Dr. Douglas Guerra,  
[douglas.guerra@oswego.edu](mailto:douglas.guerra@oswego.edu)

#### MATHEMATICS

Dr. Dele Oluwade,  
[deleoluwade@yahoo.com](mailto:deleoluwade@yahoo.com)

Dr. Christopher Baltus,  
[christopher.baltus@oswego.edu](mailto:christopher.baltus@oswego.edu)

Dr. Mark Baker,  
[mark.baker@oswego.edu](mailto:mark.baker@oswego.edu)

Dr. Monday Nnakwe,  
[mondaynnakwe@gmail.com](mailto:mondaynnakwe@gmail.com) (Anburn University)

Dr. Roza Aceska,  
[raceska@bsu.edu](mailto:raceska@bsu.edu)

Dr. Charuka Wickramasinghe,  
[gi6036@wayne.edu](mailto:gi6036@wayne.edu)

#### MEDICAL SCIENCES

Dr. Thomas Mahl,  
[Thomas.Mahl@va.gov](mailto:Thomas.Mahl@va.gov)

Dr. Jessica Amber Jennings,  
[jjennings@memphis.edu](mailto:jjennings@memphis.edu)

#### METEOROLOGY

Dr. Steven Skubis,  
[stevens.skubis@oswego.edu](mailto:stevens.skubis@oswego.edu)

#### NANOSCIENCE AND CHEMISTRY

Dr. Gary Baker,  
[bakergar@missouri.edu](mailto:bakergar@missouri.edu)

#### PHYSICS

Dr. Priyanka Rupasinghe,  
[priyanka.rupasinghe@oswego.edu](mailto:priyanka.rupasinghe@oswego.edu)

#### POLITICAL SCIENCE

Dr. Kaden Paulson-Smith,  
[Paulsonk@uwgb.edu](mailto:Paulsonk@uwgb.edu)

#### PSYCHOLOGY

Dr. Matthew Dykas,  
[matt.dykas@oswego.edu](mailto:matt.dykas@oswego.edu)

Dr. Damian Kely-Stephen,  
[kelystd@newpaltz.edu](mailto:kelystd@newpaltz.edu)

Dr. Kenneth Barideaux Jr.,  
[kbaridea@uscupstate.edu](mailto:kbaridea@uscupstate.edu)

#### SOCIAL SCIENCES

Dr. Rena Zito,  
[rzito@elon.edu](mailto:rzito@elon.edu)

Dr. Dana Atwood,  
[atwoodd@uwgb.edu](mailto:atwoodd@uwgb.edu)

#### STATISTICS

Dr. Mark Baker,  
[mark.baker@oswego.edu](mailto:mark.baker@oswego.edu)

#### TECHNOLOGY, ENGINEERING

Dr. Reg Pecen,  
[regpecen@sbsu.edu](mailto:regpecen@sbsu.edu)

#### ZOOLOGY

Dr. Chloe Lash,  
[CLash@stfrancis.edu](mailto:CLash@stfrancis.edu)

# Exploring the Link Between Mental Health Stigma and Help-Seeking Behavior in Law Enforcement Officers

Taylor Radtke\*<sup>a</sup> & Cindy Prins<sup>b</sup>

<sup>a</sup>Department of Biology, University of Central Florida College of Sciences, Orlando, FL

<sup>b</sup>Department of Population Health Sciences, University of Central Florida College of Medicine, Orlando, FL

<https://doi.org/10.33697/ajur.2026.172>

Student: [tradtke@yahoo.com](mailto:tradtke@yahoo.com)\*, [taylor.radtke@ucf.edu](mailto:taylor.radtke@ucf.edu)

Mentor: [cindy.prins@ucf.edu](mailto:cindy.prins@ucf.edu)

## ABSTRACT

Law enforcement officers work in high-stress situations with frequent exposure to traumatic events. This increases their likelihood of experiencing an impact to their mental health, yet many officers hesitate to seek professional support. Stigma persists as an important factor affecting help-seeking behaviors in many individuals. The present study surveyed 46 law enforcement officers using an anonymous Qualtrics questionnaire to assess how perceived stigma relates to willingness to seek support. Most (87.0%) reported mental health challenges as common in their field, yet few felt comfortable discussing such issues with their supervisors or therapists affiliated with their work. Conversely, 45.7% were comfortable discussing these topics with coworkers, and 52.1% with a therapist outside of their work. Additional barriers cited by officers included concerns of job security (47.8%), confidentiality (28.3%), and judgment from peers (30.4%). 52.2% of officers reported personally witnessing or experiencing stigma at their workplace, and 54.3% believed reaching out for help leads to receiving negative perceptions from peers. Spearman's rank correlation analyses found that officers who experienced stigma were less comfortable seeking help from internal resources. Additionally, qualitative data highlighted the need for confidential support resources, organizational support, and a culture change in their organization. These findings align with current research that stigma harms help-seeking behaviors and highlights the need for targeted interventions, such as anti-stigma training and peer programs.

## KEYWORDS

Mental health; Stigma; Law enforcement; Help-seeking behavior; Stigma reduction, Mental illness, Perceived stigma; Workplace culture

## INTRODUCTION

Many people around the world suffer from mental health challenges, whether or not they have an official diagnosis. One in two individuals suffers from diagnosed mental health disorders, while only one in four receives a form of treatment.<sup>1,2</sup> This underutilization of support services is a major concern among public health professionals.<sup>3-5</sup> Many studies focus on the help-seeking behaviors of different cultural or social groups, while only a few studies have focused on specific careers.<sup>6-9</sup> Some social groups studied include groups residing in low-resource settings, female veterans, and diverse ethnic groups. Several recent studies are shifting their focus to what specific factors play into the underuse of these services and avenues of support.<sup>10-13</sup> Despite this, the exploration of these factors in certain careers remains underrepresented.

Law enforcement officers work in a profession of constant exposure to threats, trauma, and high-stakes decision-making.<sup>14</sup> Many report that they face a heightened workload in addition to a lack of proper rest.<sup>15</sup> These stressors place officers at an increased risk of developing mental health disorders such as depression, post-traumatic stress disorder (PTSD), and anxiety.<sup>17,18</sup> A study surveying law enforcement in Texas found that 12% of police officers had a diagnosed mental health disorder, but 26% were currently experiencing mental illness symptoms.<sup>19</sup> Untreated mental health issues among police may contribute to increased substance use, higher rates of sick leave, and potentially higher suicide rates.<sup>17,20,21</sup> Despite these risks, utilization of mental health services among officers remains low.

One of the most persistent explanations is mental health stigma. Stigma operates at many different levels: public, self, and perceived.<sup>22,23</sup> Self-stigma specifically is associated with decreased self-esteem, willingness to seek help, and increased social withdrawal.<sup>24</sup> In policing culture, stigma may be amplified by norms of toughness and fear of being seen as unfit for duty.<sup>25,26</sup> Stigma is negatively associated with law enforcement help-seeking attitudes in several studies.<sup>16,27,28</sup> A meta-analysis in 2015 found that stigma exerts a small-to-moderate negative effect on help-seeking behavior in general populations.<sup>27</sup> However, a study that

surveyed 248 police officers in Texas and Oklahoma on their attitudes toward seeking mental health services and mental health stigma found that stigma significantly negatively affected officers' help-seeking behavior. This study also made a distinction between public stigma and self-stigma, and it was found that both were negatively associated with attitudes toward seeking mental health services.<sup>25</sup> Additionally, a study surveying multiple Wisconsin police departments found that perceived public stigma significantly reduced help-seeking intentions amongst officers and was fully mediated by self-stigma. They also suggested that low help-seeking intent among police officers could result in a buildup of stress and anxiety, which can cause many issues if not treated.<sup>29</sup>

There are several barriers outside of stigma that are often recognized among law enforcement officers. These include concerns of job security, confidentiality, and a lack of resources.<sup>14,30</sup> One study conducted a systematic review of 14 articles analyzing first responders' experience with stigma. They found that the most frequently endorsed barriers were fears of confidentiality and fears that seeking help would have negative consequences on their profession. These barriers, along with stigma, greatly affect a significant number of first responders, leading to decreased support for mental health.<sup>31</sup> Understanding how not only stigma, but also other factors affect individuals' willingness to seek help is important for creating targeted interventions to improve support services.<sup>32</sup>

Stigma likely plays a significant role in the reluctance of officers to seek out mental health support.<sup>33</sup> However, research is relatively scarce, with many studies relying on quantitative data rather than open-ended questions. Small-sample, mixed-method design approaches may supplement the general picture by providing a deeper understanding of officers' perception of stigma. This study aims to examine how perceived mental health stigma affects help-seeking behaviors among law enforcement officers. Specifically, this study explores the relationship between stigma exposure and officers' comfort seeking support from internal and external sources, as well as the organizational and structural barriers that shape these behaviors. It is hypothesized that officers who report experiencing or witnessing mental health stigma will report lower comfort seeking help from internal organizational resources.

## METHODS AND PROCEDURES

### *Study design and participants*

This cross-sectional study examined how the perception of mental health stigma influences officers' willingness to seek professional support services by utilizing an online, anonymous survey. The survey was distributed electronically to officers within a Wisconsin deputy sheriffs' association in the United States. The survey was created and administered using Qualtrics between June 20 and July 4, 2025 (Qualtrics, Provo, UT). Participation was entirely voluntary. The survey opened with an informed consent form describing the study's purpose, procedures, and participants' rights. By continuing to take the survey, participants agreed to participate in the study. To ensure anonymity, no identifying information, including name, rank, department, or exact age, was collected. To ensure confidentiality, the deputy sheriffs' association only had access to a QR code and link prompting them to fill out the survey. The responses were accessed only by the researchers through password-protected Qualtrics accounts. This study was reviewed and approved as exempt by the University of Central Florida Institutional Review Board. Participants were eligible if they were at least 18 years old and currently employed in the law enforcement profession. Recruitment was facilitated by the president of the sheriffs' association, who distributed the survey link to members via email.

### *Measures*

The survey consisted of 14 questions, roughly 5 minutes, designed to assess perceptions of mental health stigma, help-seeking behaviors, and the workplace culture within the organization (see Appendix). The survey instrument was developed independently by the researchers and did not employ standardized or previously validated stigma scales. A new instrument was created to capture law enforcement-specific experiences and perceptions of mental health stigma, including organizational, cultural, and operational factors that may not be addressed in general stigma scales. The questions were tailored to capture relevant information specific to the law enforcement culture. The survey collected both quantitative and qualitative data. Open-ended questions at the end of the survey asked officers to suggest what they think could be done to reduce stigma in the workplace, as well as to provide any additional comments or experiences if they wished.

Participants were asked to provide basic demographic data: their current profession, their age reported in ranges of 10 years, and how many years of service they have within the profession, reported as an open-ended numeric response. Race and gender were not collected as the aim of this study was to explore mental health stigma and help-seeking behaviors in the law enforcement profession, rather than to investigate potential differences associated with individual demographic characteristics. Officers were then asked to rate the perceived prevalence of mental health challenges in their profession and their level of comfort discussing mental health with different individuals. Mental health challenges were defined as "any difficulty or disruption in one's emotional, psychological, or social well-being that impacts how one thinks, feels, or behaves. This can include a range from diagnosed disorders to temporary anxiety or stress." Comfort level was assessed for coworkers, supervisors, and therapists affiliated with work and outside of work. Each item was rated on a five-point Likert scale ranging from "extremely uncomfortable" to "extremely comfortable." Participants were also asked about coworkers' reporting of mental health, as well as what barriers they

face to seeking help. The survey also asked respondents if they had personally witnessed or experienced stigma in their workplace. In the survey, stigma was defined by the researchers as “the negative attitude about a person that leads to negative actions/thoughts or discrimination.”

*Data analysis*

Quantitative data were analyzed using JASP 0.95.4. Descriptive statistics (frequencies, percentages) were calculated for categorical items. Associations between stigma exposure and comfort with various individuals were tested with Spearman’s rank correlation and chi-square analysis, where binary categories were applied. Significance was set at  $p < 0.05$ . Spearman’s rho was selected instead of Pearson’s r because several key variables in this study were ordinal or not normally distributed. This method does not assume normality or linearity, making it appropriate for assessing associations between stigma and comfort discussing mental health.

Qualitative responses to the two open-ended questions at the end of the survey were reviewed and analyzed using thematic analysis, a method suitable for identifying, organizing, and interpreting patterns within textual data. Two different rounds of inductive coding were employed on Microsoft Excel to identify recurring ideas or concepts and ensure consistency. Codes were then grouped into broader themes based on conceptual similarity, resulting in categories such as barriers, workplace culture, and recommendations. Coding was first done by one researcher, then a second researcher reviewed the coding structure and thematic assignments to ensure consistency and clarity. After the codes were independently reviewed by the researchers, they collectively determined the final themes that emerged for each response. Representative quotes were selected to illustrate each theme. This approach was chosen as it allows for systematic exploration of participants’ perspectives and understanding of attitudes and experiences related to mental health stigma.

**RESULTS**

*Descriptive Statistics*

A total of 46 law enforcement officers completed the survey. Demographic information collected in this survey included years of service and age. Age and years of service were collected, as previous research suggests that both factors can influence attitudes toward mental health and willingness to seek help. Including these categories allows for examination of whether comfort discussing mental health and experiences of stigma vary across different career stages.

Years of service ranged from two to 33, with a mean of 14.4 years and a standard deviation of 10.7 years. The ages of 45 participants were recorded, as one officer opted not to answer this question. **Table 1** presents the age ranges of the 45 participants, collected in ranges of 10 years. The mean age was 38.7 years with a standard deviation of 11.2 years. Demographic variables were treated as exploratory covariates rather than control variables. They were not used to adjust or control the primary analyses, but were examined to explore if patterns in mental health stigma differed across demographic subgroups. No significant differences were found for any demographic variables, indicating that stigma and related outcomes were generally consistent across age, years of service, and professional role in this study.

Age Range	Number of Participants
18-30 years	11
31-40 years	16
41-50 years	9
51-60 years	9
61-70 years	0
71+	0

**Table 1.** Age of participants reported in ranges of 10 years (n=45).

When asked how common mental health challenges are in their profession, the majority of respondents reported very common (52.2%) or common (34.8%), with the remainder outlining them as only somewhat common. No respondents reported rare. Comfort level when discussing mental health varied by who they discussed the topic with. **Table 2** presents the full breakdown of responses. This data was organized into categories to analyze relative willingness to discuss mental health. Roughly 45.7% of officers expressed some degree of comfort with speaking to coworkers, whereas only 23.9% felt the same towards their supervisor. Additionally, only 17.4% felt comfortable speaking with a therapist affiliated with their work. In contrast, 52.1% were comfortable discussing mental health with a therapist outside of their work.

	Extremely uncomfortable	Somewhat uncomfortable	Neutral	Somewhat comfortable	Extremely comfortable
Coworker	10.9%	21.7%	21.7%	28.3%	17.4%
Supervisor	32.6%	26.1%	17.4%	13.0%	10.9%
Internal therapist	32.6%	26.1%	23.9%	6.5%	10.9%
External therapist	6.5%	19.6%	21.7%	30.4%	21.7%

**Table 2.** Reported comfort levels discussing mental health issues with different groups (n=46).

Supervisors were included as a distinct workplace-based source of support rather than a clinical support provider. While supervisors do not provide therapeutic care, they often serve as gatekeepers to formal resources, accommodations, time off, and referrals, making them a relevant point of contact in discussions of help-seeking. Analyses examining supervisors should therefore be interpreted as reflecting organizational and relational support, not clinical mental health care.

*Inferential Analyses*

Correlational analyses revealed that the perception of stigma is significantly associated with decreased comfort when seeking help internally. Participants were asked whether they had personally experienced or witnessed mental health stigma in their workplace with a dichotomous (yes/no) question. Exposure to stigma was operationalized as 0 = no exposure and 1 = exposure. Comfort discussing mental health was measured on a Likert scale in which higher scores indicated lower comfort levels. Spearman correlational analyses indicated that stigma exposure was associated with reduced comfort discussing with coworkers ( $\rho=0.32, p=0.04$ ) and with internal therapists ( $\rho=0.38, p=0.01$ ). No significant correlations were found for comfort discussing mental health with supervisors ( $\rho=0.26, p=0.11$ ) or with external therapists ( $\rho=0.06, p=0.71$ ). The majority of officers (65.2%) reported they had been approached by a coworker with concerns about their mental health. But only 30.4% reported elevating these concerns to their supervisor.

Stigma was a widely reported barrier amongst law enforcement. More than half (52.2%) reported personally witnessing or experiencing mental health stigma. Similarly, 54.3% revealed that officers who discuss their mental health challenges are treated differently. A chi-square test of associations between exposure to stigma and perception of being treated differently for seeking help approached but did not reach significance ( $X^2=2.56, p=0.11$ ). Likewise, no significant association was found between exposure to stigma and access to mental health leave ( $X^2=0.96, p=0.33$ ).

Respondents were provided a multiple-choice list of potential barriers with the ability to select all that applied. Barriers are reported descriptively to highlight the most common concerns among participants. Due to the small sample size, analyses examining differences in barriers by stigma exposure, age, or years of service were not feasible. The most frequently reported barriers were organizational barriers, such as concerns affecting job security or promotions. Although organizational barriers were the most commonly reported, logistical barriers were also reported several times, including time constraints and financial cost. **Table 3** reflects the distribution of the most commonly reported barriers and their percentages. Three participants provided barriers using the “other” option, citing fear of their struggles being used against them, and perceiving mental health needs as personally manageable.

Barrier	Percent Endorsed
Fear of job security or promotions	47.8%
Lack of time	30.4%
Judgement from peers	30.4%
Confidentiality concerns	28.3%
Financial cost	13.0%

**Table 3.** Reported barriers to receiving mental health treatment (n=46).

*Qualitative results*

Open-ended questions were analyzed with two rounds of inductive coding then subsequently were reviewed by a second researcher. The final themes of each response were collectively decided by both researchers. These questions asked respondents what they believed could be done to reduce stigma in their field (32 responses), as well as whether they had any additional thoughts or experiences they wanted to share (20 responses). These responses were analyzed for any emerging themes to deepen the understanding beyond quantitative data. Nineteen themes emerged from these responses, with multiple themes sometimes appearing in a single response. The number of times each theme emerged is outlined in **Table 4** below. These themes were grouped into four higher-order categories: cultural and stigma-related factors, organizational and leadership influences, resource and support needs, and the need for awareness and education.

Of the themes that emerged from the qualitative responses, cultural and stigma-related factors were a primary topic. Many responses referenced workplace culture and the normalization of mental health challenges. Participants described the need to “make it okay to have feelings,” to “talk about it more,” and to create an “all-encompassing culture of real concern for our wellness.” Some noted that officers are trained to control their emotions “to the point where... It’s natural for them to suppress them even when it’s unhealthy.” Others described negative workplace environments, stating that a “toxic environment” forced them to ignore their mental health. Several responses reflected distrust in the organization, including reports of supervisors attempting to discipline officers for mental health concerns, or comments that support was promised but they do not follow through. A small number of responses suggested no change was needed or that change was unlikely, including statements such as “law enforcement will never change significantly” and “I don’t think that’s possible” when asked what they think can reduce stigma in their field.

Theme	Number of Times Emerged
Normalization of mental health challenges	12
Job stress	10
Proactive support	8
Need for time off	8
Organizational culture	7
Providing resources for health and wellness	6
Education of mental health challenges	6
Leadership engagement	6
Concerns of job security	6
Confidentiality	4
Need for reduction of stigma	4
Support for mental health treatment	3
Culture is beyond fixing	3
Uncertainty about solutions	2
Leadership accountability	2
Distrust in organization	1
Need for mental health policies	1
No change needed	1
Importance of mental health services	1

Table 4. Themes emerged from qualitative responses (n=52).

Another main topic discussed was organizational and leadership influences. Several responses called for supervisors to be “more involved,” “more in touch with their employees,” or to have “more face-to-face contact.” Others described concerns about job security, including statements that officers have been fired “depending on mental status” or that people fear “losing their badge for getting help.” Some responses emphasized the need for leadership accountability and policy changes, such as enforceable penalties for managers who violate best practices or policies guaranteeing mental health time off and protection from employment consequences.

Respondents also highlighted the need for resources and support. Participants suggested measures such as “mandatory yearly therapist visits,” “have a psychologist on staff,” and providing “mental health time off” or insurance-covered sessions. Time off and workload reduction were also frequently mentioned, including requests for “more time off...to be with family and friends to decompress,” “less forced overtime shifts,” and reduced workload expectations. Several responses also described the profession as “extremely taxing on your mental health” due to long hours and stress. Other responses emphasized proactive support, such as encouraging officers to talk to someone before problems build up or having supervisors identify employees who appear stressed.

Finally, officers emphasized the need for awareness and education. Participants suggested “recurring training,” “continuous talk and sharing of relatable resources,” and “being open about it.” Some responses described the stress of the profession, noting that the job “has such a drain on your mental health” or describing daily anxiety as “a meat grinder.” Confidentiality concerns were also mentioned, including fears that internal programs were “too closely affiliated with work” or that wellness programs might notify employers. A few responses reflected uncertainty about solutions, with participants stating, “not sure” or “if I had the answer...I would’ve offered it already.”

Collectively, the qualitative responses provided contextual detail that expanded upon patterns observed in the quantitative results. Participant narratives provided detailed descriptions of respondents’ experiences related to stigma, confidentiality, and workplace context. Across themes, respondents described structural and cultural conditions shaping their willingness to discuss mental health concerns and engage with available resources. These qualitative findings, therefore, serve to contextualize and enrich the quantitative outcomes by highlighting the lived experiences underlying reported attitudes and behaviors.

**DISCUSSION**

Altogether, the findings suggest that stigma functions as a central mechanism shaping help-seeking behaviors among officers. Internalized and perceived stigma appear to influence not only attitudes toward mental health care but also trust in organizational resources, which in turn affects willingness to seek support. These dynamics are further compounded by organizational culture and structural barriers, such as time constraints and financial cost. Framing the results within these broader mechanisms helps situate the findings within existing literature while clarifying their implications for intervention.

This study offers an insight into how law enforcement officers within a single Wisconsin agency perceived and experienced mental health stigma by integrating quantitative survey data with qualitative narratives. Rather than offering conclusions applicable to law enforcement broadly, these findings reflect the experiences and perceptions of officers within this specific organizational context. Within this sample, a core finding is that stigma remains a restraint to seeking help, particularly in relation to organizational culture.

Internal sources of support were less trusted than external ones. 52.1% of respondents were comfortable with external therapists; fewer expressed comfort with supervisors (23.9%) or internal therapists (17.4%). These results suggest that, among participants in this study, internal organizational roles were perceived as less safe for disclosure. This is consistent with broader findings in first responder and policing literature.<sup>29</sup> Many officers reported concerns about job security (47.8%) and confidentiality (28.3%), indicating that organizational and cultural factors within this setting may influence help-seeking behavior. One past study in particular also cited confidentiality as one of the most commonly reported barriers.<sup>31</sup> Interestingly, this study found logistical barriers such as time constraints and finances affecting help-seeking behavior among officers.<sup>34</sup> Officers in the present study reported that lack of time (30.4%) and financial cost (13.0%) were barriers. This data shows participants emphasized that both cultural and structural factors affect help-seeking behaviors.

Qualitative responses further contextualized these patterns by illustrating how participants described confidentiality concerns and distrust of institutional supports in their own words. Respondents frequently referenced the perceived independence of external services as a reason they might be more willing to seek help outside their organization. Similar themes have been described in prior qualitative studies of first responders, which stressed the role of institutionalized stigma and confidentiality concerns.<sup>35</sup> The present findings extend this work as exposure to stigma was found to be correlated with reduced comfort in seeking help from both coworkers ( $p=0.044$ ) and internal therapists ( $p=0.014$ ), suggesting officers in this organization experience cultural stigma in their workplace often. Importantly, these qualitative accounts should be understood as illustrative rather than exhaustive, reflecting the perspectives of respondents who chose to participate.

Participants also described a range of organizational strategies they believed could reduce stigma, including wellness programs, routine check-ins, and leadership visibility. These suggestions are consistent with interventions discussed in prior literature, such as anti-stigma training, but in this study, they represent participant perceptions rather than evaluated outcomes.<sup>11,29,36</sup> Notably, recommendations such as mandatory wellness checkups were often described conditionally, with respondents emphasizing the importance of confidentiality and independence from disciplinary oversight, such as an external staff psychologist and a confidential reporting line. Qualitative responses indicate that officers may be more likely to seek help when secure and reliable resources are available. Findings also showed that stigma is highly internalized and shapes the perceptions of disclosure amongst officers. Many officers emphasized the prevalence of mental health challenges (52.2% very common, 34.8% common), yet reported the negative perception of individuals who do seek help (54.3%). As such, these findings do not imply that existing internal services are effective but rather underline the types of structural features participants believed would be necessary for such supports to be acceptable within their organizational context, highlighting a potential area for further study. Peer support networks led by other officers were another suggestion by participants as a mechanism to potentially reduce stigma by offering a trusted point to receive support. Several peer programs have been implemented in other first responder organizations, offering insight into how other organizations are approaching stigma in the workplace.<sup>36-38</sup> While the present study cannot determine the effectiveness of these strategies, the data indicate that officers in this Wisconsin agency perceive organizational and peer-based resources as potentially important for mental health support.

Although participants reported substantial distrust toward internal mental health resources, suggestions for mandatory wellness checkups were not framed as an endorsement of existing internal services. Rather, respondents described mandatory check-ins as a potential mechanism to normalize mental health care and remove the burden of self-disclosure in a stigmatized environment. Several participants emphasized that such measures would only be acceptable if confidentiality were guaranteed and services were independent from disciplinary or administrative oversight. These concerns highlight the broader need for structural supports that promote confidential and stigma-free access to mental health resources within law enforcement agencies. Federally, programs such as the Law Enforcement Mental Health and Wellness Act (LEMHWA) offer funding to build peer support, training, and reduce stigma for organizations.<sup>39</sup> These programs provide organizations with resources they may utilize to develop confidential wellness support services, though the present study cannot determine their effectiveness.<sup>40</sup>

This study offers several contributions to the literature on mental health stigma in law enforcement. By integrating quantitative measures of stigma and qualitative narratives, this study provides a more nuanced understanding of how stigma manifests in officers' daily experiences than quantitative-only studies. The qualitative findings add context to observed patterns in help-seeking comfort and perceived barriers, clarifying how organizational culture and confidentiality concerns shape these outcomes. This study uniquely highlights the disparity between trust in internal versus external mental health supports, suggesting that organizational culture may play a key role in shaping help-seeking behaviors. Additionally, capturing both cultural and structural barriers offers a comprehensive view of factors limiting help-seeking among officers. Findings also suggest that stigma may be socially transmitted within the workplace, influencing perceptions of disclosure and mental health challenges. By identifying the types of support officers perceive as trustworthy, this study informs the design of future interventions that align with officers' preferences, potentially enhancing uptake and effectiveness. Finally, these findings contribute law enforcement-specific data to the broader first responder literature, which has largely focused on other occupational groups, while reinforcing and expanding upon prior research on organizational and cultural determinants of stigma.

Despite its contributions, this study has limitations. The small sample size ( $n=46$ ) restricts generalizability to all officers within both the organization surveyed and the profession of law enforcement itself. The small sample also restricts statistical power, making subgroup analyses of age or years of service not feasible. A formal power analysis was not conducted, which may have constrained the study's ability to detect smaller effects. Additionally, degrees of freedom and confidence intervals for correlations were not calculated, which limits the precision of reported associations. The lack of survey reminders due to the short period of data collection may have contributed to the lack of responses. Additionally, no response rate was calculated as the total number of officers invited to participate in this study was not provided to the researchers. The self-selected nature of participation also introduces the possibility of bias, as officers interested in mental health services and making support proactive may have been overrepresented. By relying on self-report and recall, responses to questions are subject to the ability to accurately recall past experiences and potentially even concerns of confidentiality, even though anonymity was assured. As the survey itself employed single-item measures and was developed independently, it was not pilot tested or tested for internal consistency. This may have caused individuals to respond in ways that were not anticipated. Additionally, without pre-validated stigma scales, such as POSS, our binary measurement of stigma may not capture self-stigma or perceived public stigma accurately.<sup>41</sup> The qualitative component of this study did not directly address the potential link between mental health stigma and help-seeking behavior, limiting the ability to fully illuminate the psychological mechanisms underlying this relationship. This study did not collect race or gender information, nor did it examine the role of culturally related factors on mental health stigma or help-seeking behaviors, which constrains the interpretability and generalizability of the current findings. Lastly, this study design is cross-sectional and correlational; we cannot infer causality or directionality, limiting conclusions of real-world behavior.

Future research should acknowledge these limitations. Larger, wider-ranging, and multi-agency samples may improve external validity and applications. Incorporating validated stigma scales could increase the insight gained from responses. Longitudinal designs could examine how stigma evolves over the course of officers' careers and whether organizational supports influence perceived stigma. These experimental studies could directly test how help-seeking behavior among law enforcement changes with interventions.

## CONCLUSIONS

This study emphasizes the critical need for law enforcement organizations to address the deep-rooted stigma surrounding mental health in the workplace. Participants suggested several mechanisms to potentially reduce this stigma and increase the normalcy of mental health issues within this organization. Future research should investigate how specific factors, such as organizational policies and leadership attitudes, influence help-seeking behaviors across departments of varying sizes and regions. Longitudinal and intervention-based studies could also evaluate if implementing external counseling, routine wellness checkups, or stigma reduction training leads to measurable changes in the well-being of officers.

Broadly, this work contributes to the growing recognition that mental health stigma within law enforcement is not only an individual issue but a cultural one. By organizing these findings with the current body of literature on occupational mental health, future studies can build a foundation for national standards. Consistent and validated mental wellness programs could ensure the normalization of mental health care as an integral component of professional integrity in policing.

## ACKNOWLEDGMENTS

The authors thank the surveyed deputy sheriffs' association for their participation in this study, as well as the University of Central Florida College of Medicine for funding to T. Radtke.

## REFERENCES

1. Terlizzi, E., and Schiller, J. (2022) Mental Health Treatment Among Adults Aged 18–44: United States, 2019–2021, *National Center for Health Statistics (U.S.)*. doi:10.15620/cdc:120293
2. McGrath, J.J., Al-Hamzawi, A., Alonso, J., Altwajri, Y., Andrade, L., Bromet, E., Bruffaerts, R., Miguel Caldas de Almeida, J., Chardou, S., Tat Chiu, W., Degenhardt, L., Demler, O., Ferry, F., Gureje, O., Maria Haro, J., Karam, E., Karam, G., Khaled, S., Kovess-Masfety, V., Magno, M., Elena Medina-Mora, M., Moskalewicz, J., Navarro-Mateu, F., Nishi, D., Plana-Ripoll, O., Posada-Villa, J., Rapsey, C., Sampson, N., Carlos Stagnaro, J., Stein, D., Ten Have, M., Torres, Y., Vladescu, C., Woodruff, P., Zarkov, Z., and Kessler, R. (2023) Age-of-onset and cumulative risk of mental disorders: A cross-national analysis of population surveys data based on 156,331 respondents from 29 countries, *Lancet Psychiatry* 10(9), 668–681. doi:10.1016/S2215-0366(23)00193-1
3. Yoon, H. (2019) Willingness to seek professional mental health service in three ethnic groups of Asian Americans, *J Ethn Cult Divers Soc Work* 30(5), 445–461. doi:10.1080/15313204.2019.1680589
4. Abe-Kim, J., Takeuchi, D.T., Hong, S., Zane, N., Sue, S., Spencer, M.S., Appel, H., Nicdao, E., and Alegria, M. (2007) Use of Mental Health-Related Services Among Immigrant and US-Born Asian Americans: Results From the National Latino and Asian American Study, *Am. J. Public Health* 97(1), 91–98. doi:10.2105/AJPH.2006.098541

5. Sorkin, D.H., Nguyen, H., and Ngo-Metzger, Q. (2011) Assessing the Mental Health Needs and Barriers to Care Among a Diverse Sample of Asian American Older Adults, *J Gen Intern Med* 26(6), 595–602. doi:10.1007/s11606-010-1612-6
6. Mekonen, T., Chan, G.C.K., Belete, T., Menberu, M., Davidson, L., Hides, L., and Leung, J. (2022) Mental health service utilization in a lowresource setting: A qualitative study on perspectives of health professionals in Northwest Ethiopia, *PLoS ONE* 17(11). doi:10.1371/journal.pone.0278106
7. Bañada, R., Jang, Y., Palinkas, L.A. (2024) Willingness to use mental health counseling in diverse groups of Asian Americans, *PLoS ONE* 19(7). doi:10.1371/journal.pone.0306064
8. Kimerling, R., Bastian, L.A., Bean-Mayberry, B.A., Bucossi, M.M., Carney, D.V., Goldstein, K.M., Phibbs, C.S., Pomernacki, A., Sadler, A.G., Yano, E.M., and Frayne, S.M. (2015) Patient-Centered Mental Health Care for Female Veterans, *Psychiatr Serv* 66(2), 155–162. doi:10.1176/appi.ps.201300551
9. Luo, B.A., Li, S., Chen, S., Qin, L., Chen, Y., Shu, M., and Liao, X. (2022) Mental health resources and its equity in Central South of China: A case study of Hunan Province, *PLoS ONE* 17(10). doi:10.1371/journal.pone.0272073
10. Jetelina, K.K., Molsberry, R.J., Gonzalez, J.R., Beauchamp, A.M., and Hall, T. (2020) Prevalence of Mental Illness and Mental Health Care Use Among Police Officers, *JAMA Netw Open* 3(10). doi:10.1001/jamanetworkopen.2020.19658
11. van den Brink, R.H., Broer, J., Tholen, A.J., Winthorst, W.H., Visser, E., and Wiersma, D. (2012) Role of the police in linking individuals experiencing mental health crises with mental health services, *BMC Psychiatry* 12. doi:10.1186/1471-244X-12-171
12. Schauman, O., Aschan, L.E., Arias, N., Beards, S., and Clement, S. (2013) Interventions to increase initial appointment attendance in mental health services: a systematic review, *Psychiatr Serv* 64(12), 1249–1258. doi:10.1176/appi.ps.201200540
13. Patel, R., Harvey, J., and Forrester, A. (2018) Systemic limitations in the delivery of mental health care in prisons in England, *Int. J. Law Psychiatry* 60, 17–25. doi:10.1016/j.ijlp.2018.06.003
14. Richards, N.K., Suarez, E.B., and Arocha, J.F. (2021) Law Enforcement Officers' Barriers to Seeking Mental Health Services: a Scoping Review, *J Police Crim Psychol* 36(3), 351–359. doi:10.1007/s11896-021-09454-x
15. Okhrimenko, I.M., Barko, V.V., Vavryk, L.V., Chornous, V.D., Okhrimenko, S.S., Aleksandrov, Y.V., and Onishchuk, L.M. (2023) THE IMPACT OF PROFESSIONAL STRESS ON THE MENTAL HEALTH OF LAW ENFORCEMENT OFFICERS, *Wiad Lek* 76(6), 1428–1435. doi:10.36740/WLek202306115
16. Papazoglou, K., and Pizarro Andersen, J. (2014) A Guide to Utilizing Police Training As a Tool to Promote Resilience and Improve Health Outcomes Among Police Officers, *ResearchGate* 20(2), 103–111. doi:10.1037/b0099394
17. Lane, J., Le, M., Martin, K., Bickle, K., Campbell, E., and Ricciardelli, R. (2021) Police Attitudes Toward Seeking Professional Mental Health Treatment, *J Police Crim Psychol* 37, 123–131. doi:10.1007/s11896-021-09467-
18. Rodriguez, S., Ferrell, B., Cipriano, R.J., and Van Hasselt, V.B. (2024) Law enforcement mental health: Strategies and issues in prevention and treatment, *Pract Innov* 9(1), 52–64. doi:10.1037/pri0000210
19. Jetelina, K., Molsberry, R., Gonzalez, J., Beauchamp, A., and Hall, T. (2020) Prevalence of Mental Illness and Mental Health Care Use Among Police Officers, *JAMA Netw Open* 3(10). doi:10.1001/jamanetworkopen.2020.19658
20. Violanti, J.M., Charles, L.E., McCanlies, E., Hartley, T.A., Baughman, P., Andrew, M.E., Fekedulegn, D., Ma, C.C., Mnatsakanova, A., and Burchfiel, C.M. (2017) Police stressors and health: a state-of-the-art review, *Policing: A Journal of Policy and Practice* 40(4), 642–656. doi:10.1108/PJPSM-06-2016-0097
21. Boyd, J., Boyd, S., and Kerr, T. (2015) Visual and Narrative Representations of Mental Health and Addiction by Law Enforcement, *Int. J. Drug Policy* 26(7), 636–644. doi:10.1016/j.drugpo.2015.04.007
22. Corrigan, P.W., and Rao, D. (2012) On the Self-Stigma of Mental Illness: Stages, Disclosure, and Strategies for Change, *Can J Psychiatry* 57(8), 464–469. doi:10.1177/070674371205700804
23. Corrigan, P. (2004) How stigma interferes with mental health care, *Am Psychol* 59(7), 614–625. doi:10.1037/0003-066X.59.7.614
24. Sun, J., Yin, X., Li, C., Liu, W., and Sun, H. (2022) Stigma and Peer-Led Interventions: A Systematic Review and Meta-Analysis, *Front Public Health* 13. doi:10.3389/fpubh.2022.915617
25. Karaffa, K.M., and Koch, J.M. (2016) Stigma, Pluralistic Ignorance, and Attitudes Toward Seeking Mental Health Services Among Police Officers, *Crim Justice Behav* 43(6), 759–777. doi:10.1177/0093854815613103
26. Santre, S. (2024) Mental Disorders and Mental Health Promotion in Police Officers, *Health Psychol Res* 12. doi:10.52965/001c.93904
27. Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rusch, N., Brown, J.S.L., and Thornicroft, G. (2015) What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies, *Psychol Med* 45(1), 11–27. doi:10.1017/S0033291714000129
28. Vogel, D.L., Wade, N.G., and Haake, S. (2006) Measuring the self-stigma associated with seeking psychological help, *J Couns Psychol* 53(3), 325–337. doi:10.1037/0022-0167.53.3.325
29. Grupe, D.W. (2023) Mental Health Stigma and Help-Seeking Intentions in Police Employees, *J Community Saf Well-Being* 8(1), 32–39. doi:10.35502/jcswb.290
30. Newell, C.J., Ricciardelli, R., Czarnuch, S.M., and Martin, K. (2022) Police staff and mental health: barriers and recommendations for improving help-seeking, *Police Pract Res* 23(1), 111–124. doi:10.1080/15614263.2021.1979398

31. Haugen, P.T., McCrillis, A.M., Smid, G.E., and Nijdam, M.J. (2017) Mental health stigma and barriers to mental health care for first responders: A systematic review and meta-analysis, *J Psychiatr Res* 94, 218–229. doi:10.1016/j.jpsychires.2017.08.001
32. Park, M. (2022) Mental Health Stigma and Law Enforcement Officers, *ProQuest*.
33. Whittington, M., and Basham, S. (2024) The Role of Stigma and Confidentiality in Police Officer Utilization of Stress Relief Programs, *J Police Crim Psychol* 39, 402–410. doi:10.1007/s11896-024-09681-y
34. Paoline, E.A., Wiercioch, A.G., and Gau, J.M.. (2025) Police Officer Mental Health: Predictors of and Barriers to Seeking Help, *Police Q* 0. doi:10.1177/10986111251407953
35. MacLean, S.E., Edgar, N.E., Ahluwalia, C., Testa, V., and Hatcher, S. (2025) A Qualitative Study on the Design and Implementation of a First Responder Operational Stress Injury Clinic, *Health Serv Insights* 18. doi:10.1177/11786329251333019
36. Hansson, L., and Markström, U. (2014) The effectiveness of an anti-stigma intervention in a basic police officer training programme: a controlled study, *BMC Psychiatry* 14. doi:10.1186/1471-244X-14-55
37. Wittmann, L., Dorner, R., Heuer, I., Bock, T., and Mahlke, C. (2021) Effectiveness of a contact-based anti-stigma intervention for police officers, *Int. J. Law Psychiatry* 76. doi:10.1016/j.ijlp.2021.101697
38. Horan, K.A., Marks, M., Ruiz, J., Bowers, C., and Cunningham, A. (2021) Here for My Peer: The Future of First Responder Mental Health, *Int. J. Environ Res Public Health* 18(21). doi:10.3390/ijerph182111097
39. Community Oriented Policing Services, *LAW ENFORCEMENT MENTAL HEALTH AND WELLNESS ACT (LEMHWA) PROGRAM*, <https://cops.usdoj.gov/lembwa> (accessed October 7, 2025)
40. Fisher, M.P., and Lavender, C.D. (2023) Ensuring Optimal Mental Health Programs and Policies for First Responders: Opportunities and Challenges in One U.S. State, *Community Ment Health J* 59(7), 1341–1351. doi:10.1007/s10597-023-01121-1
41. Burzee, Z., Bowers, C., and Beidel, D. (2022) A re-evaluation of Stuart’s police officer stigma scale: Measuring mental health stigma in first responders, *Front Public Health* 10. doi:10.3389/fpubh.2022.951347

#### ABOUT THE STUDENT AUTHOR

Taylor Radtke is currently attending the University of Central Florida. She plans to graduate in December 2026 with a Bachelor’s degree in Biology on the Pre-Health track. She aspires to attend medical school to become a physician.

#### PRESS SUMMARY

Law enforcement officers regularly face high-stress and traumatic situations, placing them at greater risk for mental health challenges. This study surveyed 46 officers to examine how stigma influences their willingness to seek professional support. Results showed that while most officers recognize mental health struggles as common in their field, many avoid seeking help due to fears about job security, confidentiality, and stigma. These findings highlight the urgent need for confidential resources, organizational support, and anti-stigma initiatives to promote a healthier culture within law enforcement.



# Clustering in High-Dimension – Tools and Challenges

Lucy Liu\* & Haim Bar

Department of Statistics, University of Connecticut, Storrs, CT

<https://doi.org/10.33697/ajur.2026.173>

Students: [lucy.liu@uconn.edu](mailto:lucy.liu@uconn.edu)\*

Mentor: [haim.bar@uconn.edu](mailto:haim.bar@uconn.edu)

## ABSTRACT

Dimensionality reduction methods such as Multidimensional scaling (MDS) or t-Distributed Stochastic Neighbor Embedding (t-SNE) are often followed by clustering in the reduced plot. To examine whether or to what extent these methods affect clustering, we simulate several data structures and apply clustering methods. We first perform clustering using the data in the original space, where we know the true clusters, then perform MDS and t-SNE to scale the data down to two dimensions, cluster on this projected data, and compare differences in the results. We find that MDS and t-SNE can either increase or decrease clustering performance, and are unable to correctly represent data structures with certain shape structures, or in the presence of noise. We examine several clustering methods and show that their performance depends to a large extent on the structure of the data, original dimension, and the noise level, even before we perform dimensionality reduction via MDS and t-SNE. No method among the ones considered here dominates the others in terms of clustering accuracy.

## KEYWORDS

K-means; Hierarchical clustering; Ward linkage; Complete linkage; Average linkage; Single linkage; Persistence Diagram; Simulation

## 1. INTRODUCTION

Clustering plays a vital role in extracting meaningful patterns and structures from unlabeled data. The challenge of identifying coherent groupings within data spans diverse disciplines, including biology, finance, and social sciences. For example, clustering algorithms are often used in medical imaging, where the objective is to perform segmentation and identify tumors. However, the quality of clustering outcomes can be influenced by the latent characteristics of the data, such as the shape, dimensionality, and distribution of clusters.

The central problem addressed in this paper is: given data embedded in a high-dimensional space with unknown cluster geometry, which combination of clustering algorithm and dimensionality reduction method yields the most reliable results? This question has direct practical importance because selecting an inappropriate clustering method can lead to misleading interpretations—from misidentifying cell types in genomics to incorrectly segmenting populations in social science research. To address this problem, we conduct a systematic simulation-based study that highlights the behavior of existing clustering methods under controlled, restricted conditions. We vary three key factors—cluster shape (convex vs. non-convex), ambient dimensionality, and noise level—to provide empirical guidance for practitioners who must choose among these methods without knowing the true underlying structure of their data.

When presented with a dataset to cluster, is there an optimal clustering algorithm that consistently outperforms others, or does the quality of clustering of a certain method depend on the structural layout of the data? Different algorithms, such as  $K$ -means and hierarchical clustering, each have unique assumptions and mechanisms. A comparative understanding of these methods in relation to varying data configurations is crucial for guiding methodological choices.

Do certain algorithms demonstrate robustness, or do they require significant adaptation to handle such complexities? As datasets grow increasingly complex, high-dimensional data presents a unique set of challenges for clustering algorithms. The “curse of dimensionality” can obscure meaningful groupings, which makes it necessary to investigate how classical and modern clustering techniques perform in high-dimensional spaces. Understanding algorithmic behavior in these scenarios helps practitioners anticipate potential pitfalls and select more reliable clustering strategies.

Does dimensionality reduction invariably enhance clustering by removing noise, or does it introduce distortions that compromise results? Projecting high-dimensional data to a lower dimension offers several advantages, such as enabling visualization in two dimensions and improving computational efficiency. However, there is also a risk of misrepresentation, potentially introducing distortions that affect the outcome of clustering. Exploring these trade-offs is essential for determining when and how dimensionality reduction should be applied.

Furthermore, how does the spatial arrangement of data influence the effectiveness of these combined strategies? Dimensionality reduction techniques, such as Multidimensional Scaling (MDS)<sup>1</sup> and t-Distributed Stochastic Neighbor Embedding (t-SNE)<sup>2</sup>, are often employed to mitigate high-dimensional challenges. However, their impact on clustering quality remains an open question, particularly for non-convex or complex data structures.

The process of dimensionality reduction may lead to misplacement of points. Namely, points that are close in the original data may end up far apart in the projected data, or vice versa. The impact of such misplacements on clustering quality has not been fully addressed.

This paper is organized as follows. In Section 2, we review related work and the application areas that motivate this study. In Section 3, we establish the necessary notation, justify our selection of clustering and dimensionality reduction methods, and describe each method in detail. In Section 4, we present results from an extensive simulation study. In Section 5, we summarize our main observations, and Section 6 contains practical advice and suggestions for future research.

## 2. RELATED WORK

The methods studied in this paper— $K$ -means, hierarchical clustering, topological data analysis (TDA), MDS, and t-SNE—are well-established tools that have been applied across a wide range of domains. Understanding how they have been used in practice motivates the need for the systematic comparative study presented here.

In computational biology, dimensionality reduction followed by clustering is standard practice for analyzing single-cell RNA sequencing (scRNA-seq) data<sup>3</sup>. Researchers reduce the gene-expression space (often thousands of dimensions) to two or three dimensions using MDS or t-SNE before inferring cell lineages and cluster identities. The implicit assumption is that the dimensionality reduction step faithfully preserves cluster separability—an assumption our study directly examines under controlled simulation conditions.

In political science, Diaconis et al.<sup>4</sup> applied MDS to the 2005 U.S. House of Representatives roll call votes, producing a 3-D mapping that reveals voting patterns. Their work illustrates MDS as an exploratory visualization tool, but does not evaluate how subsequent clustering is affected by the projection step, nor how this changes as the original dimension increases. Our study fills this gap by measuring the effect of MDS and t-SNE on clustering accuracy across a range of dimensions.

Concept mapping, as used in policy evaluation<sup>5</sup>, applies MDS to “pile sorting” data to identify thematic clusters in survey responses. Again, the quality of the final clustering depends on whether the low-dimensional embedding preserves the true groupings—a question we address empirically for multiple data configurations and cluster shapes.

These applications collectively demonstrate that the pipeline of dimensionality reduction followed by clustering is widely

practiced yet its reliability is rarely evaluated in a controlled manner. This is the precise gap our simulation study addresses.

### 3. METHODS AND PROCEDURES

Cluster analysis is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups. These groups are called *clusters*. The effectiveness of clustering algorithms is highly dependent on the type of data used.

In this section, we describe three approaches to clustering. The first is called *K*-means, the second is a family of clustering algorithms called hierarchical, and the third approach is called topological data analysis (TDA). Regardless of the approach, in this paper we will assume that the data consists of  $N$  points in an  $n$ -dimensional Euclidean space, and the distance between two points,  $x, y \in \mathbb{R}^n$  is computed in the usual way, by summing the squared differences in all the coordinates:

$$d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}.$$

In the following, when we refer to distance, we mean this Euclidean norm. However, other distance metrics may be used for clustering.

We now elaborate on some well-known clustering methods, all of which require the number of clusters to be specified. The following clustering algorithms are included in scikit-learn<sup>6</sup>, a Python machine-learning package. We selected *K*-means and hierarchical clustering (with all four linkage criteria available in scikit-learn) because they are among the most widely used clustering algorithms in practice, and because they span a broad spectrum of geometric assumptions—from the spherical cluster assumption of *K*-means to the flexible, shape-sensitive behavior of single linkage. We include TDA (specifically, the ToMATo algorithm) as a representative topology-based approach that is theoretically capable of capturing non-convex cluster shapes without relying on distance-to-centroid measures. This selection directly reflects the research questions posed above, enabling a direct comparison of centroid-based, linkage-based, and topology-based clustering strategies.

#### 3.1. *K*-Means

A popular algorithm for clustering is *K*-means. It has been used across a wide range of application areas in many different disciplines. The *K*-means algorithm works by assigning data points to  $K$  clusters, each represented by a centroid, usually the mean of the points in that cluster.

The algorithm follows these steps:

1. Initialization: Select  $K$  initial centroids from the dataset.
2. Assignment: Assign each data point to the nearest centroid.
3. Update: Compute new centroids by averaging the points in each cluster.
4. Iteration: Repeat the assignment and update steps until the centroids stabilize or the change is below a defined threshold.

The *K*-means algorithm aims to choose centroids that minimize a criterion known as the *inertia* or within-cluster sum of squares. That is, if  $x_i \in C_j$ , the set of all data points assigned to the  $j$ -th cluster, and  $\mu_j$  is the centroid of the  $j$ -th cluster, then the *K*-means algorithm aims to find  $\mu_1, \dots, \mu_K$  such that

$$I = \min_{\mu_1, \dots, \mu_K} \sum_{j=1}^K \sum_{i: x_i \in C_j} \|x_i - \mu_j\|^2.$$

Inertia measures cluster compactness, with lower values indicating more cohesive clusters.

$K$ -means has a few limitations. First, it performs poorly in the presence of elongated clusters and manifolds with irregular shapes due to its reliance on the Euclidean distance. Another limitation of  $K$ -means is that different random initializations may give different results. A third limitation is that, if the data we use for clustering is given in a high-dimensional space,  $\mathbb{R}^n$ , but the true clusters are defined in terms of just  $m$  features, where  $m < n$ , then each additional feature (coordinate) adds some noise which inflates the distances, and this has a negative impact on clustering performance. (This applies to other methods, and not just  $K$ -means). Suppose that the noise in each coordinate that has no role in defining the true clusters has a standard normal distribution,  $N(0, 1)$ . The sum of squares of such superfluous coordinates has a chi-square distribution with  $n - m$  degrees of freedom and, importantly, a mean of  $n - m$ . Thus, when  $n$  is much greater than  $m$ , the distances are dominated by noise rather than by the actual cluster configuration. This makes distance calculations less meaningful. It is often recommended to run a dimensionality reduction algorithm prior to  $K$ -means clustering to alleviate this problem and speed up computations.

### 3.2. Hierarchical Clustering

Hierarchical clustering is a method in which clusters are formed sequentially – either *top-down* (divisive clustering), where in the first step all points are put in one cluster, and then in each step one cluster is split into two until each point is its own cluster; or *bottom-up* (agglomerative clustering), where each observation starts in its own cluster, and in each step two clusters are merged together until all the points are in one cluster.

The hierarchy resulting from this process is represented as a tree, or dendrogram. Hierarchical clustering is deterministic and will give the same result each time.

We use agglomerative clustering in this paper. With this method, joining the next two closest clusters in each step depends on how one defines ‘closest’, because, even when one decides the distance metric between points (the Euclidean, in our case), there is more than one way to define the distance between sets. The literature also uses the term *linkage criteria* to refer to the chosen definition of distance between sets, since this choice determines which two clusters will be linked (merged) in the next step.

Scikit-learn’s AgglomerativeClustering object offers four possible linkage criteria, which we describe now. We use the notation  $A$  and  $B$  to define sets (clusters), and  $i$  and  $j$  to denote points in these sets.

**Single linkage** minimizes the distance between the closest pair of observations from different clusters. In other words, the distance between two clusters is defined as the smallest distance between *any* pair of points from disjoint clusters. Using mathematical notation, the single linkage distance between sets is defined as:

$$D(A, B) = \min_{i \in A, j \in B} d(i, j),$$

where  $d(i, j)$  is the Euclidean distance between elements  $i$  and  $j$  in clusters  $A$  and  $B$ , respectively.

**Complete linkage** minimizes the maximum distance between pairs of points from different clusters. Complete linkage is thus based on the following definition of distance between sets:

$$D(A, B) = \max_{i \in A, j \in B} d(i, j).$$

**Average linkage** minimizes the *average* of the distances between all pairs of points from two disjoint clusters.

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{i \in A} \sum_{j \in B} d(i, j)$$

where  $|A|$  and  $|B|$  are the number of elements in clusters A and B, respectively.

**Ward linkage** minimizes the sum of squared differences within all clusters<sup>7</sup>. It minimizes variance similar to the  $K$ -means objective function. The distance between sets for Ward linkage is expressed as:

$$D(A, B) = \sqrt{\frac{|A| \cdot |B|}{|A| + |B|}} \cdot \|\bar{a} - \bar{b}\|,$$

where  $|A|$  and  $|B|$  are the number of elements in clusters A and B, respectively,  $\bar{a}$  and  $\bar{b}$  are the centroids (mean points) of clusters A and B respectively, and  $\|\bar{a} - \bar{b}\|$  is the Euclidean distance between the centroids of clusters A and B.

### 3.3. Topological Data Analysis

Topological data analysis (TDA) is a technique for identifying structural patterns in complex datasets using persistent homology, a method that captures the underlying shape of data across multiple dimensions. TDA can quantify shape and structure in data, making it a topology-based method which is in theory resistant to noise, scalable to larger data sets, and capable of capturing complex geometric shapes. We find, however, that it is not necessarily so, and TDA may be severely affected by the dimensionality of the data and the noise level.

We will use the Topological Mode Analysis Tool (ToMATo) Algorithm for TDA clustering<sup>8</sup>. ToMATo operates in three main steps:

1. **Density Estimation and Neighborhood Graph Construction:** The algorithm begins by estimating the density of points in the dataset. It then constructs a neighborhood graph, where each data point is connected to its nearest neighbors based on a chosen similarity measure.
2. **Mode-Seeking and Initial Cluster Formation:** ToMATo assigns each point to a local density maximum by following the paths of connected points in the neighborhood graph. This process groups points into initial clusters, each centered around a mode, a local maximum of the estimated data density that serves as the center of an initial cluster, representing a dense and potentially meaningful region of the data. A persistence diagram is created to measure the persistence of each mode before it gets merged with another. The birth time of a mode is when it first appears as an independent cluster. The death time is when it merges into a more dominant cluster. The difference (persistence = death - birth) quantifies the cluster's significance.
3. **Hierarchical Merging of Clusters:** Clusters are merged based on their persistence. Modes with low persistence, short-lived peaks, are considered less important and merged first. High-persistence modes, stable clusters, remain in the final result. Since merging follows a hierarchical structure, each new cluster is formed by combining two existing ones, producing a nested sequence of cluster formations.

### 3.4. Evaluating Clustering Performance

When evaluating the performance of a clustering algorithm, one must first determine which pairs of points were clustered correctly, and which were not. Since labels are assigned to clusters arbitrarily by a clustering algorithm, these labels may not match the true labels directly. Instead of relying on the arbitrary labels assigned by an algorithm, pairwise relationships between points are used. In other words, if two points that are assigned to the same cluster (regardless of the label) are also in the same cluster in the true clustering configuration, then this pair is considered correctly assigned. This leads to four possible outcomes for each pair, and we count the number of occurrences of each outcome over all the pairs:

- $a$  is the number of pairs of points that are in the same cluster in both the predicted clusters and true ones (true positives).
- $b$  is the number of pairs that are in the same cluster in the predicted configuration but not in the true ones (false positives).
- $c$  is the number of pairs that are in the same cluster in the true configuration but not in the predicted one (false negatives).
- $d$  is the number of pairs that are not in the same cluster, neither in the true configuration nor in the predicted one (true negatives).

One way to evaluate clustering performance is the *accuracy*, defined as the number of correct predictions over total predictions, or  $(a + d)/(a + b + c + d)$ . There are two problems with accuracy. One is that it may be inflated when the data is not balanced. That is, when some clusters are much larger than others. In such unbalanced cases, a simple clustering rule that puts all points in one cluster will exhibit high accuracy. The other problem is that when the number of true negatives ( $d$ ) is large, it dominates both the numerator and the denominator and yields an accuracy value close to 1. In clustering tasks, it is often the case that  $d$  is indeed large, since most pairs are not in the same cluster.

Thus, in the following we use the Jaccard index, which is defined as:

$$J = \frac{a}{a + b + c}.$$

Note that it is similar to the accuracy measure except that the true negatives are dropped from the calculation. The Jaccard Index<sup>9</sup> considers all pairs of elements and determines whether they are clustered together in both clusterings (intersection, the numerator) or in at least one of the clusterings (union, the denominator), which is why it is also referred to as the intersection over union (IoU) in the literature. The Jaccard index ranges from 0 to 1, where 0 means no agreement (completely different clusterings), and 1 indicates perfect agreement.

### 3.5. Dimensionality Reduction Methods

As mentioned above, clustering is especially challenging when the data is embedded in a high-dimensional space, but the true clusters are defined by lower-dimensional shapes. Thus, before clustering is performed, the first step is to try and reduce the dimensionality of the data. Because clustering is especially useful when the clusters can be visualized, it is often the case that the data is projected into a 2-D Euclidean space.

Although dimensionality reduction is an essential first step when performing clustering, it is important to note that it results in some loss of information. To illustrate, consider projecting the surface of the earth onto a 2-D map. In this process, distances are inevitably distorted. For example, on a standard world map Alaska and Russia appear on opposite sides, suggesting they are far apart, when in fact they are relatively close. This shows that when high-dimensional data are reduced to lower dimensions, distances between points can be distorted. Since there is already distortion from three dimensions to two, it stands to reason that there would be even greater distortion when reducing from much higher dimensions. Thus, measuring the distortion caused by the projection step is essential to evaluate the degree of misplacement of data points, and its impact on the ultimate task of clustering.

In this paper we will use two dimensionality reduction methods: Multidimensional scaling (MDS) and t-Distributed Stochastic Neighbor Embedding (t-SNE).

We begin with a brief description of MDS. Given a dissimilarity or distance matrix  $D$  representing the differences between all pairs of points, such that  $D_{ij}$  denotes the distance between data points  $i$  and  $j$  in the high-dimensional space,

multidimensional scaling (MDS) seeks a configuration of points in a low-dimensional space such that the corresponding Euclidean distances  $d_{ij}$  are as close as possible to  $D_{ij}$ , across all pairs  $i, j$ . Metric MDS achieves this by minimizing a loss function known as stress. We use Kruskal's<sup>1</sup> Stress-1, defined as

$$\text{Stress-1} = \sqrt{\frac{\sum_{i < j} (d_{ij} - D_{ij})^2}{\sum_{i < j} D_{ij}^2}}.$$

The summation is taken over all unordered pairs  $i < j$ . The Stress-1 criterion measures the normalized discrepancy between interpoint distances in the original and embedded spaces, with smaller stress values indicating a more faithful low-dimensional representation of the original distance structure.

Scikit-learn provides an MDS algorithm that is implemented in Python.

Next, we briefly describe t-SNE<sup>2</sup>. t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction method widely used for visualizing high-dimensional data in two or three dimensions. This method is based on converting pairwise similarities between data points into probability distributions and minimizing the divergence between these distributions in the high- and low-dimensional spaces.

In the high-dimensional space, similarities between points are modeled using conditional probabilities that reflect the likelihood of one point being a neighbor of another under a Gaussian distribution centered at the point of interest. In the low-dimensional embedding, similarities are modeled using a Student's t-distribution with one degree of freedom, which has heavier tails than a Gaussian and thus better preserves local structures while avoiding the "crowding problem." The optimization objective is to minimize the Kullback-Leibler (KL) divergence between the two distributions. This minimization problem is typically solved via gradient descent.

The scikit-learn library provides an efficient implementation of t-SNE through the class `sklearn.manifold.TSNE`. This implementation supports multiple optimization algorithms, enabling scalability to moderately large datasets.

We restrict our study to MDS and t-SNE because they are the dimensionality reduction methods most commonly encountered in the application domains described in Section 2, and because they represent two fundamentally different methodological families: metric (distance-preserving) methods exemplified by MDS, and distribution-based (neighbor-embedding) methods exemplified by t-SNE. This contrast is sufficient to illustrate the full range of behaviors relevant to our research questions. Other methods, most notably Uniform Manifold Approximation and Projection (UMAP)<sup>10</sup>, have gained increasing popularity and typically produce high-quality embeddings; evaluating UMAP alongside MDS and t-SNE is a natural direction for future work.

#### 4. RESULTS

We present simulations that were performed in order to assess whether or to what extent clustering algorithms are affected by three factors:

1. Noise level – the dispersion of points belonging to a distinct cluster. The larger the dispersion, the more likely it is that points will be found in the neighborhood, or even within, another cluster.
2. The dimensionality – as stated above, if the true dimension of the clusters,  $m$ , is smaller than the dimension of the data,  $n$ , then the superfluous dimensions contribute to noise and when  $n - m$  is large the noise overwhelms the signal, making clustering impossible.
3. The shapes of the clusters – in particular, we use both convex and non-convex examples. Several clustering algorithms rely on convexity (or even on sphericity) and are known to perform poorly when some clusters are non-convex.

We use three configurations in our simulations. In the first scenario we generate five clusters as ‘Gaussian clouds’ (or balls) in  $\mathbb{R}^4$ . In the second scenario, we use hyperspheres, which are similar to the clouds, except that the data is concentrated close to the surface of the five balls. The third configuration consists of two horseshoes in  $\mathbb{R}^2$  such that they do not overlap, but their convex hulls do.

Each scenario is repeated with varying degrees of noise. Furthermore, the data from each scenario is embedded in a higher dimensional space  $\mathbb{R}^n$ , with  $n = 4, 8, 16, 32, 64, 128, 256$ .

We compare clustering results in these configurations (in terms of the IOU, or the Jaccard index) before and after MDS and t-SNE are applied.

#### 4.1. Gaussian Clouds

We simulate five Gaussian clouds, each with 600 points, for a total of 3,000 points. The centers of the clouds span a four-dimensional space.

The centers are embedded in an  $n$ -dimensional space, and the points are sampled from an  $n$ -dimensional multivariate normal distribution around each center, with the cluster standard deviation set to  $\sqrt{n}/4$ . A 2-D depiction of the data after applying MDS is shown in Figure 1.

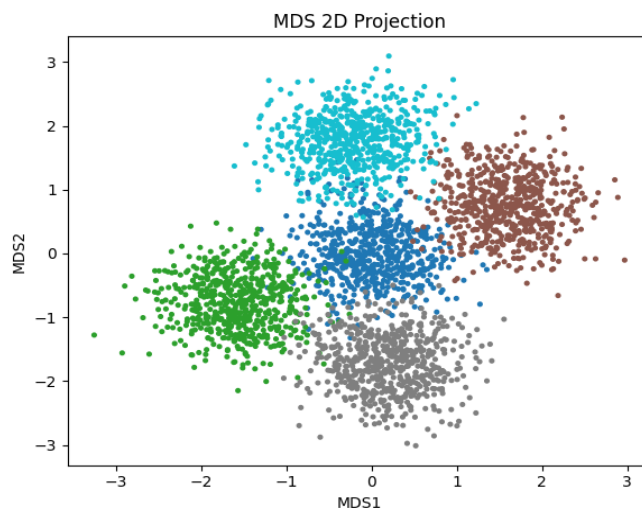
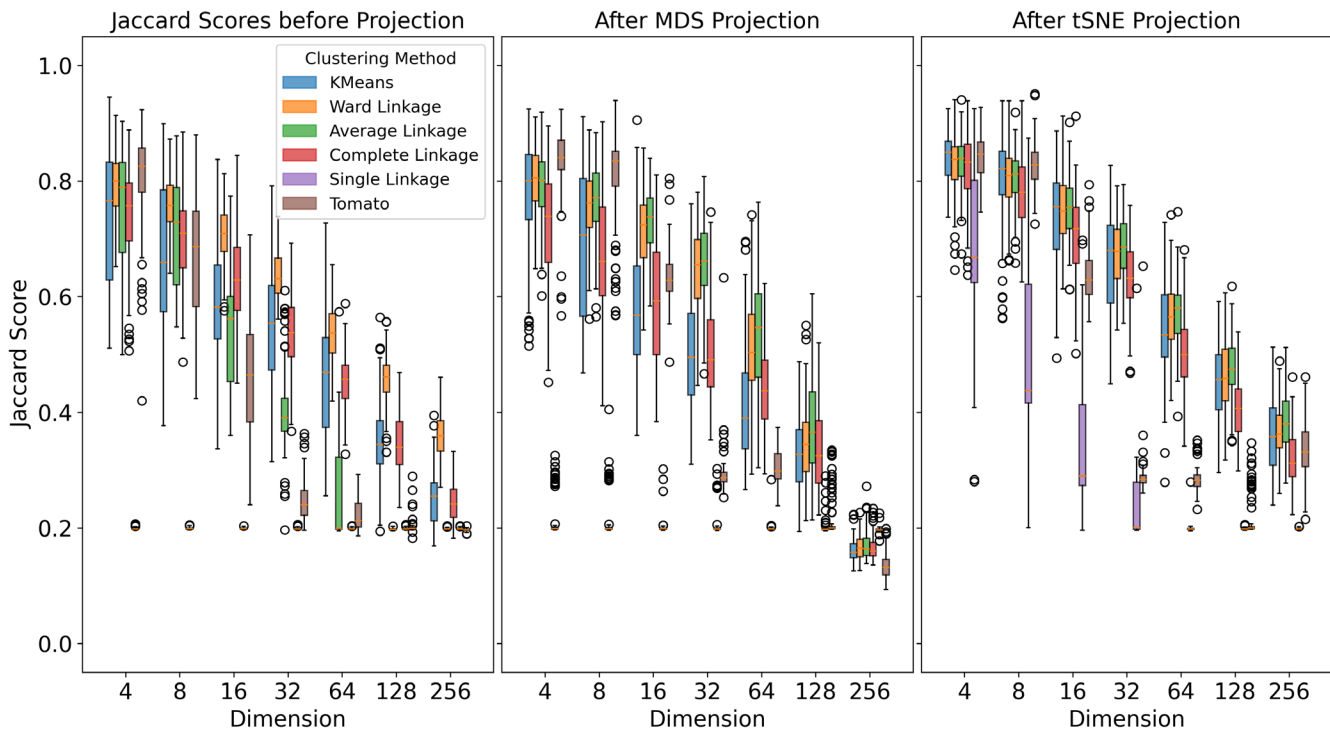


Figure 1. Visualization of the simulated dataset consisting of five Gaussian clusters, each with 600 points (3,000 points total). 2-D embedding obtained using MDS.

For each iteration in our simulation, we randomly draw 300 points from each cluster (out of its 600) to form a sample of 1,500 points. Here, and in the following scenarios, when using the cluster.KMeans function, we set the parameter `max_iter=1` to allow a valid comparison with the other clustering methods.

The results of six clustering methods are shown in Figure 2, for different values of  $n$  (the dimension of the space). The left figure shows boxplots of the Jaccard index obtained from 100 simulations when the six clustering methods were applied to the original data in  $\mathbb{R}^n$ . The other two figures show the results when clustering is applied after reducing the dimension to  $\mathbb{R}^2$  with MDS (middle figure) and t-SNE (right figure). Table 1 shows only the median Jaccard index for the six methods for  $n = 4$  and  $n = 64$ .

Figure 2 shows that all of the clustering methods decreased in performance as the dimension increased. This holds true both with the original (unscaled) data, and after dimensionality reduction. The rate at which the performance deteriorates is smallest when using t-SNE to reduce the dimension. The plots show that with this configuration TDA had a



**Figure 2.** Comparison of six clustering methods applied to the simulated clouds dataset across varying data dimensions  $n$  (from 4 to 256). The left panel shows boxplots of the Jaccard index from 100 simulations on the original  $\mathbb{R}^n$  data. The middle and right panels show performance after reducing the data to  $\mathbb{R}^2$  using MDS and t-SNE, respectively. For all methods, the KMeans algorithm was limited to a single iteration per sample to ensure a fair comparison.

Clouds	Dim 4			Dim 64		
	In $\mathbb{R}^n$	After MDS	After t-SNE	In $\mathbb{R}^n$	After MDS	After t-SNE
<b>K-means</b>	0.766	0.801	<b>0.850*</b>	0.469	0.390	0.534
<b>Ward linkage</b>	0.800	0.805	0.838	0.537*	0.503	0.565
<b>Average linkage</b>	0.789	0.801	0.839	0.199	0.546*	<b>0.580*</b>
<b>Complete linkage</b>	0.757	0.739	0.833	0.457	0.437	0.499
<b>Single linkage</b>	0.198*	0.199*	0.668*	0.198*	0.198*	0.198*
<b>TDA</b>	0.826*	0.840*	0.847	0.213	0.298	0.282

**Table 1.** Median Clustering Performance – Gaussian clouds. A superscript/subscript asterisk denotes the best/worst performance in a column, respectively. Bold-face denotes the best performance overall, for a fixed  $n$ .

slightly better performance than the other methods when  $n = 4$ , and also when  $n = 8$  but only after dimensionality reduction. However, in higher dimensions the performance of TDA (with and without dimensionality reduction) quickly deteriorates.

The worst performance among all clustering methods in this scenario was that of single linkage. Although with  $n = 4$  t-SNE greatly improved the median Jaccard index (from 0.198 to 0.668, see Table 1), it is still far below all other methods, and in higher dimension it performed very poorly.

Without dimensionality reduction, Ward linkage had the best performance in higher dimensions. With MDS as the dimensionality reduction method, average linkage was as good as Ward’s. K-means and complete linkage were not as good, especially as  $n$  increased.

In dimension 4, MDS offered a slight improvement over clustering without scaling, except for complete linkage, where the median was decreased a little, and single linkage which performed poorly in this configuration. Running t-SNE prior to clustering improved the median Jaccard index for all methods. In dimension 64, running t-SNE prior to clus-

tering improved the median Jaccard index for all methods except for single linkage, and the most dramatic improvement was for average linkage. MDS improved the average linkage and TDA performance, but for the other four methods the median score decreased.

4.2. Hyperspheres

The second scenario is similar to the previous one, except that the points in each cluster are concentrated near the surface, as opposed to being scattered uniformly in the ball. Note that unlike the previous example, hyperspheres are non-convex sets. This is easy to see from the triangle inequality. As before, we simulate five spheres, each with 600 points, for a total of 3,000 points. Points are drawn uniformly on a sphere with a fixed radius of  $\sqrt{n}/2$ .

In our simulation, we increase the dimension in which the clusters are embedded from 4 to 256. We run the simulation 100 times, each time drawing a sample size of 300 points out of the 600 from each cluster. Figure 3 shows the boxplot of the Jaccard index for the six clustering methods, over a range of dimensions. The left figure shows the results without a dimensionality reduction step. The middle/right figures show the results when clustering is performed after MDS/t-SNE, respectively. Table 2 shows the medians for the cases of  $n = 4$  and  $n = 64$ .

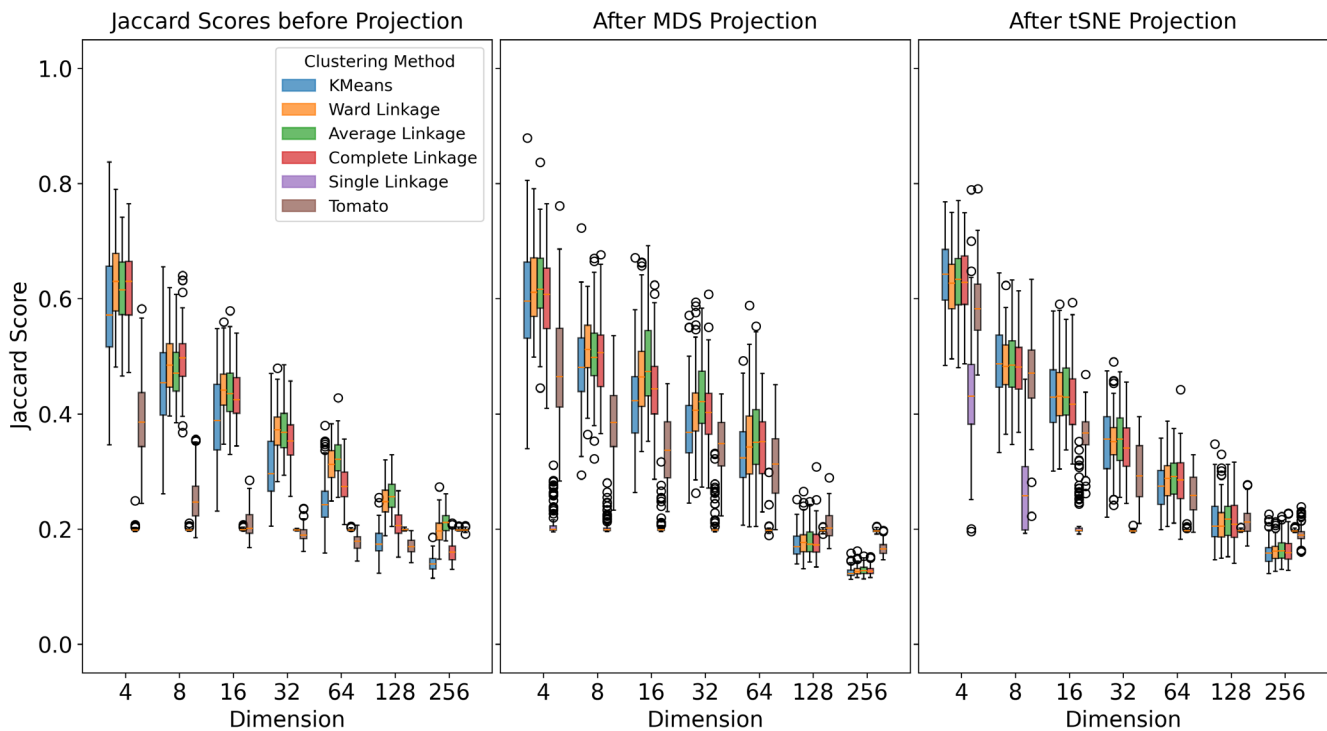


Figure 3. Comparison of clustering methods applied to the simulated hyperspheres dataset for varying dimensions  $n$  (from 4 to 256). The left panel shows boxplots of the Jaccard index from 100 simulations on the original  $\mathbb{R}^n$  data. The middle and right panels show performance after reducing the data to  $\mathbb{R}^2$  using MDS and t-SNE, respectively.

Hyperspheres	Dim 4			Dim 64		
	In $\mathbb{R}^n$	After MDS	After t-SNE	In $\mathbb{R}^n$	After MDS	After t-SNE
K-means	0.572	0.596	<b>0.642*</b>	0.242	0.323	0.274
Ward linkage	0.630*	0.611	0.627	0.312	0.342	0.287
Average linkage	0.615	0.616*	0.633	0.321*	0.350	0.291*
Complete linkage	0.630*	0.608	0.628	0.273	<b>0.351*</b>	0.285
Single linkage	0.198 <sub>*</sub>	0.199 <sub>*</sub>	0.431 <sub>*</sub>	0.198	0.197 <sub>*</sub>	0.197 <sub>*</sub>
TDA	0.385	0.464	0.582	0.179 <sub>*</sub>	0.313	0.258

Table 2. Median Clustering Performance – hyperspheres. A superscript/subscript asterisk denotes the best/worst performance in a column, respectively. Bold-face denotes the best performance overall, for a fixed  $n$ .

In some regards the results are similar to the previous scenario: all of the clustering methods decreased in performance

as the dimension increased, and the worst performer was single linkage. Also, the performance of TDA deteriorates most rapidly, followed by *K*-means, which is also negatively affected by the increase in *n*.

We see that although this scenario is similar to the previous one, the fact that the clusters are not uniformly distributed in a ball, but are concentrated near the surface, presents a challenge to all clustering methods. The overall scores are lower than in the previous case, and deteriorate rather quickly. In this scenario, MDS was somewhat effective in improving the clustering in high dimension, while t-SNE resulted in lower scores when compared with no dimensionality reduction (except for a slight improvement for complete linkage; see Table 2).

In dimension 4, Ward linkage had the highest median Jaccard index before dimensional scaling, followed closely by complete linkage. Average linkage had the highest Jaccard index after MDS. *K*-means had the highest median Jaccard index after t-SNE. In dimension 64, average linkage had the highest median Jaccard index before dimensional scaling, and also after t-SNE. Complete linkage was best after MDS (0.351), and overall in this scenario with *n* = 64.

### 4.3. Two Horseshoes

For another non-convex scenario, we simulate two horseshoes, each with 2,000 points, for a total of 4,000 points. Figure 4 shows the data in 2-D, after performing MDS.

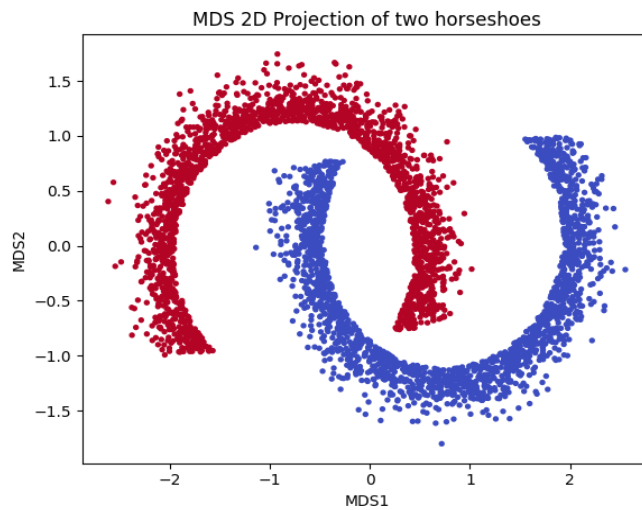


Figure 4. Visualization of the simulated dataset consisting of two non-convex clusters, each with 2,000 points (4,000 points total). 2-D embedding obtained using MDS.

In our simulations, we increase the dimension from 4 to 64. In each of 100 iterations we draw a random sample of size of 200 points from each cluster. The medians from each method for *n* = 4 and *n* = 64 are shown in Table 3, and the boxplots from 100 iterations with each *n* are shown in Figure 5.

Two Horseshoes	Dim 4			Dim 64		
	In $\mathbb{R}^n$	After MDS	After t-SNE	In $\mathbb{R}^n$	After MDS	After t-SNE
<b>K-means</b>	0.440*	0.447*	0.645	0.441*	0.438*	0.643
<b>Ward linkage</b>	0.537	0.536	0.558	0.525	0.546*	0.576
<b>Average linkage</b>	0.524	0.524	0.525	0.528	0.526	0.539
<b>Complete linkage</b>	0.538*	0.550*	0.581	0.543*	0.551	0.586
<b>Single linkage</b>	0.498	0.498	<b>1.000*</b>	0.498	0.500	<b>1.000*</b>
<b>TDA</b>	0.476	0.486	0.500	0.461	0.487	0.483

Table 3. Median Clustering Performance – two horseshoes. A superscript/subscript asterisk denotes the best/worst performance in a column, respectively. Bold-face denotes the best performance overall, for a fixed *n*.

The results in this scenario are strikingly different from the previous examples. First, regardless of *n*, all methods struggle to find the correct clusters, as indicated by a Jaccard index which is approximately 0.5 for most methods. The one

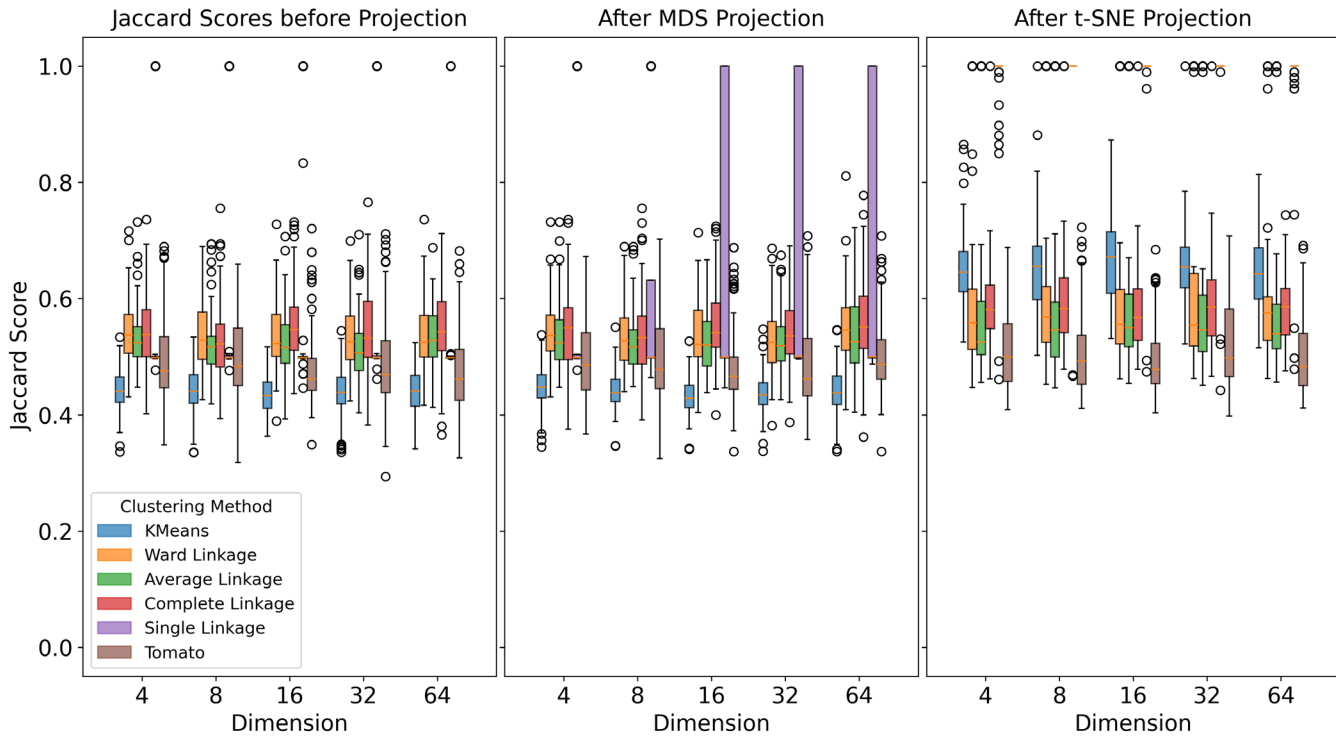


Figure 5. Comparison of clustering methods applied to the two-horseshoes dataset for varying dimensions  $n$  (from 4 to 64). The left panel shows boxplots of the Jaccard index from 100 simulations on the original  $\mathbb{R}^n$  data. The middle and right panels show performance after reducing the data to  $\mathbb{R}^2$  using MDS and t-SNE, respectively.

method that appears to perform better in some of the simulations is the single linkage, which was the worst method in the previous examples. However, to achieve this improved performance one must start with t-SNE. In this scenario  $K$ -means has the worst performance when no dimensionality reduction is performed, or if MDS is performed. On the other hand, when t-SNE is performed,  $K$ -means performs better (increasing from 0.44 to 0.64), but not nearly as good as single linkage with t-SNE.

### 5. DISCUSSION

The purpose of this study was to evaluate how different clustering algorithms perform across a variety of data structures and to determine whether or to what extent clustering is affected by the dimension of the space containing the clusters. We also investigated in what way dimensionality reduction methods influence the accuracy of these clustering outcomes. We used the Jaccard index (also known as intersection over union, or IoU) to assess the quality of each clustering algorithm.

First, we observe that the simulations demonstrate clustering performance depends strongly on the geometric structure of the data and the presence of noise, even before dimensionality reduction is applied. We used three configurations, and found that when the clusters are convex, the performance was better, overall, when compared with non-convex shapes. Clustering in the Gaussian clouds scenario (4.1) had better IOUs than the similar, but non-convex setting of hyperspheres (4.2), and both had much better results than the third scenario (4.3), in which not only the shapes are non-convex, but their convex hulls intersect.

Second, we find that no single clustering method consistently performs well across all data configurations or outperforms the other methods. For example,  $K$ -means and Ward linkage tended to perform well in the first two scenarios when the underlying clusters were approximately spherical. However, their performance was poor in the third scenario. In contrast, single linkage showed sensitivity to noise and performed the worst among all methods for both the clouds

and spheres scenarios, yet had the best performance for the non-convex horseshoe data after t-SNE. Topological data analysis (TDA), using the ToMATo algorithm, did not have the best clustering performance in any of the three scenarios we have tested.

Third, we find that the higher the dimension in which the clusters are embedded, the harder the clustering task, regardless of the method. In some cases, performing a dimensionality reduction before clustering does improve the result. However, this is not always the case (e.g., the two horseshoes) and the positive effect of the dimensionality reduction step depends on the shapes. In general, between the two methods that we used, namely, MDS and t-SNE, the latter is preferred, but that is not always the case. For example, with the hyperspheres in  $\mathbb{R}^{64}$ , complete linkage after MDS yielded the best result (albeit, a rather poor one), while in the clouds scenario, the better results were achieved if t-SNE was used. See Figures 2 and 3.

## 6. CONCLUSIONS

The results highlighted here have important practical consequences. It is important to keep in mind that with real data we do not know whether the clusters are convex or not, or whether their convex hulls intersect, nor do we know the actual intrinsic dimension of the data, nor the magnitude or distribution of noise. As we have demonstrated, there is no way to choose the ‘best’ clustering algorithm, as each one depends on the (unknown) shapes in different ways. The effectiveness of each method depends fundamentally on the geometry and noise characteristics of the data. Methods like  $K$ -means, Ward linkage, average linkage, and complete linkage have better clustering performance in settings with compact, spherical clusters, while single linkage offers advantages for clustering well-separated, non-convex data.

Furthermore, if low-dimensional clusters are embedded in a high-dimensional space, the contribution of superfluous coordinates to the distance between points is substantial and obscures meaningful geometric relationships as the dimension of the space becomes much larger than the intrinsic dimension of the clusters. Dimensionality reduction techniques may alleviate this problem in some cases, but not always. In some cases, clustering in the original high-dimensional space provided more accurate results than after projection to lower dimensions. Our findings suggest that there is no clear winner between the dimensionality reduction methods we have used.

Finally, we point out that in our analysis we provided the correct number of clusters, but in general this is another unknown and has to be estimated by the user. The general approach is to vary  $k$  and choose the one that gives the best information criterion (e.g., AIC, BIC). This is beyond the scope of this study, but it is yet another challenge facing practitioners who perform clustering.

We recommend that analysts perform cross-validation; splitting the data into training and testing data, and then evaluate multiple clustering methods and dimensionality reduction methods for any given dataset, rather than relying on a preselected clustering method or on MDS or t-SNE plots.

Future work includes determining the intrinsic dimension of datasets for better clustering performance (and possibly a better way to select the correct number of clusters) and leveraging AI to improve current clustering methods. Such improvements have the potential to improve deep-learning-based segmentation methods such as nnU-Net<sup>11</sup> to detect tumors in medical imaging.

## ACKNOWLEDGEMENTS

The authors thank the Holster Scholar Grant for supporting this project.

## REFERENCES

1. Kruskal, Joseph B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* 29.1 : 1-27. <https://doi.org/10.1007/BF02289565>
2. Maaten, L.V., and Hinton, G.E. (2008) Visualizing Data using t-SNE, *Journal of Machine Learning Research* 9, 2579-

2605.

3. Van den Berge, K., Roux de Bézieux, H., and Street, K. (2020) Trajectory-based differential expression analysis for single-cell sequencing data, *Nature Communications* 11, 1201. <https://doi.org/10.1038/s41467-020-14766-3>
4. Diaconis, P., Goel, S., and Holmes, S. (2008) Horseshoes in multidimensional scaling and local kernel methods, *The Annals of Applied Statistics* 2(3), 777–807. <https://doi.org/10.1214/08-AOAS165>
5. Trochim, W. M., and McLinden, D. (2017) Introduction to a Special Issue on Concept Mapping, *Evaluation and Program Planning* 60, 166–175. <https://doi.org/10.1016/j.evalprogplan.2016.10.006>
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* 12, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
7. Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association* 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
8. Chazal, F., Guibas, L. J., Oudot, S. Y., and Skraba, P. (2013) Persistence-Based Clustering in Riemannian Manifolds, *Journal of the ACM* 60(6), Article 41. <https://doi.org/10.1145/2535927>
9. Jaccard, P. (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura, *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 547–579.
10. McInnes, L., Healy, J., and Melville, J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv preprint arXiv:1802.03426*. <https://arxiv.org/abs/1802.03426>
11. Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, *Nat Methods* 18, 203–211. <https://doi.org/10.1038/s41592-020-01008-z>
12. Dexter, E., Rollwagen-Bollens, G., and Bollens, S. M. (2018) The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling, *Limnology and Oceanography: Methods* 16(7), 434–443. <https://doi.org/10.1002/lom3.10257>
13. Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007) Stability of Persistence Diagrams, *Discrete and Computational Geometry* 37, 103–120. <https://doi.org/10.1007/s00454-006-1276-5>

#### ABOUT THE STUDENT AUTHOR

Lucy Liu is a junior majoring in both Statistics and Applied Mathematics, and she plans to graduate in the Spring of 2027.

#### PRESS SUMMARY

Clustering is a widely used technique for finding patterns or groups within data. This simulation study investigates the performance of several clustering methods when the clusters are embedded in high dimensional spaces, and whether clustering performance is improved if one first performs dimensionality reduction. By testing various data scenarios and dimensions, and applying several clustering methods, we observed that (i) dimensionality reduction methods such as MDS and t-SNE sometimes weaken clustering accuracy, (ii) higher dimensions of data reduce clustering performance, and (iii) no single clustering method works best in all settings.

# Framing the News: How an Article's Headline and Content Valence Shape Readers' Understanding and Impressions

Kyle Smoak\*, Kenneth Barideaux Jr., & Christa Christ

Department of Psychology, University of South Carolina Upstate, Spartanburg, SC

<https://doi.org/10.33697/ajur.2026.174>

Student: [kmsmoak@email.uscupstate.edu](mailto:kmsmoak@email.uscupstate.edu)

Mentors: [kbaridea@uscupstate.edu](mailto:kbaridea@uscupstate.edu); [cchrist@uscupstate.edu](mailto:cchrist@uscupstate.edu)

## ABSTRACT

Media outlets often use “clickbait” or misleading headlines to increase engagement, but these strategies may distort reader understanding. Although prior research shows that negative information is more likely to capture attention and spread online, most studies have focused on the dissemination of news on social media rather than on readers’ cognitive processing. Fewer studies have examined how headline features, such as emotional valence and congruence with article content, influence judgments, comprehension, and behavioral intentions. Across two experiments, we investigated how headline valence and headline–article congruence influenced readers’ impressions and comprehension. In Study 1, college students read an article with a positive, negative, or neutral headline and rated their impressions of the article. In Study 2, participants read articles written in a positive or negative tone, paired with either congruent or incongruent headlines, and completed the same impression measures plus comprehension questions. Study 2 also included a replication of Study 1 as well as an independent replication sample for the unique Study 2 analysis using a national sample obtained using Prolific. Across both studies, headline valence did not significantly affect participants’ impressions. However, congruency between the headline and article did influence comprehension in one sample. Together, these findings suggest that the relationship between headline framing and news comprehension is complex and may depend on headline–article alignment as well as contextual and individual factors.

## KEYWORDS

Headlines; News Articles; Valence; Comprehension; Negativity Bias; Clickbait; Reader Impressions; Article Tone

## INTRODUCTION

With the rise of technology, people can now easily access news and information with just the click of a button. Recent research indicates that 86% of U.S. adults obtain news through a smartphone, computer, or tablet.<sup>1</sup> While this accessibility allows for greater public engagement with current events, it has also intensified competition among media outlets. As a potential response to this competition, many online news platforms have introduced new methods of attracting audiences, including clickbait headlines designed to grab the reader's attention.<sup>2</sup> Previous research has demonstrated that headlines play a critical role in guiding readers' attention and influencing their decisions about which articles to read.<sup>3</sup> The limited background information that is provided by these clickbait headlines urges readers to pursue corresponding articles to provide context.<sup>4</sup> Unfortunately, emerging evidence suggests that strategies such as using clickbait headlines may undermine public understanding by distorting how readers process information.<sup>5</sup> Additionally, strategic use of epistemic language in headlines could undermine public trust or interest in the news.<sup>6</sup> Therefore, more research is needed to uncover which headline characteristics are related to how one cognitively and emotionally interacts with news content.

The influence of news headlines stems from the fact that it is typically the first element of an article that is encoded and is often interpreted through the lens of the reader's prior knowledge.<sup>7</sup> In other words, the extent to which a headline activates relevant background knowledge can shape how new information is integrated with preexisting knowledge, ultimately affecting what information the reader retains.<sup>8,9</sup> For instance, Surber and Schroeder<sup>9</sup> found that headings improved recall of high-importance information, particularly for individuals with greater prior knowledge of the topic.

While headlines help activate the necessary mental schema for understanding an article, the way a headline is framed also plays an important role in shaping comprehension and overall interest in the content, especially when considering the emotional valence of the language used. Readers tend to fixate more on<sup>10</sup> and have higher discrimination for<sup>11</sup> negatively valenced text compared to neutral or positive content. Furthermore, readers show a stronger preference for negative headlines and additionally skew the negativity of the headline when sharing the information.<sup>12</sup> This negativity bias appears consistent across modalities, such as

reading versus listening to the news.<sup>13</sup> Moreover, the effect of negative language on processing speed may vary depending on individual differences.<sup>14,15</sup> For instance, Lei *et al.*<sup>15</sup> found that individuals with a high need for affect and strong narrative absorption read positive words more slowly when engaged with a nonfiction story about a missing person, suggesting that emotional predispositions may also influence how headline valence affects information processing.

From a theoretical perspective, negative stimuli are thought to attract more attention due to their evolutionary significance.<sup>16</sup> This phenomenon, known as the negativity bias, suggests that negative information may signal potential threats, leading to heightened attention as a survival mechanism. However, some researchers argue that emotional arousal, rather than valence alone, may play a more important role in capturing attention.<sup>17</sup> Emotional arousal refers to an increased physiological and psychological state in response to a stimulus, which naturally enhances alertness and focus. As a result, some studies indicate that individuals are more likely to attend to and remember stimuli that are highly emotionally arousing, regardless of whether they are positive or negative in valence.<sup>18,19</sup>

Given that negative or emotionally arousing information tends to be preferentially processed, it is reasonable to predict that such content would influence both memory and comprehension. Research supports this prediction, showing that negatively charged emotional content often narrows attentional focus to the most salient aspects of an event (e.g., blood at a crime scene), thereby enhancing memory for those details.<sup>20</sup> This effect, known as “emotional memory narrowing,” has been observed across a variety of contexts, including natural disasters,<sup>21</sup> childhood abuse and maltreatment,<sup>22</sup> and physical injuries.<sup>23</sup> However, this focused attention comes at a cost: while memory for emotionally negative features may be strengthened, memory for neutral or positive aspects of the experience may be diminished or even impaired.<sup>20</sup> The ability to use negative text to capture attention and improve reader retention provides an incentive for news outlets to use negativity to drive engagement, despite the potential ramifications for how such negativity bias could influence society more broadly.

For example, Mujica and Bachmann<sup>24</sup> studied how exposure to different levels of melodramatization in news stories affected participants’ recall and understanding. Their results showed that non-melodramatic versions of news reports resulted in better understanding of the causes and effects described, regardless of the topic. The authors suggest that highly dramatic content may trigger strong emotional responses that, when negative, hinder understanding. Although their study did not directly manipulate headlines, headlines often set the emotional valence and shape how a story is initially interpreted. When headlines carry a negative valence, they may prompt readers to approach the content less analytically, reducing their ability to differentiate and integrate information, ultimately impairing comprehension.<sup>25</sup>

Though the purpose of a headline is to summarize the main idea of an article and help readers choose which articles to read,<sup>26</sup> there are many instances where the headline and the article’s content are incongruent. These misleading headlines are usually associated with fake news and contribute to the spread of misinformation online. In a study by Carcioppolo *et al.*,<sup>5</sup> participants who read an article with an accurate headline had higher recognition and comprehension scores than those reading a clickbait question headline (e.g., “Does green tea extract cause liver damage?”) and a clickbait exaggerated headline (e.g., “White wine has a scary link to skin cancer.”). The study also showed that reading corrected information within an article was not enough to counteract the negative effects of a misleading headline. In a similar study, Ecker *et al.*<sup>26</sup> found that even when misleading headlines are corrected within the article, the initial misinformation often continues to influence readers’ understanding.

### *The Current Study*

While existing research has investigated the preference for negative information in news reading, current approaches have primarily focused on the dissemination process itself.<sup>12</sup> This has been driven by the rise of digital social media platforms, where reposting and sharing news articles have become a dominant form of engagement. Although the dissemination process helps us to understand social behavior, like the spread of information in online environments, more research is needed to better understand the information-processing aspect of news consumption.

While Ecker *et al.*<sup>26</sup> demonstrated that factually misleading headlines continue to influence comprehension even when corrected in the article text, less is known about how emotionally incongruent headlines, in which the emotional tone differs from the article content while facts remain consistent, affect information processing. The present research addresses this gap by examining whether emotional valence incongruence between headlines and articles influences comprehension and reader impressions. Like prior research on headline effects, we selected article topics relevant to our participant population (i.e., U.S. college students), as well as the population more broadly, to maintain ecological validity and engagement.

The present research aimed to address these gaps through two experiments. Study 1 sought to confirm the negativity bias by investigating how the emotional valence of a news headline (positive, negative, or neutral) influences readers’ judgments and behavioral intentions, such as interest in the article, perceived informativeness, and willingness to share. We hypothesized that

participants in the negative valence group would report significantly higher levels of interest, perceived informativeness, and likelihood of sharing the article than those in the neutral or positive valence groups, consistent with prior findings that negative information is more captivating.<sup>12</sup> Additionally, we hypothesized that participants in the negative valence group would rate the article as more biased, as negatively framed headlines may be perceived as more opinionated or emotionally charged than neutral or positive ones.

Study 2 focused on the impact of headline–article congruency on reading comprehension. Participants read an article preceded by a headline that was either congruent (i.e., aligned in valence and content) or incongruent (i.e., contradictory or misleading relative to the article’s content). We hypothesized that the negative headline groups would have higher reading comprehension than the positive headline groups. We also hypothesized that participants exposed to congruent headlines would demonstrate significantly better comprehension of the article than those exposed to incongruent headlines, based on the idea that congruency facilitates smoother cognitive integration and reduces confusion or misdirection caused by misleading cues,<sup>26</sup> and that this difference would be larger for the negative headline group.

## METHODS AND PROCEDURES

### STUDY 1

#### *Participants*

All 116 participants were sourced from a university in the south-eastern U.S. using the school’s research recruitment system (SONA, 2025). Of these participants, 89 (76.7%) identified as female, 23 (19.8%) identified as male, 2 (1.7%) identified as gender non-binary, 1 (0.9%) identified as transgender male, and 1 (0.9%) identified as genderfluid. Racial demographics included 58 (50.0%) Caucasian/White, 39 (33.6%) African American/Black, 10 (8.6%) Hispanic/Latin, 6 (5.2%) Asian American, 1 (0.9%) Native American, and 2 (1.7%) other. The age of participants was between 16 and 45 ( $M = 19.23$ ,  $SD = 3.09$ ). Participant political ideology included 34 (29.3%) conservatives, 55 (47.4%) moderates, and 26 (22.4%) liberals. When asked how often they pay attention to the news, 21 (18.1%) said daily, 53 (45.7%) said weekly, 30 (25.9%) said monthly, and 12 (10.3%) said yearly. When asked where they get most of their news, 26 (22.4%) said online news outlets, 80 (69.0%) said social media, and 10 (8.6%) said television stations. When asked about their familiarity with the topic of censorship, 9 (7.8%) said they were very unfamiliar, 6 (5.2%) said they were unfamiliar, 45 (38.8%) said they were neither unfamiliar nor familiar, 41 (35.3%) said they were familiar, and 15 (12.9%) said they were very familiar. Before completing the analysis, one participant was excluded due to having taken more than 80 minutes to complete the 10–15-minute study.

#### *Materials*

The independent variable manipulation involved two experimental groups and one control group. All content focused on censorship, a highly debated issue in today’s political climate that has received extensive coverage across various online media outlets.<sup>27</sup> The experimental groups included a positive valence group, where the headline “Censorship, and The Balancing of Protected Rights with Wanting to Improve Information Security” included the two distinct positively valenced words “protected” and “improve”, and a negative valence group, where the headline “Censorship, and the Potential Slaughter of Rights Versus Avoiding the Hurt of Unsecured Information” included the two distinct negatively valenced words “slaughter” and “hurt”. Similarly, the control group headline “Censorship, and Sentiments Towards People’s Rights While Considering the Privacy and Availability of Information” included the two neutral words “sentiments” and “availability.” The valence of the selected words was determined using the Affective Norms for English Words (ANEW) by Bradley and Lang.<sup>28</sup> Along with the headline, participants were presented with the same article on the topic of censorship, which was developed by generative AI using a prompt for a non-biased news article on censorship (Grammarly, 2024) and was 222 words in length.

To measure the dependent variables, four items were developed to measure the participants’ interest in the article, the likelihood that they would share the article, how informative they felt the article was, and how biased they felt the article was. Each item was measured on a 6-point scale, with 1 being (not very) and 6 being (very), and the item measuring perceived bias was reverse-coded. Participants responded to a manipulation check question asking them to rate the headline’s language valence on a scale of 1 (negative) to 9 (positive), along with standard demographic questions (i.e., age, race, gender). Finally, they answered the additional descriptive questions to better understand the population composition. The first question asked how often they pay attention to the news, either every day, weekly, monthly, or yearly. The second question asked where they get their news from—online news outlets, social media, television stations, or printed media. The third asked whether they identified their political ideology as conservative, moderate, or liberal. The last question asked how familiar they were with the topic of censorship on a scale of 1 (very unfamiliar) to 5 (very familiar).

#### *Procedure*

This research was approved by the institution’s IRB as an exempt study. Because some individuals attend college prior to the age of 18, minors who wished to participate in exempt studies are required to submit a signed parental consent form to the

department prior to participating in any exempt studies. Participants self-selected to participate in the online study using the school's research recruitment system and were initially presented with an informed consent document. After reading and indicating consent, participants were then randomly assigned to one of the two experimental groups or the control group. Regardless of their group, they were presented with instructions to read the following article carefully, with the phrase "starting with the title" underlined, bolded, and in a larger text. Depending on their assigned group, these instructions were followed by the headline corresponding to their assigned condition. All three groups were provided with the same AI-generated article on censorship. Because the reading rate of college students can range from 100-400 words per minute (Wake Forest University, n.d.), participants had to spend a minimum of three minutes reading the headline and article before moving on with the study to account for a potential slower rate of reading and distractions present in an online environment. They then responded to all survey items for the dependent variables, the manipulation check question, and other demographic questions. Participants were then debriefed on the experiment.

## STUDY 2

### *Participants – Original Sample*

For Study 2, although an effort was made to collect data in person to encourage participant attention, due to constraints on data collection only during course sessions, data from 85 participants were collected in person, while data from 58 participants were collected online. All 143 participants were recruited from the same university used for Study 1. Of the 143 participants recruited, 112 (78.3%) identified as female, 27 (18.9%) identified as male, and 4 (2.8%) identified as gender non-binary. Racial demographics included 63 (44.1%) African American/Black, 60 (42.0%) Caucasian/White, 13 (9.1%) Hispanic/Latinx, 5 (3.5%) Asian American, and 1 Native American (0.7%). The ages of participants reported were between 16 and 53 years ( $M = 20.41$ ,  $SD = 4.67$ ). The reported political ideology for participants included 35 (24.5%) conservatives, 75 (52.4%) moderates, and 33 liberals (23.1%). When asked how often they pay attention to the news, 19 (13.3%) said daily, 68 (47.6%) said weekly, 42 (29.4%) said monthly, 3 (2.1%) said yearly, and 11 (7.7%) said never. When asked where they get most of their news, 27 (18.9%) said online news outlets, 98 (68.5%) said social media, 12 (8.4%) said television stations, and 6 (4.2%) said other. When asked about their familiarity with the topic of censorship, 13 (9.1%) said they were very unfamiliar, 19 (13.3%) said they were unfamiliar, 58 (40.6%) said they were neither unfamiliar nor familiar, 39 (27.3%) said they were familiar, and 14 (9.8%) said they were very familiar.

### *Participants – Replication Sample*

To replicate the findings of Study 2, a secondary sample was obtained using Prolific from a sample of individuals who identified as college students in the U.S. All 191 participants completed the study online. Of the 191 participants recruited, 81 (42.4%) identified as female, 97 (50.8%) identified as male, 4 (2.1%) transgender male, 1 (.5%) transgender female, 7 (3.7%) identified as gender non-binary, and 1 (.5%) opted not to disclose. Racial demographics included 39 (20.4%) African American/Black, 105 (55%) Caucasian/White, 22 (11.5%) Hispanic/Latinx, 22 (11.5%) Asian American, and 3 indicated other (1.6%). The ages of participants reported were between 18 and 45 years ( $M = 28.50$ ,  $SD = 7.28$ ). The reported political ideology for participants included 31 (16.2%) conservatives, 55 (28.8%) moderates, and 105 liberals (55.0%). When asked how often they pay attention to the news, 72 (37.7%) said daily, 96 (50.3%) said weekly, 22 (11.5%) said monthly, and 1 (0.5%) said never. When asked where they get most of their news, 70 (36.6%) said online news outlets, 110 (57.6%) said social media, 9 (4.7%) said television stations, and 2 (1.0%) said other. When asked about their familiarity with the topic of censorship, 1 (.5%) said they were very unfamiliar, 10 (5.2%) said they were unfamiliar, 29 (15.2%) said they were neither unfamiliar nor familiar, 105 (55.0%) said they were familiar, and 46 (24.1%) said they were very familiar.

### *Materials*

Study 2 employed a 2x2 factorial design, manipulating both the valence of the headline (positive or negative) and the valence of the article (congruent or incongruent with headline). Participants were assigned to one of four experimental conditions: positive headline congruent with article valence, positive headline incongruent with article valence, negative headline congruent with article valence, and negative headline incongruent with article valence. Generative AI (Copilot, 2025) assisted in the development of two headlines with opposing valences: a positive valence (i.e., Censorship: Protecting Users from Harm) and a negative valence (i.e., Censorship: A Threat to Freedom Disguised as Security). The article headlines were revised from the original ones used in the first study to make them more appealing to the reader. The same AI tool was used to develop corresponding article text in both positive and negative valence, and was further edited by two researchers. Edits made involved making sure that differences in language valence occurred at the same relative locations in both articles and to similar degrees in order to ensure manipulation was consistent between the different articles.

To ensure successful valence manipulation, we validated the article text using Linguistic Inquiry and Word Count (LIWC; Tausczik & Pennebaker<sup>30</sup>). LIWC analysis revealed that the negative-valence article had a negative emotion score of 5.02, substantially higher than the typical average of 1.38 for formal written language. Similarly, the positive-valence article produced a positive emotion score of 4.11, compared to the typical average of 2.33 for formal language. To control for potential confounds

related to article length, both versions of the article, including their respective headlines, were matched for length, with each totaling 227 words. Furthermore, to address inattention issues related to headline valence exposure from Study 1, Study 2 included a manipulation check that involved presenting participants with the article headline on a separate page and asking them to type it into a response box. On the following page, they were again presented with the headline that was now paired with the article's text. Afterwards, participants were asked to rate the valence of the article headline on a Likert-type scale ranging from 1 to 9, with 1 being *negative* and 9 being *positive*.

Along with questions that gauged interest in the article content, one's political ideology, and standard demographics (i.e., age, race, gender) that were included in Study 1, Study 2 also included five additional questions to measure the dependent variable of reading comprehension. Each reading comprehension question was multiple choice and involved information that could be found in the provided text. A total score of the number of comprehension questions answered accurately was calculated for each participant.

### *Procedure*

This research was approved by the institution's IRB as an exempt study. Because some individuals attend college prior to the age of 18, minors who wished to participate in exempt studies are required to submit a signed parental consent form to the department prior to participating in any exempt studies. Participants for the original sample self-selected into the study using the school's research recruitment system. Participants that signed up for in-person appointment came to the lab at a participant-selected time to take part in the in-person study that was facilitated using Qualtrics software (Qualtrics, 2025). Each session typically consisted of no more than 5 participants. While others signed up for an online option and completed the study in an area of convenience. For the replication sample, participants self-selected into the study using Prolific, an international research recruitment platform. Participants who signed up on Prolific were immediately redirected to Qualtrics to complete the study online.

For all participants, an informed consent document was presented to each participant providing them with basic information about the study. After indicating their consent, each participant was randomly assigned to one of the four experimental conditions. To ensure participants were correctly exposed to the stimulus, they were presented with the headline for the group they were assigned to (i.e., positive or negative valence), as well as asked to type out this headline in a box below. The headline was then presented again on the following page along with the article text associated with their assigned group (i.e., congruent or incongruent). Participants could only proceed with the study after a minimum of two minutes. This time was reduced from Study 1 since the in-person nature of the study should have allowed for fewer distractions, and even at a slower rate of reading (i.e., 100 words/minute), only 2.27 minutes would be needed to sufficiently read the article. After reading the article, participants were asked to answer reading comprehension questions related to the article's content, followed by a question asking them to rate the headline's valence. Along with this, they were presented with questions to gauge the reader's interest in the article as well as questions to observe demographic data (i.e., age, gender, race, political ideology). Finally, participants were debriefed on the experiment.

## **RESULTS**

### *STUDY 1*

#### *Manipulation Check*

To have passed the manipulation check, those who received the negatively valenced headline had to indicate a rating between 1-3, those who received the neutral headline had to indicate a rating between 4-6, and those who received the positively valenced headline had to indicate a rating between 7-9. Only 47 of the participants (40.5% of the 116 participants) correctly identified the headline as the intended valence, and thus all others were excluded from subsequent analyses due to possible lack of attention.

Of these 47 participants, 41 (87.2%) identified as female, 5 (10.6%) identified as male, and 1 (2.1%) identified as gender non-binary. Racial demographics included 29 (61.7%) Caucasian/White, 15 (31.9%) African American/Black, 2 (4.3%) Asian American, and 1 (2.1%) Hispanic/Latin. The age of participants was between 16 and 45 ( $M = 19.59$ ,  $SD = 4.32$ ). Participant political ideology included 16 (34.0%) conservatives, 21 (44.7%) moderates, and 9 (19.2%) liberals. When asked how often they pay attention to the news, 7 (14.9%) said daily, 22 (46.8%) said weekly, 13 (27.7%) said monthly, and 5 (10.6%) said yearly. When asked where they get most of their news, 13 (27.7%) said online news outlets, 28 (59.6%) said social media, and 6 (12.8%) said television stations. When asked about their familiarity with the topic of censorship, 5 (10.6%) said they were very unfamiliar, 2 (4.3%) said they were unfamiliar, 16 (34.0%) said they were neither unfamiliar nor familiar, 18 (38.3%) said they were familiar, and 6 (12.8%) said they were very familiar.

An analysis of variance (ANOVA) was conducted to examine the relationship between the headline valence rating and the IV condition of participants. There was a significant difference in ratings of headline valence between the positive headline ( $M =$

5.53,  $SD = 1.18$ ), neutral headline ( $M = 5.00, SD = 1.19$ ), and negative headline ( $M = 4.71, SD = 1.51$ ),  $F(2,111) = 3.85, MSE = 1.69, p = .024, \eta_p^2 = .065$ . Pairwise comparison indicated that positive headlines were rated as more positive compared to negative headlines ( $p = .007$ ), while neither was different in rating compared to the neutral headline ( $p = .080, p = .333$ , respectively).

*Potential Confounds*

To account for a potential confounding variable and determine if covariates were needed for our analysis, we examined whether the participants' familiarity with censorship, political ideology, and frequency with which they attend to the news differed across assigned IV conditions while excluding those who failed the manipulation check. There was a marginally significant difference in familiarity with censorship between IV conditions,  $F(2,44) = 3.20, MSE = 1.13, p = .05, \eta_p^2 = .13$ . However, there was no significant difference across the IV conditions for political ideology [ $F(2,43) = .35, MSE = .55, p = .707, \eta_p^2 = .02$ ] or frequency with which they attend to the news [ $F(2,44) = .08, MSE = .78, p = .924, \eta_p^2 = .004$ ]. Therefore, given our limited sample size and that familiarity with censorship was only marginally significant, none of these variables were included as covariates in our analysis.

*Hypothesis Testing*

A between-group ANOVA was conducted with only the participants who passed the manipulation check to examine the impact of headline valence condition (i.e., positive, neutral, negative) on participant impressions (i.e., interest, shareability, bias, informative) of the article. Contrary to the research hypothesis, how interesting participants found the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2,44) = 1.66, MSE = 1.44, p = .202, \eta_p^2 = .07$ . Contrary to the research hypothesis, the likelihood that participants would share the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2,44) = .59, MSE = 1.40, p = .560, \eta_p^2 = .03$ . Contrary to the research hypothesis, how informative participants found the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2,44) = .63, MSE = 1.48, p = .538, \eta_p^2 = .03$ . Contrary to the research hypothesis, how biased participants felt the article was in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2,44) = .35, MSE = 1.83, p = .706, \eta_p^2 = .02$ . See **Table 1** for univariate statistics.

Variable	Positive Headline		Negative Headline		Neutral Headline	
	M	SD	M	SD	M	SD
Interest	3.25	1.67	3.71	1.11	2.84	1.08
Share	2.00	1.20	1.43	0.79	1.94	1.24
Informative	4.38	1.48	3.86	0.90	3.84	1.22
Bias	3.13	1.46	2.57	1.40	2.75	1.32

Note: M and SD represent mean and standard deviation, respectively

**Table 1.** Study 1: Impact of Headline Valence (positive, negative, neutral) on Participant Impressions

*STUDY 2: ORIGINAL SAMPLE RESULTS*

*Manipulation Check*

To have passed the manipulation check, those who received the negatively valenced headline had to indicate a rating between 1-3 and those who received the positively valenced headline had to indicate a rating between 7-9. We assumed in-person data collection and requesting participants to type the article title prior to reading the article would increase attention. However, only 37 participants (25.9% of the 143 participants) correctly identified the headline as the intended valence. Rates of accuracy were similar among those who participated in the study in person, with only 23 of the participants (27.1% of the 85 participants) correctly identified the headline as the intended valence. An ANOVA was conducted to examine the relationship between the headline valence rating and the IV condition of participants. Although participants were not rating the article headline within the anticipated range (i.e., 1-3 for negative valence, 7-9 for positive valence), there was a significant difference between the ratings of the two headlines, such that the negative headline ( $M = 4.44, SD = 1.62$ ) was rated as more negatively valenced than the positive headline ( $M = 5.29, SD = 1.23$ ),  $F(1,139) = 12.28, MSE = 2.07, p < .001, \eta_p^2 = .08$ . There was also a significant difference in ratings of headline valence when accounting for the article valence, such that the negative headline with congruent text ( $M = 4.15, SD = 1.58$ ) was rated as more negatively valenced than both the positive headline with congruent article text ( $M = 5.35, SD = 1.43; p < .001$ ) and the positive headline with incongruent article text ( $M = 5.22, SD = 1.02; p = .002$ ),  $F(3,137) = 4.97, MSE = 2.06, p = .003, \eta_p^2 = .10$ . There were no significant differences between the other groups. Because the inability to pass the manipulation check no longer appears to be due to inattention, considering the changes made between Study 1 and Study 2, the rating of the headline valence was included as a covariate in subsequent analyses instead of excluding those who failed the manipulation check.

*Potential Confounds*

To account for a potential confounding variable and determine if covariates were needed for our analysis, we examined whether the participants’ familiarity with censorship, political ideology, and the frequency with which they attend to the news differed across the assigned IV conditions. There was no significant difference across IV conditions for familiarity with censorship,  $[F(3,139) = .90, MSE = 1.15, p = .445, \eta_p^2 = .02]$ , political ideology  $[F(3,139) = .53, MSE = .48, p = .660, \eta_p^2 = .01]$  or frequency with which they attend to the news  $[F(3,139) = 1.38, MSE = 1.48, p = .253, \eta_p^2 = .03]$ . Therefore, none of these variables were included as covariates.

*Hypotheses on Impressions on Article Content (Study 1 Replication)*

To readdress the hypothesis of Study 1, a between-group ANCOVA (i.e., analysis of covariance) was conducted to examine the impact of headline valence condition (i.e., positive, negative) on participant impressions (i.e., interest, shareability, bias, informative) of the article, while controlling for participant ratings on headline valence. Contrary to the research hypothesis, how interesting participants found the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,138) = .01, MSE = 1.52, p = .931, \eta_p^2 < .001$ . Contrary to the research hypothesis, the likelihood that participants would share the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,138) = .28, MSE = 1.79, p = .595, \eta_p^2 = .002$ . Contrary to the research hypothesis, how informative participants found the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,138) = .06, MSE = 1.05, p = .804, \eta_p^2 < .001$ . Contrary to the research hypothesis, how biased participants felt the article was in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,138) = .02, MSE = 1.70, p = .904, \eta_p^2 < .001$ . See **Table 2** for univariate statistics.

Variable	Positive Headline		Negative Headline	
	M	SD	M	SD
Interest	3.19	1.31	3.08	1.18
Share	2.59	1.45	2.32	1.25
Informative	4.24	1.03	4.18	1.05
Bias	2.73	1.22	2.80	1.40

Note: M and SD represent mean and standard deviation, respectively

**Table 2.** Study 2: Impact of Headline Valence (positive, negative) on Participant Impressions

Additionally, because the changes to Study 2 procedures did not eliminate the issue of valence ratings matching anticipated valence, we reran the analyses for Study 1 to include the entire sample, while controlling for participant ratings on headline valence to directly compare the results to Study 2. How interesting participants found the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2,111) = .12, MSE = 1.62, p = .884, \eta_p^2 = .002$ . The likelihood that participants would share the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2, 111) = 1.31, MSE = 1.39, p = .273, \eta_p^2 = .02$ . How informative participants found the article in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2, 111) = .71, MSE = 1.59, p = .494, \eta_p^2 = .01$ . How biased participants felt the article was in the negative valence headline group was not significantly different from the neutral valence headline group or the positive valence headline group,  $F(2, 111) = .85, MSE = 1.56, p = .430, \eta_p^2 = .02$ . All findings were consistent with Study 2 results but were also contrary to hypotheses. Univariate statistics are presented in **Table 3**.

Variable	Positive Headline		Negative Headline		Neutral Headline	
	M	SD	M	SD	M	SD
Interest	2.95	1.21	3.05	1.45	2.95	1.12
Share	1.95	1.00	1.97	1.30	2.26	1.33
Informative	3.74	1.37	3.76	1.26	3.97	1.22
Bias	2.56	1.19	2.62	1.16	2.90	1.37

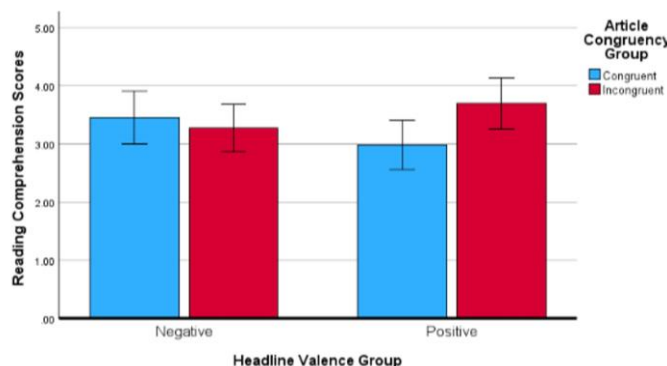
Note: M and SD represent mean and standard deviation, respectively

**Table 3.** Study 1: Impact of Headline Valence (positive, negative, neutral) on Participant Impressions including all participants

*Hypotheses on Reading Comprehension*

An 2x2 between-groups ANCOVA was conducted to examine the effects of headline valence and article valence on reading comprehension, while controlling for participants’ perceived headline valence, which was entered as a covariate. Contrary to what was hypothesized, the results indicated that the main effect of headline was not significant,  $F(1,135) = .02, MSE = 1.57, p = .902, \eta_p^2 < .001$ . There was also no significant main effect of article valence,  $F(1,135) = 1.57, MSE = 1.57, p = 0.211, \eta_p^2 = 0.01$ .

However, there was a significant interaction between headline valence and article valence,  $F(1,135) = 4.39$ ,  $MSE = 1.57$ ,  $p = .038$ ,  $\eta_p^2 = 0.03$ . Contrary to what was hypothesized, and as shown in **Figure 1**, follow-up simple effects tests revealed that when the headline was negative, there was no significant difference in reading comprehension ( $p = 0.555$ ), but that when the headline was positive, reading comprehension improved when article text was incongruent (i.e., negatively valenced) compared to when it was congruent [ $F(1,135) = 5.59$ ,  $MSE = 1.57$ ,  $p = 0.019$ ,  $\eta_p^2 = 0.04$ ].



**Figure 1.** Study 2: Participants' Mean Comprehension Score (y-axis) for Valence Congruency Groups (x-axis). Please note, error bars represent +/- 2 standard error of each condition mean. 'Congruent' signifies the headline valence matched the valence of the article while 'incongruent' signifies the headline valence did not match the valence of the article. Covariate in the model (i.e., participant rating of headline valence) are evaluated at the mean of 4.86.

#### Post-Hoc Analysis of Political Ideology

Although opinions on the article and political ideology did not vary across valence conditions, these variables may still be related to reading comprehension, such that only those with an interest in the article or of a particular political ideology had reading comprehension impacted by valence. Therefore, both were examined for their relationship with reading comprehension, controlling for participants' perceived headline valence. Bivariate correlations indicated that reading comprehension was positively associated with interest in the article [ $r(137) = .24$ ,  $p = .005$ ] and article informativeness [ $r(137) = .28$ ,  $p = .001$ ]. However, reading comprehension was not associated with likelihood to share the article [ $r(137) = -.01$ ,  $p = .886$ ] or ratings of article bias [ $r(137) = -.02$ ,  $p = .816$ ]. A between-groups ANCOVA revealed that political ideology was not associated with reading comprehension,  $F(2, 136) = 1.94$ ,  $MSE = 1.58$ ,  $p = .147$ ,  $\eta_p^2 = .03$ . Given our small sample size, the potential moderating effect of political ideology on the relationship between valence, congruency, and reading comprehension was not explored.

#### STUDY 2: REPLICATION SAMPLE RESULTS

##### Manipulation Check

An ANOVA was conducted to examine the relationship between the headline valence rating and the IV condition of participants. There was a significant difference between the ratings of the two headlines, such that the negative headline ( $M = 4.13$ ,  $SD = 1.49$ ) was rated as more negatively valenced than the positive headline ( $M = 5.01$ ,  $SD = 1.50$ ),  $F(1,188) = 16.64$ ,  $MSE = 2.23$ ,  $p < .001$ ,  $\eta_p^2 = .08$ . There was also a significant difference in ratings of headline valence when accounting for the article valence, such that the negative headline with congruent text ( $M = 3.96$ ,  $SD = 1.68$ ) was rated as more negatively valenced than both the positive headline with congruent article text ( $M = 5.14$ ,  $SD = 1.10$ ;  $p < .001$ ) and the positive headline with incongruent article text ( $M = 4.89$ ,  $SD = 1.82$ ;  $p = .003$ ),  $F(3,187) = 6.36$ ,  $MSE = 2.23$ ,  $p < .001$ ,  $\eta_p^2 = .09$ . There was also a significant difference such that negative headline with incongruent text ( $M = 4.30$ ,  $SD = 1.27$ ) was rating as more negatively valenced than the positive headline with congruent article text ( $M = 5.14$ ,  $SD = 1.10$ ;  $p = .006$ ). There were no other significant differences between the groups. Because the inability to pass the manipulation check no longer appears to be due to inattention, considering the changes made between Study 1 and Study 2, the rating of the headline valence was included as a covariate in subsequent analyses instead of excluding those who failed the manipulation check.

##### Potential Confounds

To account for a potential confounding variable and determine if covariates were needed for our analysis, we examined whether the participants' familiarity with censorship, political ideology, and the frequency with which they attend to the news differed across the assigned IV conditions. There was no significant difference across IV conditions for familiarity with censorship, [ $F(3,187) = .18$ ,  $MSE = .66$ ,  $p = .907$ ,  $\eta_p^2 = .003$ ], political ideology [ $F(3,187) = .47$ ,  $MSE = .57$ ,  $p = .703$ ,  $\eta_p^2 = .008$ ] or frequency with which they attend to the news [ $F(3,187) = 1.37$ ,  $MSE = .48$ ,  $p = .255$ ,  $\eta_p^2 = .02$ ]. Therefore, none of these variables were included as covariates.

*Hypotheses on Impressions on Article Content (Study 1 Replication)*

To readdress the hypothesis of Study 1, a between-group ANCOVA (i.e., analysis of covariance) was conducted to examine the impact of headline valence condition (i.e., positive, negative) on participant impressions (i.e., interest, shareability, bias, informative) of the article, while controlling for participant ratings on headline valence. Contrary to the research hypothesis, how interesting participants found the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,187) = 1.18, MSE = 1.41, p = .280, \eta_p^2 = .006$ . Contrary to the research hypothesis, the likelihood that participants would share the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,187) = .04, MSE = 2.04, p = .838, \eta_p^2 < .001$ . Contrary to the research hypothesis, how informative participants found the article in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,187) = .11, MSE = 1.03, p = .740, \eta_p^2 = .001$ . Contrary to the research hypothesis, how biased participants felt the article was in the negative valence headline group was not significantly different from the positive valence headline group,  $F(1,187) = .72, MSE = 1.35, p = .396, \eta_p^2 = .004$ . See **Table 4** for univariate statistics.

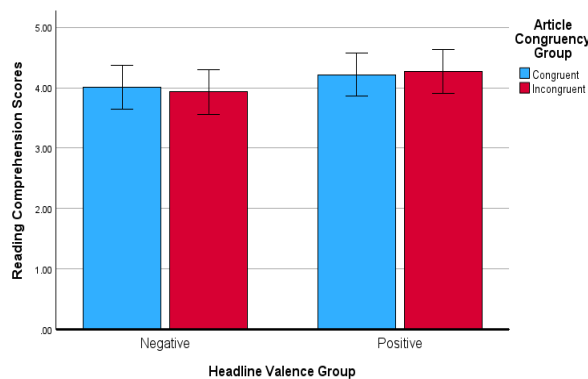
Variable	Positive Headline		Negative Headline	
	M	SD	M	SD
Interest	3.96	1.28	4.05	1.11
Share	3.16	1.38	3.07	1.50
Informative	4.43	1.05	4.37	1.01
Bias	2.71	1.22	2.77	1.21

Note: M and SD represent mean and standard deviation, respectively

**Table 4.** Prolific Sample: Impact of Headline Valence (positive, negative) on Participant Impressions

*Hypotheses on Reading Comprehension*

An 2x2 between-groups ANCOVA was conducted to examine the effects of headline valence and article valence on reading comprehension, while controlling for participants' perceived headline valence, which was entered as a covariate. Contrary to what was hypothesized, the results indicated that the main effect of headline was not significant,  $F(1,184) = 2.13, MSE = 1.52, p = .146, \eta_p^2 = .01$ . There was also no significant main effect of article valence,  $F(1,184) = .01, MSE = 1.52, p = .938, \eta_p^2 < .001$ . Lastly, contrary to what was hypothesized, there was no significant interaction between headline valence and article valence,  $F(1,184) = .13, MSE = 1.52, p = .719, \eta_p^2 = .001$ , see **Figure 2**.



**Figure 2.** Prolific Sample: Participants' Mean Comprehension Score for Valence Congruency Groups. Please note, error bars represent +/- 2 standard error of each condition mean. 'Congruent' signifies the headline valence matched the valence of the article while 'incongruent' signifies the headline valence did not match the valence of the article. Covariate in the model (i.e., participant rating of headline valence) are evaluated at the mean of 4.57.

*Post-Hoc Analysis of Political Ideology*

Although opinions on the article and political ideology did not vary across valence conditions, these variables may still be related to reading comprehension, such that only those with an interest in the article or of a particular political ideology had reading comprehension impacted by valence. Therefore, both were examined for their relationship with reading comprehension, controlling for participants' perceived headline valence. Bivariate correlations indicated that reading comprehension was positively associated with article informativeness [ $r(187) = .15, p = .043$ ]. However, reading comprehension was not associated with interest in the article [ $r(187) = .13, p = .068$ ], likelihood to share the article [ $r(187) = -.03, p = .676$ ] or ratings of article bias [ $r(187) = -.01, p = .918$ ]. A between-groups ANOVA revealed that political ideology was not associated with reading comprehension,  $F(2, 186) = 2.39, MSE = 1.49, p = .094, \eta_p^2 = .03$ . Given our small sample size, the potential moderating effect of political ideology on the relationship between valence, congruency, and reading comprehension was not explored.

## DISCUSSION

### STUDY 1

The goal of this study was to observe how headline valence impacted reader impressions on a neutral article. Results from this study indicated that there was no significant difference in audience impressions of an article between varying headline valence groups. This contradicts what was hypothesized as well as previous literature that indicated there to be a preference towards negatively valenced headlines<sup>12</sup>.

Previous research suggests a medium effect size (Cohen's  $F = 0.25$ ) would have been expected<sup>10, 12</sup> and that a sample size closer to 158 would have been needed to achieve power of 0.80 at  $p = 0.05$ . Therefore, a major limitation of this study was the high rate at which participants failed the manipulation check (i.e., incorrectly identifying the valence of the headline). Despite the headline being in a bold and larger font, 59.5% were inaccurate in their identification of the headline valence, and as such, were excluded from the study, which significantly reduced the study's power. Therefore, our sample size was likely not sufficient to test our hypothesis. Additional data was not collected for Study 1 due to the study occurring as part of a class project and was thus constrained by the semester timeline. Therefore, to improve the sample size in a follow-up project during the following semester, we sought to change the procedures to increase the saliency of the article headline in Study 2, thus hopefully eliminating the need to exclude a large portion of participants. As such, the procedure was adjusted to require participants to engage with the article's headline, and study participation occurred in person rather than online.

Additionally, the neutral condition was excluded in Study 2 to investigate the impact of congruency between the headline and article content on reading comprehension, given that "clickbait" headlines may be more extreme than the content itself to initially hook readers. Therefore, a comparison of negative to positive valence seemed warranted. Eliminating the neutral valence condition also allows us to maintain greater power while accommodating the addition of more conditions. In Study 1, participants were assigned to one of three conditions (i.e., positive, negative, neutral headline valence), while Study 2 included four conditions (i.e., positive congruent, positive incongruent, negative congruent, and negative incongruent). This focus on congruency was informed by prior findings indicating that when the headline and article are paired congruently, reading comprehension scores tend to be higher.<sup>5, 26</sup> A power analysis for Study 2, assuming a medium effect size (Cohen's  $F = .25$ ), indicates 179 is needed to test the interaction effect and 128 to test a bivariate relationship with power of .80 at  $p = .05$ .

### STUDY 2

Results of Study 2 examining reader impressions of the article were consistent with Study 1, showing no impact of headline valence on reader impressions. This finding contrasts with our hypothesis and previous literature indicating a preference towards negative headlines.<sup>12</sup> One possible explanation for the contradictory results is that, unlike Study 1, participants were not all exposed to a single neutral article. Instead, they were presented with either a positive or a negative article. Therefore, a negative stimulus was presented to participants even when the impact of a positive headline was meant to be observed (e.g., the positive incongruent group), possibly mitigating any differences in impressions resulting from positive language in the headline. Additionally, issues with the manipulation check continued in Study 2. Due to such a large number of participants rating the headlines within a more neutral valence range (i.e., 4-6), it cannot be determined if headline valence can truly be classified as positive and negative, even though the negative headline was reported by participants to be more negatively valenced than the positive headline. Attention to the headline was likely not the main reason for headline valence ratings, as participants in study 2 were asked to read their assigned headline and then type it into a text box. Pilot testing headlines may be useful for future research to ensure an ideal level of valence, rather than relying on expected valence of individual words selected from the Affective Norms for English Words.<sup>28</sup>

It is also possible that the topic of censorship was simply not good for this study on valence. Adults under 50 appear to be more moderate as a population in their opinions on censorship when compared to adults over 50.<sup>31</sup> The difference could possibly hint towards less investment in this topic in younger generations. It may be advantageous to use a more widely talked about and divisive topic. For example, the topic of abortion may serve as a stronger candidate for future research, as young people tend to be more opinionated towards one side of the argument than any other age group, which is contrary to censorship.<sup>32</sup>

Results from the original Study 2 sample, concerning the valence effect on reading comprehension, did not align with our hypotheses either. Findings indicated that headline valence alone did not impact reading comprehension, but the mismatch between headline and article valence did. This effect was most evident when a positive headline was paired with a negative article. The result that the group with positive headlines and incongruent valence had the highest reading comprehension scores was contrary to the expectation that the negative headline with congruent valence would lead to the best scores. However, these results do not eliminate the possibility of negativity bias, as participants in the positive headline incongruent group still encountered a negative article, despite the headline's positivity. Although these findings differ from previous research, it may be because earlier studies focused on the mismatch between factual and misleading headlines rather than word valence.<sup>5, 26</sup> In this

study, all groups received factual information, with the only difference being the language used to describe it. As mentioned earlier, there might be little interest in censorship, and instead of confusing participants with contradictory information, using conflicting language around the same core facts may actually increase attention. However, the replication of Study 2 with a larger sample size indicated there was no longer a significant effect of pairing a positive headline with a negative article. Initial Study 2 findings could have been a type I error as the sample size was underpowered. However, there were also differences between the demographics of the initial study and its replication that could have impacted the difference in findings. A slightly older average age, less familiarity with censorship, frequency that participants attended to the news, and a larger portion of liberal participants in the replication could have contributed to the different findings. More specifically, since the replication sample seemed to indicate less familiarity with censorship, it's possible that they lacked preconceived notions of censorship and therefore placed greater effort on attending to the news article content. Additionally, less preconceived notions could mean a lack of already existing biases that the headline and article valences could act on. Less familiarity with the topic and its relation to lower potential for preexisting biases could lower variability in participant investment as well as perception in the text.

#### GENERAL DISCUSSION

The present set of studies examined the effects of headline valence on reader impressions, behavioral intentions, and reading comprehension. Study 1 primarily tested how headlines influence reader impressions and behavioral intentions, while Study 2 sought to replicate and extend these findings to reading comprehension. Although neither study demonstrated a significant impact of headline valence on reader impressions, contrary to the hypotheses, we believe our findings have important implications for future research.

First, the consistent null effects across both studies may indicate that perceptions of online news and behavioral intentions with online media are highly individualized. For example, Janét *et al.*<sup>33</sup> found that headlines did not predict engagement scores (e.g., likelihood of sharing, reading, commenting on, or ignoring a news article). Instead, individuals' scientific curiosity and political views were stronger predictors. This implies that the negativity bias in media engagement may not apply to all types of content or could be influenced by other factors.

Past research suggests that emotionally charged stimuli capture attention more strongly than neutral stimuli.<sup>34</sup> This effect has also been demonstrated in neuroscience research, showing that emotional (both positive and negative) material elicits greater neural activation than neutral material, thereby enhancing cognitive processing and memory.<sup>35</sup> Although our theoretical framework was grounded in negativity bias, this broader literature highlights the possibility that emotional arousal, rather than valence-specific effects, may influence interest in news headlines. Thus, future research should directly examine whether general emotionality, independent of positive or negative valence, drives headline engagement.

Second, it is well established that exposure to a relevant headline before reading can activate background knowledge and support comprehension.<sup>8</sup> In Study 2, we initially observed higher comprehension when a positive headline was preceded by a negative article; however, this effect did not replicate in a secondary sample, suggesting that the relationship is not straightforward. These results partially align with Lagerwerf *et al.*,<sup>36</sup> who observed that incongruity between headlines and accompanying photos enhanced information processing. Despite methodological differences between their study and ours, both sets of findings suggest that moderate incongruity may induce cognitive dissonance that heightens attention and encourages deeper processing. Still, the interaction effect found with one sample may not be driven solely by emotional valence. Because the headlines in our study differed semantically (i.e., emphasizing freedom and security versus protection and harm), they may have activated different conceptual frames and knowledge schemas independent of emotional tone. These differences might have impacted understanding and congruency effects, potentially confounding efforts to isolate valence-specific influences. Future research should therefore more carefully control headline stimuli so that they vary primarily in emotional valence while holding other features constant, including conceptual content, emotional intensity, arousal level, and clickbait quality. Such efforts are especially important given that these findings also did not replicate within our secondary sample.

The current study had several strengths and limitations. The study uniquely contributed to previous literature by implementing word valence manipulation into the observation of headline and article congruency. In this area of study, previous literature had focused on contradictory information rather than on contradictory ways of presenting that information. Another strength was that the current study replicated the findings of the headline valence's impact on readers' opinions of the article, showing consistency between Study 1 and Study 2 in this respect.

However, it is important to note that some samples were underpowered, which may have contributed to the non-significant findings. Study 1 required 158 participants to achieve 80% power for a medium-sized effect. Due to time constraints during the semester, as this study was a class project, we were only able to recruit 116 participants. A power analysis for Study 2, for a medium effect size, required 179 participants to test the interaction effect and 128 to test a bivariate relationship with 80% power.

Although our replication sample included 191 participants, the original sample for Study 2 included only 143 participants, which is sufficient for the bivariate main effect analyses but not for testing the interaction effect. Given that the interaction effect was significant only with the smaller sample, those results should be interpreted with caution. The power issue, stemming from small sample sizes, was a limitation that may have produced null effects and prevented additional analyses from being conducted.

Another limitation was the lack of validated measures. Participants were specifically asked to indicate their political ideology rather than complete a questionnaire to better assess their affiliation. Due to time constraints and the unique aspects of this study, it was difficult to find a validated measure of reader interest in an article. Additionally, single-item dependent variable measures were created to capture factors indicative of interest, though they may not have accurately gauged the reader's interest as intended. Validity for single-item measures is typically more difficult and time-consuming to establish, as common methods used for multiple-item measures, such as confirming internal consistency, are not applicable. Similarly, the lack of multiple items may prevent a measure from capturing the full scope of more complex topics and lead to further inconsistency in the measurements. However, the simplicity of a single item can be beneficial, as it may be more straightforward for the participant when measuring a simple construct, provided the proper steps for validation are taken.<sup>38</sup> A multiple-item approach to address the lack of validated measures could be more beneficial, as it would allow for a more comprehensive measurement of reader interest. In future studies, this could be addressed by using a general reader interest measure with the specific article in the study.

## CONCLUSIONS

The findings have broader implications beyond those observed in this study. Current events and topics are constantly reported in the news and can significantly impact people's lives. How individuals perceive and understand certain information may be influenced by how it is presented to them. Misunderstandings caused by the way information is presented could lead people to support perspectives they might not have if they had a clearer understanding. However, results from this study do not suggest that these misunderstandings could arise from the use of language valence, as valence does not appear to shape opinions and perceptions on important issues. Although concerns exist about news sources' use of valenced language to increase engagement and profit, the likelihood that it affects readers' perceptions appears slim. However, there is still importance for trustworthy news outlets to be careful about how they might mislead readers. Still, efforts to mislead the audience may manifest in various forms. Therefore, readers should be aware of how they might be misled and consult a variety of sources to obtain the most accurate understanding of a topic.

## ACKNOWLEDGEMENTS

The authors thank the Office of Sponsored Awards and Research Support at USC Upstate for providing funding to be able to collect data on Prolific that allowed us to obtain a replication sample for Study 2.

## DATA AVAILABILITY

The materials and anonymized data for both studies are publicly available on the Open Science Framework (OSF): <https://osf.io/yb94x>

## REFERENCES

1. Pew Research Center, News platform fact sheet, <https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/> (accessed July 2025)
2. Scacco, J. M., and Muddiman, A. (2020) The curiosity effect: Information seeking in the contemporary news environment. *New Media & Society* 22, 429–448. <https://doi.org/10.1177/1461444819863408>
3. Ifantidou, E. (2009) Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics* 41, 699–720. <https://doi.org/10.1016/j.pragma.2008.10.016>
4. Scott, K. (2021) You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics* 175, 53–66. <https://doi.org/10.1016/j.pragma.2020.12.023>
5. Carcioppolo, N., Lun, D., and McFarlane, S. J. (2021) Exaggerated and questioning clickbait headlines and their influence on media learning. *Journal of Media Psychology: Theories, Methods, and Applications* 34, 30–41. <https://doi.org/10.1027/1864-1105/a000298>
6. Chuey, A., Luo, Y., and Markman, E. M. (2024) Epistemic language in news headlines shapes readers' perceptions of objectivity. *PNAS Proceedings of the National Academy of Sciences of the United States of America* 121. <https://doi.org/10.1073/pnas.2314091121>
7. León, J. A. (1997) The effects of headlines and summaries on news comprehension and recall. *Reading and Writing* 9, 85–106. <https://doi.org/10.1023/A:1007928221187>
8. Bransford, J. D., and Johnson, M. K. (1972) Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior* 11, 717–726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)

9. Surber, J. R., and Schroeder, M. (2007) Effect of prior domain knowledge and headings on processing of informative text. *Contemporary Educational Psychology* 32, 485-498. <https://doi.org/10.1016/j.cedpsych.2006.08.002>
10. Arfé, B., Delatorre, P., and Mason, L. (2022) Effects of negative emotional valence on readers' text processing and memory for text: An eye-tracking study. *Reading and Writing: An Interdisciplinary Journal*. <https://doi.org/10.1007/s11145-022-10362-7>
11. Zhang, M., Ding, S., Liu, G., Xu, Y., Fu, X., Zhang, W., and Xin, Z. (2021) Negativity bias in emergent online events: Occurrence and manifestation. *Acta Psychologica Sinica* 53, 1361–1375. <https://doi.org/10.3724/SP.J.1041.2021.01361>
12. Zhang, M., Wu, H., Huang, Y., Han, R., Fu, X., Yuan, Z., and Liang, S. (2024) Negative news headlines are more attractive: Negativity bias in online news reading and sharing. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. <https://doi.org/10.1007/s12144-024-06646-6>
13. Melumad, S., and Meyer, R. J. (2025) How listening versus reading alters consumers' interpretations of news. *Journal of Marketing Research* 62, 342-361. <https://doi.org/10.1177/00222437241280068>
14. Knickerbocker, H., Johnson, R. L., and Altarriba, J. (2015). Emotion effects during reading: Influence of an emotion target word on eye movements and processing. *Cognition and Emotion* 29, 784-806. <https://doi.org/10.1080/02699931.2014.938023>
15. Lei, A., Willems, R. M., and Eekhof, L. S. (2023) Emotions, fast and slow: processing of emotion words is affected by individual differences in need for affect and narrative absorption. *Cognition and Emotion* 37, 997-1005. <https://doi.org/10.1080/02699931.2023.2216445>
16. Pratto, F., and John, O. P. (1991) Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology* 61, 380–391.
17. Zsidó, Andres N. (2024) The effect of emotional arousal on visual attentional performance: A systematic review. *Psychological Research* 88, 1–24. <https://doi.org/10.1007/s00426-023-01852-6>
18. Kauschke, C., Bahn, D., Vesker, M., and Schwarzer, G. (2019) The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in Psychology* 10, 1654. <https://doi.org/10.3389/fpsyg.2019.01654>
19. Kousta, S. T., Vinson, D. P., and Vigliocco, G. (2009) Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* 112, 473-481. <https://doi.org/10.1016/j.cognition.2009.06.007>
20. Kensinger, E. A., Garoff-Eaton, R. J., and Schacter, D. L. (2007) Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language* 56, 575-591. <https://doi.org/10.1016/j.jml.2006.05.004>
21. Bahrack, L. E., Parker, J. F., Fivush, R., and Levitt, M. (1998) The effects of stress on young children's memory for a natural disaster. *Journal of Experimental Psychology: Applied* 4, 308–331. <https://doi.org/10.1037/1076-898X.4.4.308>
22. Goodman, G. S., Quas, J. A., and Ogle, C. M. (2010) Child maltreatment and memory. *Annual Review of Psychology* 61, 325-351. <https://doi.org/10.1146/annurev.psych.093008.100403>
23. Peterson, C., and Whalen, N. (2001) Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology* 15, S7-S24. <https://doi.org/10.1002/acp.832>
24. Mujica, C., and Bachmann, I. (2018) The impact of melodramatic news coverage on information recall and comprehension. *Journalism Studies* 19, 334-352. <https://doi.org/10.1080/1461670X.2016.1190661>
25. Bebbington, K., MacLeod, C., Ellison, T. M., and Fay, N. (2017) The sky is falling: evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior* 38, 92-101.
26. Ecker, U. K., Lewandowsky, S., Chang, E. P., and Pillai, R. (2014) The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology* 20, 323–335. <https://doi.org/10.1037/xap0000028>
27. Kubin, E., von Sikorski, C., and Gray, K. (2025) Political censorship feels acceptable when ideas seem harmful and false. *Political Psychology*, 46(2), 279-299. <https://doi.org/10.1111/pops.13011>
28. The Center for Research in Psychophysiology, University of Florida, Affective norms for English words (ANEW): Instruction manual and affective ratings,. <https://pdodds.w3.uvm.edu/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf> (Accessed October 2024)
29. Center for the Advancement of Teaching, Wake Forest University, Estimation Details, <https://cat.wfu.edu/resources/workload/estimationdetails/> (Accessed August 2025)
30. Tausczik, Y. R., and Pennebaker, J. W. (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 24-54. <https://doi.org/10.1177/0261927X09351676>
31. Pew Research Center, Most Americans favor restrictions on false information, violent content online, <https://www.pewresearch.org/short-reads/2023/07/20/most-americans-favor-restrictions-on-false-information-violent-content-online/> (Accessed August 2025)
32. Pew Research Center, Public Opinion on Abortion, <https://www.pewresearch.org/religion/fact-sheet/public-opinion-on-abortion/> (Accessed August 2025)
33. Janét, K., Richards, O., and Landrum, A. R. (2022) Headline format influences evaluation of, but not engagement with, environmental news. *Journalism Practice* 16, 35-55. <https://doi.org/10.1080/17512786.2020.1805794>
34. Yiend, J. (2010) The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion* 24, 3-47. <https://doi.org/10.1080/02699930903205698>

35. Inaba, M., Nomura, M., and Ohira, H. (2005) Neural evidence of effects of emotional valence on word recognition. *International Journal of Psychophysiology* 57, 165–173. <https://doi.org/10.1016/j.ijpsycho.2005.01.002>
36. Lagerwerf, L., Timmerman, C., and Bosschaert, A. (2016). Incongruity in news headings: Readers' choices and resulting cognitions. *Journalism Practice* 10, 782-804.
37. Allen, M. S., Iliescu, D., and Greiff, S. (2022) Single item measures in psychological science: A call to action. *European Journal of Psychological Assessment* 38, 1–5. <https://doi.org/10.1027/1015-5759/a000699>

#### **ABOUT THE STUDENT AUTHOR**

Kyle Smoak is a psychology major at USC Upstate. He plans to graduate in the Fall of 2026. Upon graduation, he hopes to pursue graduate studies in Industrial-Organizational Psychology.

#### **PRESS SUMMARY**

News headlines are designed to grab attention, but when they don't match the article's content, they can alter how readers understand and remember the story. We tested how positive, negative, and misleading headlines influence people's impressions and comprehension of news articles. Across two studies, we found that whether a headline sounded positive or negative did not significantly impact readers' impressions. However, when headlines did not match the article's tone, readers' comprehension was affected in one sample, suggesting that misleading or incongruent headlines may influence how people process news. These findings highlight the complexity of headline effects and the importance of careful headline framing to support an accurate understanding of news content.

# Physics-Informed Neural Network Frameworks for the Analysis of Engineering and Biological Dynamical Systems Governed by Ordinary Differential Equations

Andrew Particka<sup>\*a</sup>, Tyrus Whitman<sup>a</sup>, Christopher Diers<sup>b</sup>, Ian Griffitt<sup>a</sup>, Charuka Wickramasinghe<sup>c</sup>, & Pradeep Ranaweera<sup>d</sup>

<sup>a</sup> Department of Engineering, Siena Heights University, Adrian, MI

<sup>b</sup> Department of Computer Information Systems, Siena Heights University, Adrian, MI

<sup>c</sup> Department of Oncology, Wayne State University, Detroit, MI

<https://doi.org/10.33697/ajur.2026.175>

Students: [apartick@sienabeights.edu](mailto:apartick@sienabeights.edu)<sup>\*</sup>, [twhitman@sienabeights.edu](mailto:twhitman@sienabeights.edu), [cdiers@sienabeights.edu](mailto:cdiers@sienabeights.edu), [igriffi1@sienabeights.edu](mailto:igriffi1@sienabeights.edu)

Mentors: [gi6036@wayne.edu](mailto:gi6036@wayne.edu), [pranawee@sienabeights.edu](mailto:pranawee@sienabeights.edu)

## ABSTRACT

This study presents and validates the predictive capability of the Physics-Informed Neural Networks (PINNs) methodology for solving a variety of engineering and biological dynamical systems governed by ordinary differential equations (ODEs). While traditional numerical methods are effective for many ODEs, they often struggle to achieve convergence in problems involving high stiffness, shocks, irregular domains, singular perturbations, high dimensions, or boundary discontinuities. Alternatively, PINNs offer a powerful approach for handling challenging numerical scenarios. In this study, classical ODE problems are employed as controlled testbeds to systematically evaluate the accuracy, training efficiency, and generalization capability of the PINNs framework under controlled conditions. Although not a universal solution, PINNs can achieve superior results by embedding physical laws directly into the learning process. The existence and uniqueness properties of several benchmark problems are first analyzed, and the PINNs methodology is subsequently validated on each model system. The results demonstrate that, for complex problems to converge to correct solutions, the loss function components data loss, initial condition loss, and residual loss must be appropriately balanced through careful weighting. It was further established that systematic tuning of hyperparameters including network depth, layer width, activation functions, learning rate, optimization algorithms, weight initialization schemes, and collocation point sampling plays a crucial role in achieving accurate solutions. Additionally, embedding prior knowledge and imposing hard constraints on the network architecture, without losing the generality of the ODE system, significantly enhances the predictive capability of PINNs.

## KEYWORDS

Physics-Informed Neural Networks; Ordinary Differential Equations; DeepXDE; Dynamical Systems; Activation functions; Adams Algorithm; Picard–Lindelöf Theorem; Grönwall's Inequality

## INTRODUCTION

Analytical solutions to ordinary and partial differential equations (ODEs/PDEs) are often unattainable, making numerical, discretization based methods the standard alternative.<sup>1-3</sup> While established techniques like Euler's method and the Runge-Kutta method, for ODEs, or the finite element method and finite difference method for PDEs, have enabled the simulation of complex systems, yet they are hampered by significant limitations. Those include: the challenges of mesh generation, inefficiency with high-dimensional problems, and difficulties in resolving stiffness, complex boundary conditions, and singular perturbations.<sup>4-7</sup> This study considers several benchmark dynamical systems (the Lorenz system, Lotka-Volterra equation, coupled mass–spring system, and RLC circuit equation) that are usually modeled using ODEs that describe the time evolution of a quantity which needs to be solved effectively from noisy measurements or sparse data. To explore more about other interesting dynamical systems, the following references are recommended.<sup>8-12</sup>

Training deep neural networks for scientific applications is often hindered by a lack of large datasets. While purely data-driven models can fit existing data, they frequently produce physically implausible results and generalize poorly due to extrapolation and biases. A more robust solution is physics-informed learning, which integrates prior knowledge from physical laws to enhance model performance. PINNs have emerged as a leading method in this area, embedding differential equations directly into the learning process. By leveraging full physical knowledge and effectively adding constraints to the optimization algorithm PINNs can be trained with minimal to no labeled data to create accurate surrogate models, where the loss function quantifies the

discrepancy between the model's predictions, and the physical constraints. This study presents and validates PINNs approaches with problem specific strategies, optimizing the solution to dynamical systems governed by ODEs.

In the last eight years, PINNs have been used effectively in diverse applications,<sup>13-16</sup> such as compartmental brain modeling, engineering, and computer vision. To obtain the approximate solution to a differential equation via deep learning, a key step is to constrain the neural network to minimize the residual (of the differential equation). Compared to the traditional mesh-based methods, deep learning could offer a mesh free approach by taking advantage of the automatic differentiation,<sup>17</sup> and could break the curse of dimensionality.<sup>18</sup> Another attractive feature of PINNs is that it can be used to solve inverse problems with minimum change of the code for forward problems.<sup>19</sup> In addition, PINNs have been further extended to solve integro-differential equations, fractional differential equations,<sup>20</sup> and stochastic differential equations.<sup>21</sup>

This study, demonstrates several benchmark examples using the PINNs framework. The implementation of the PINNs framework is carried out with the DeepXDE library in Python: which serves as both an educational and research platform for computational science and engineering applications. DeepXDE<sup>22</sup> can handle multi-physics problems and complex geometries through the constructive solid geometry approach, eliminating the need for intricate and time-consuming geometric preprocessing. Moreover, it allows users to solve time-dependent and steady-state differential equations with ease by specifying appropriate initial conditions. The framework also provides flexible callback functionalities for monitoring and customizing the training process.

The remainder of this paper is structured as follows. *Section 2* presents the general existence and uniqueness theorem for nonlinear system of ODEs and validates the theorem for the benchmark problems discussed in the study. *Section 3* outlines the PINNs methodology employed to solve these ODEs. *Section 4* provides numerical examples that demonstrate the predictive capability of the PINNs approach. Finally, *Section 5* concludes the results and highlights potential directions for future research.

## METHODS AND PROCEDURES

Before addressing methods for solving differential equations: whether analytically, qualitatively, or numerically, it is essential to first examine the fundamental mathematical questions of existence and uniqueness. The first concern is whether a solution exists at all; without existence, the search for a solution is meaningless. The second concern is uniqueness: if multiple solutions arise from the same problem formulation, the equation loses its predictive power and may have limited applicability to physical systems. Since differential equations often admit infinitely many solutions, uniqueness can only be guaranteed by imposing appropriate initial or boundary conditions. Thus, as a preliminary step, a set of theorems and conditions is stated to provide the theoretical foundation for verifying the existence and uniqueness of solutions to the ODE systems presented in this study to explore the predictive power of the PINNs framework. The proof of the theorems can be found from.<sup>23</sup>

### Existence and Uniqueness

**Theorem 1** (Existence and Uniqueness Theorem for Linear Equations) Let  $p_1(t), \dots, p_{n-1}(t)$  and  $g(t)$  are continuous on an interval  $(a, b)$  containing the point  $t_0$ . Then for every choice of initial values  $y_0, y_1, \dots, y_{n-1}$  there exists a unique solution on the whole interval  $(a, b)$  to the initial value problem.

$$y^{(n)}(t) + p_1(t)y^{(n-1)}(t) + \dots + p_n(t)y(t) = g(t);$$

$$y(t_0) = y_0, y'(t_0) = y_1, \dots, y^{(n-1)}(t_0) = y_{n-1}.$$

**Theorem 2** (Picard–Lindelöf Existence and Uniqueness Theorem for Systems) If  $\mathbf{f}$  and  $\frac{\partial \mathbf{f}}{\partial x_i}$ ,  $i = 1, 2, \dots, n$ , are continuous function in a ‘rectangle  $R = \{(t, x_1, \dots, x_n): a < t < b, c_i < x_i < d_i, i = 1, \dots, n\}$  that contains the point  $(t_0, \mathbf{x}_0)$ , then the initial values problem  $\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x})$ ,  $\mathbf{x}(t_0) = \mathbf{x}_0$ , where  $\mathbf{x}_0 = \text{col}(x_{1,0}, \dots, x_{n,0})$ , has a unique solution in some interval  $[t_0 - h, t_0 + h]$ ,  $h$  is a positive constant.

**Theorem 3** (Continuation of Solution) Let  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{f}(t, \mathbf{x})$  denote the vector function  $\mathbf{f}(t, \mathbf{x}) = (f_1(t, x_1, \dots, x_n), \dots, f_n(t, x_1, \dots, x_n))$ . Suppose  $\mathbf{f}$  and  $\frac{\partial \mathbf{f}}{\partial x_i}$ ,  $i = 1, 2, \dots, n$ , are continuous on the strip  $R = \{(t, \mathbf{x}): a \leq t \leq b, \mathbf{x} \text{ arbitrary}\}$  containing the point  $(t_0, \mathbf{x}_0)$ . Assume further that there exist a positive constant  $L$  such that for  $i = 1, 2, \dots, n$ ,  $\left| \frac{\partial \mathbf{f}}{\partial x_i} \right| \leq L$  for all  $t, \mathbf{x}$  in  $R$ . Then the initial value problem  $\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x})$ ,  $\mathbf{x}(t_0) = \mathbf{x}_0$ , has a unique solution on the entire interval  $a \leq t \leq b$ .

**Lemma 1** (Grönwall's inequality) Let  $I$  denote an interval of the real line of the form  $[a, b]$  with  $a < b$ . Let  $\beta$  and  $u$  be real valued continuous functions defined on  $I$ . If  $u$  is differentiable in the interior of  $I$  and satisfies the differential inequality  $u'(t) \leq \beta(t)u(t)$ ,  $t \in (a, b)$  then  $u$  is bounded by the solution of the corresponding differential equation  $v'(t) = \beta(t)v(t)$ :  $u(t) \leq u(a) \exp\left(\int_a^t \beta(s)ds\right)$  for all,  $t \in I$ .

*The Lorenz System*

The Lorenz system is a set of three ordinary differential equations, first developed by the meteorologist Edward Lorenz while studying atmospheric convection. It is a classic example of a system that can exhibit chaotic behavior, meaning its output can be highly sensitive to small changes in its starting conditions. For certain values of its parameters, the system's solutions form a complex, looping pattern known as the Lorenz attractor as shown in **Figure 1**. The model describes how three key properties of this system change over time as follows:

$$\begin{aligned} \frac{dx}{dt} &= \delta(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \quad t \in [a, b]$$

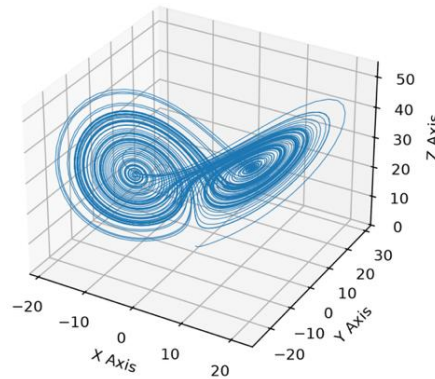


Figure 1. The Lorenz attractor.

with initial conditions  $x(0) = x_0$ ,  $y(0) = y_0$ ,  $z(0) = z_0$ , where,  $x$  is proportional to the intensity of the convection (the rate of fluid flow),  $y$  is proportional to the temperature difference between the rising and falling air currents and  $z$  is proportional to the distortion of the vertical temperature profile from a linear one. To verify the existence and the uniqueness of the solution, let

$$f(t, x) = \begin{bmatrix} \delta(y - x) \\ x(\rho - z) - y \\ xy - \beta z \end{bmatrix} \text{ and thus } \frac{\partial f}{\partial x} = \begin{bmatrix} -\delta & \delta & 0 \\ \delta - z & -1 & -x \\ y & x & -\beta \end{bmatrix}. \text{ Each component of } f(t, x) \text{ is a polynomial in } (x, y, z) \text{ and thus continuous as}$$

well as all the entries of  $\frac{\partial f}{\partial x}$  are continuous. Then by the Picard–Lindelöf theorem:  $f$  is locally Lipschitz, hence for any initial condition  $x(t_0) = x_0$  there exists a unique local solution on some interval  $t \in [t_0 - \epsilon, t_0 + \epsilon]$ . To extend the solution globally, define the energy-like function  $V(x, y, z) = x^2 + y^2 + (z - \rho)^2 \geq 0$ . Differentiating along solutions:  $\dot{V} = 2x\dot{x} + 2y\dot{y} + 2(z - \rho)\dot{z}$ . Then we substitute the Lorenz equations to  $\dot{V}$  and obtain:

$$\dot{V} = 2x\delta(y - x) + 2y(x(\rho - z) - y) + 2(z - \rho)(xy - \beta z)$$

After expansion and applying standard inequalities ( $2ab \leq a^2 + b^2$ ) and completing the square for the  $Z$ -term:

$$\dot{V} \leq KV + C_0$$

where  $K > 0$  and  $C_0 > 0$  are constants depending on  $\sigma, \rho, \beta$ . Then by Grönwall's inequality, it follows that,  $V(x, y, z)$  and therefore  $(x(t), y(t), z(t))$ , remains bounded for all  $t \geq t_0$ .

*Lotka -Volterra Equation*

The Lotka–Volterra equations, also known as the Lotka–Volterra predator–prey model, are a pair of first-order nonlinear differential equations, frequently used to describe the dynamics of biological systems in which two species interact, one as a predator and the other as prey. The populations change through time according to the pair of equations:

$$\frac{dx}{dt} = \alpha x - \beta xy, \quad \frac{dy}{dt} = -\gamma y + \delta xy; \text{ with initial conditions } x(0) = x_0, y(0) = y_0.$$

To verify the existence and the uniqueness of the solution, let  $f(t, x) = \begin{bmatrix} \alpha x - \beta xy \\ -\gamma y + \delta xy \end{bmatrix}$  and thus  $\frac{\partial f}{\partial x} = \begin{bmatrix} \alpha - \beta y & -\beta x \\ -\delta y & -\gamma + \delta x \end{bmatrix}$ . Each component of  $f(t, x)$  is a polynomial in  $(x, y, z)$  and thus continuous as well as all the entries of  $\frac{\partial f}{\partial x}$  are continuous. Then by the

Picard–Lindelöf theorem:  $f$  is locally Lipschitz, hence for any initial condition  $\mathbf{x}(t_0) = \mathbf{x}_0$  there exists a unique local solution on some interval  $t \in [t_0 - \epsilon, t_0 + \epsilon]$ .

*Coupled Mass - Spring System*

A coupled mass–spring system is considered, consisting of two masses connected by springs, where the motion of each mass affects the other. The governing equation for the mass–spring system can be written as:

$$[inertia]x'' + [damping]x' + [stiffness]x = F_{external}$$

The variables  $x$  and  $y$  typically represent the displacements of two masses from their equilibrium positions. The primes ( $x'', y''$ ) represent acceleration of mass 1, and mass 2 respectively. **Figure 2** shows on a smooth horizontal surface, a mass is attached to a fixed wall by a spring, and another mass is attached to the first object by a spring. The objects are aligned horizontally so that the springs are their natural lengths. In this study,  $F_{external} = 0$  is set, and the coupled mass–spring system is defined by the following system of second-order linear differential equations:

$$m_1x'' + (k_1 + k_2)x - k_2y = 0 \qquad m_2y'' + (k_2y - k_2)x = 0$$

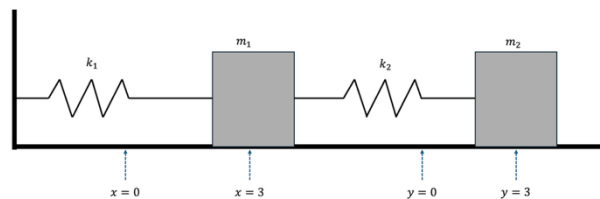
For simplicity set,  $m_1 = 2, m_2 = 1, k_1 = 1,$  and  $k_2 = 1$  with  $x(0) = 3, x'(0) = 0, y(0) = 3,$  and  $y'(0) = 0$  to discuss the existence uniqueness. To this end, first convert Mass–Spring System into a system of first order differential equation using the transformations  $u_1 = x, u_2 = x', u_3 = y$  and  $u_4 = y'$ . Then obtain the order reduction system as follows:

$$u' = f(t, u) = \begin{bmatrix} u_2 \\ u_3 - 3u_1 \\ u_4 \\ 2u_1 - 2u_3 \end{bmatrix}, \qquad u(0) = u_0 = \begin{bmatrix} 3 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

Since  $f(t, u)$  is linear in  $u$  and independent of  $t, f(t, u)$  is globally Lipschitz. Therefore, by Picard–Lindelöf theorem there exists a unique global solution for all  $t$ . Additionally, consider the linear system in matrix form:

$$u' = Au, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & -2 & 0 \end{bmatrix}$$

then the eigen values of  $A$  are  $\lambda_1 = +2i, \lambda_2 = -2i, \lambda_3 = +i$  and  $\lambda_4 = -i$ . Since all eigenvalues are purely imaginary, the solution is oscillatory, which will be demonstrated in *Section 3*.



**Figure 2.** The coupled mass–spring system.

*RLC Circuit Equation*

An RLC circuit is an electrical circuit consisting of a resistor (R), an inductor (L), and a capacitor (C), connected in series or in parallel. The equation governing the voltage  $v(t)$  across the capacitor were given as follows:

$$C \frac{d^2v}{dt^2} + \frac{1}{R} \frac{dv}{dt} + \frac{1}{L} v = f(t) \quad v(0) = v_0 \quad v'(0) = v'_0$$

For constants,  $C, R (\neq 0),$  and  $L (\neq 0)$  with a continuous function  $f(t)$  RLC circuit equation has a unique solution by **Theorem 1**. If  $f(t) = 0$  for  $t > 0$  RLC circuit is in free oscillation meaning how the circuit behaves after it’s energized and then left to oscillate or decay with no external forcing function. The oscillation is called underdamped if  $R > \sqrt{L/4C},$  overdamped if  $R < \sqrt{L/4C},$  and critically damped if  $R = \sqrt{L/4C};$  an example for underdamping is demonstrated in *Section 3*. **Figure 3** exhibits a parallel RLC circuit.

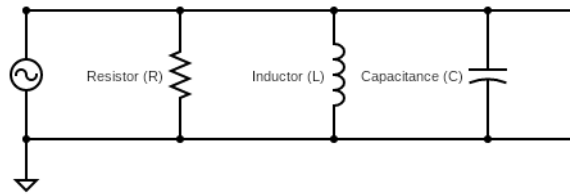


Figure 3. A parallel RLC circuit.

Physics Informed Neural Networks (PINNs)

In this section, a brief overview of deep neural networks is provided, along with the procedure of PINNs for solving differential equations. Mathematically, a deep neural network is a particular choice of compositional function. The simplest neural network is the feed forward neural network (FNN), also called multilayer perceptron (MLP), which applies linear and nonlinear transformations to the inputs recursively. Although many different types of neural networks have been developed in the past decades, this paper consider FNN, which is sufficient for most differential equation problems. Let  $\mathcal{N}^L(x): \mathbb{R}^{d_{in}} \rightarrow \mathbb{R}^{d_{out}}$  be a  $L$ -layer neural network, with  $N_l$  neurons in the  $l$ -th layer. The weight matrix and bias vector in the  $l$ -th layer is denoted by  $\mathbf{W}^l \in \mathbb{R}^{N_l \times N_{l-1}}$  and  $\mathbf{b}^l \in \mathbb{R}^{N_l}$ , respectively. Given a nonlinear activation function  $\sigma$ , which is applied element-wisely, the FNN is recursively defined as follows:

- Input layer:  $\mathcal{N}^0(x) = x \in \mathbb{R}^{d_{in}}$ ;
- Hidden layers:  $\mathcal{N}^l(x) = \sigma(\mathbf{W}^l \mathcal{N}^{l-1}(x) + \mathbf{b}^l) \in \mathbb{R}^{N_l}$ , for  $1 \leq l \leq L - 1$ ;
- Output layer:  $\mathcal{N}^L(x) = \mathbf{W}^L \mathcal{N}^{L-1}(x) + \mathbf{b}^L \in \mathbb{R}^{d_{out}}$

Commonly used activation functions including the sigmoid ( $1/(1 + e^{-x})$ ), the hyperbolic tangent (tanh), sine ( $\sin(x)$ ), and the rectified linear unit (ReLU,  $\max(x, 0)$ ) are shown in **Figure 4**. To measure the discrepancy between the neural network solution  $\hat{u}$  and the constraints, consider the loss function defined as the weighted summation of the  $L_2$  norm of residuals for the equation and boundary/initial conditions and the experimental data. The composite PINNs loss function was defined as:

$$\begin{aligned} \mathcal{L}(\theta) &= \lambda_{data} \mathcal{L}(\theta)_{data} + \lambda_{ode} \mathcal{L}(\theta)_{ode} + \lambda_{ic} \mathcal{L}(\theta)_{ic}; \text{ where,} \\ \mathcal{L}(\theta)_{data} &= \sum_{j=1}^m \lambda_{data}^j \mathcal{L}_{data}^j = \sum_{j=1}^m \lambda_{data}^j \left[ \frac{1}{n} \sum_{i=1}^n \left( x_{obs}^j(t_i) - x_{pred}^j(t_i; \theta) \right)^2 \right], \\ \mathcal{L}(\theta)_{ode} &= \sum_{j=1}^s \lambda_{ode}^j \mathcal{L}_{ode}^j = \sum_{j=1}^s \lambda_{data}^j \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{dx_j}{dt} \Big|_{t_i} - f_j(x_j(t_i)) \right)^2 \right], \\ \mathcal{L}(\theta)_{ic} &= \sum_{j=1}^s \lambda_{ic}^j \mathcal{L}_{ic}^j = \sum_{j=1}^s \lambda_{ic}^j \left[ \frac{1}{s} \left( x_{obs}^j(t_0) - x_{pred}^j(t_0; \theta) \right)^2 \right]. \end{aligned}$$

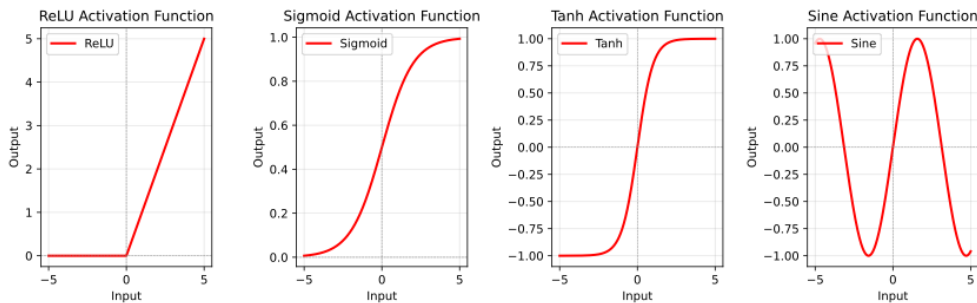


Figure 4. Commonly used activation functions.

The loss function in PINNs is designed to enforce both data fidelity and physical consistency by combining multiple objective terms. The data loss ( $\mathcal{L}_{data}$ ) ensures that the neural network’s predictions align with observed measurements, minimizing discrepancies at known data points. Data loss is calculated by the sum of the square differences of the predicted concentrations and the observed data. The ODE loss ( $\mathcal{L}_{ODE}$ ), embedding the underlying dynamics directly into the learning process makes sure to fulfill the law described by ODEs. The initial condition loss ( $\mathcal{L}_{IC}$ ), guarantees that the solution adheres to prescribed initial

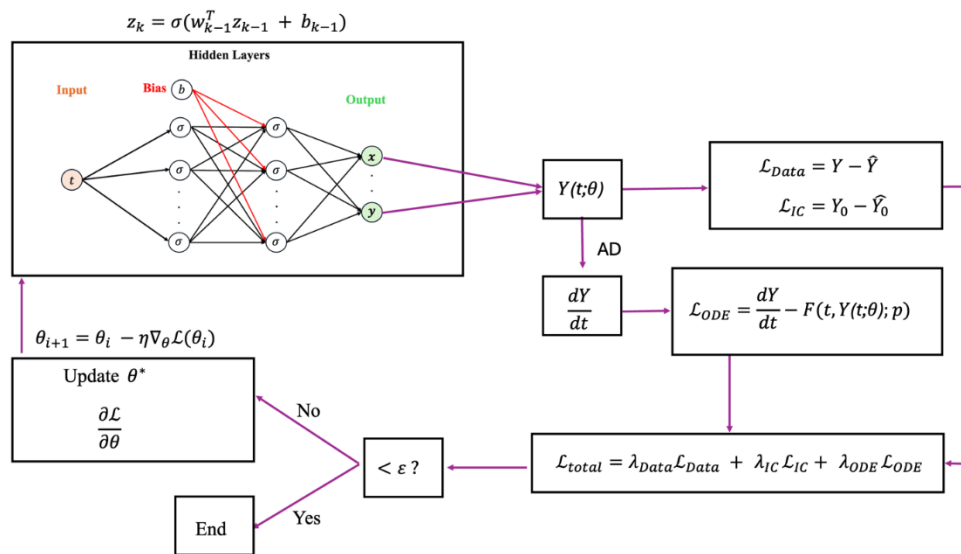
constraints. weights  $\lambda_{data}$ ,  $\lambda_{ODE}$ ,  $\lambda_{IC}$  balance contributions from data fidelity, ODE residuals, and initial conditions. The derivatives  $\frac{d\hat{x}_i}{dt}|_{t_i}$  are analytically calculated with automatic differentiation. By optimizing the composite loss function ( $\mathcal{L}(t; \theta)$ ),

$$\theta^* = \operatorname{argmin}[\mathcal{L}(t; \theta)],$$

the network not only interpolates sparse data but also generalizes as a physics-compliant surrogate model that is robust in regions where measurements are unavailable. **Figure 5** illustrates the PINN framework, where time (t) serves as input and outputs are optimized to minimize total loss. During the minimization we infer the neural network parameters  $\theta$  (weights and bias) via gradient-based optimizers, such as Adam and L-BFGS.<sup>24,25</sup> The convergence and accuracy of the PINNs are critically dependent on the selection of its architecture (i.e. number of layers and the neurons), optimizer (Adam or L-BFGS), learning rate ( $\eta$ ), weights ( $w$ ) and bias ( $b$ ) initializers (Glorot normal, Glorot uniform), number of optimization iterations, type of activation function (such as Tanh, Sigmoid, ReLU, Swish, Sine, etc.), and loss weights ( $\lambda_{data}$ ,  $\lambda_{ode}$ ,  $\lambda_{ic}$ ). In this study, the loss weight coefficients are manually selected according to the problem of interest, ensuring that the weighted losses remain of the same order of magnitude during network training. An open issue is that PINNs can produce negative values, even for strictly non-negative quantities. To handle this issue, a transformation is defined to ensure that the output is positive during training. If the nature of the solution curves is known, this information is incorporated into the network by adding a feature layer, enhancing accuracy and producing biologically meaningful solutions. The physics-informed neural network procedure for solving ODEs through the DeepXDE library is presented in the following.

*PINNs procedure for solving differential equations through DeepXDE*

- Step 1 Define the computational domain using the **geometry** module.
- Step 2 Define the differential equation.
- Step 3 Define the boundary and initial conditions.
- Step 4 Combine geometry, differential equations, and boundary/initial conditions together into **data.PDE**.
- Step 5 Specify the network architecture through layers, neurons, activation functions, inputs, outputs and weight initializers into **nn.FNN**.
- Step 6 Combine **data.PDE** and the **nn.FNN** into **dde.Model**.
- Step 7 Call **Model.compile** to set the optimization hyperparameters, such as optimizer and learning rate. The weights can be set here by loss weights.
- Step 8 Call **Model.train** to train the network where the number of iterations can be specified.
- Step 9 Call **Model.predict** to predict the solution at user defined, time points.



**Figure 5.** The physics-informed neural network model starts with time as inputs and then the outputs go to an optimization block where it minimizes the total loss  $\mathcal{L}_{Total}$  by optimizing the parameters  $\theta$  via gradient-based optimizers.

**RESULTS AND DISCUSSION**

In this section, several numerical examples are presented to evaluate the predictive capabilities of PINNs. In Example 1, the influence of network architecture, activation functions, and loss weights on the solution is investigated. *Example 2* illustrates the

ability of PINNs to approximate solutions when additional features are incorporated. *Example 3* is designed to assess prediction accuracy in the presence of noisy data. Finally, *Example 4* demonstrates the performance of PINNs in scenarios without experimental data. The total loss is computed as described in *Theorem 2*. For a function  $u(x)$  (the true solution) and its PINN approximation  $\hat{u}(x)$ , the relative  $L_2$  error is calculated at  $N$  discrete points  $x_i$  as follows:

$$L_2 \text{ error} = \frac{\sqrt{\sum_{i=1}^N (\hat{u}(x_i) - u(x_i))^2}}{\sqrt{\sum_{i=1}^N (u(x_i))^2}}$$

**Example 1.  $L_2$  Error and Computational Efficiency**

In this example, the performance of PINNs is evaluated under varying network configurations, including different numbers of layers, neurons, and activation functions. The significance of appropriate weighting through components of the loss function is further emphasized. For each simulation, the  $L_2$  error and computational time are recorded. To this end, the RLC circuit is considered and is given by:

$$C \frac{d^2v}{dt} + \frac{1}{R} \frac{dv}{dt} + \frac{1}{L} v = f(t) \quad v(0) = v_0 \quad v'(0) = v'_0$$

where,  $R = 20000$ ,  $L = 8$ ,  $C = 0.125 e^{-6}$ , and  $f(t) = 0$ . **Table 1** provides a detailed summary of the simulation results. All simulations presented in **Table 1** were conducted using the Adam optimizer with a learning rate of 0.0001. The network weights were initialized using the Glorot normal initializer, and each model was trained for 100,000 epochs. Due to the singularly perturbed nature of the *RLC Circuit Equation*, the PINNs failed to accurately capture the true solution when trained with arbitrary hyperparameters. Therefore, a systematic analysis was performed, as summarized in **Table 1**. The results indicate that the PINNs successfully captured the true solution when configured with 3 layers, 25 neurons per layer, sine activation function, and loss weights of  $[10^{-7}, 10^{+3}, 10^0, 10^0]$ . The  $L_2$  error and the corresponding computational time (in seconds) are summarized in **Table 1**. The minimum  $L_2$  error achieved was 0.0012, obtained after 114.7 seconds of training, which corresponds to the entry *\*0.0012 (114.7)* in **Table 1**. After determining the optimal configuration for obtaining an acceptable solution based on **Table 1**, we conducted simulations under identical settings while varying the number of training iterations. The corresponding results are summarized in **Table 2**. As shown, the accuracy of the predicted solution improves with an increased number of iterations; however, this improvement comes at the expense of significantly higher computational cost. **Figure 6** illustrates the network architecture, the comparison between the predicted and reference solutions, the individual components of the loss function, and the evolution of the  $L_2$  error after one million iterations.

L	AF	Number of Neurons				
		Loss Weights	N=5	N=25	N=75	N=100
1	Relu	$[10^0, 10^0, 10^0, 10^0]$	1.0(42.8)	1.0(42.8)	1.0(43.3)	1.0(48.7)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9999(42.8)	0.9999(43.5)	0.9999(47.3)	0.9999(48.7)
	Tanh	$[10^0, 10^0, 10^0, 10^0]$	1.0(46.3)	1.0(61.8)	1.0(61.8)	1.0(69.3)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	1.0295(48.3)	1.1669(61.7)	1.6172(69)	1.875(72.5)
	Sigmoid	$[10^0, 10^0, 10^0, 10^0]$	1.0(46.5)	1.0(48.5)	1.0(53.9)	1.0(56.3)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	1.0069(47.6)	1.0307(48.1)	1.1188(53.5)	1.1671(56.3)
Sine	$[10^0, 10^0, 10^0, 10^0]$	1.0(46.5)	1.0(65.7)	1.0(78.8)	1.0(89.8)	
	$[10^{-7}, 10^{+3}, 10^0, 10^0]$	1.0274(48)	1.1679(67.4)	1.6197(82.4)	1.8791(89.2)	
3	Relu	$[10^0, 10^0, 10^0, 10^0]$	1.0(64.6)	1.0(71.9)	1.0(106)	0.9999(120.8)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9999(64.1)	0.9999(70.8)	0.9999(101.2)	0.9999(117.3)
	Tanh	$[10^0, 10^0, 10^0, 10^0]$	1.0(75.8)	1.0(109)	0.9999(137.3)	0.9999(137.3)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9974(75.9)	0.9964(104)	0.9675(145)	0.9008(153.3)
	Sigmoid	$[10^0, 10^0, 10^0, 10^0]$	1.0(80)	1.0(86.9)	1.0(150.6)	1.0(187.2)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	1.0336(78.3)	1.1947(86.4)	1.4457(144.4)	1.374(189.1)
Sine	$[10^0, 10^0, 10^0, 10^0]$	1.0(77.5)	1.0(122.4)	1.0(176.7)	1.0(288.9)	
	$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.5275(79.8)	<i>*0.0012(114.7)</i>	0.0128(172.1)	0.0053(184.0)	
9	Relu	$[10^0, 10^0, 10^0, 10^0]$	1.0(128.5)	1.0(142.1)	1.0(256.9)	1.0(290.2)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9999(129.7)	0.9999(147.6)	0.9999(247.5)	0.9999(287.7)
	Tanh	$[10^0, 10^0, 10^0, 10^0]$	0.9442(159.8)	0.8923(241)	0.9992(375.8)	1.0(817.3)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9936(158.5)	0.9888(233.1)	0.9775(350.6)	0.9841(600.1)
	Sigmoid	$[10^0, 10^0, 10^0, 10^0]$	1.0(165)	1.0(196.1)	1.0(402.7)	1.0(534.4)
		$[10^{-7}, 10^{+3}, 10^0, 10^0]$	0.9939(164.4)	0.9999(194.3)	1.0(381.9)	1.0001(533.8)
Sine	$[10^0, 10^0, 10^0, 10^0]$	0.9998(2422.2)	0.7753(260.5)	1.0(470.8)	1.0(519)	
	$[10^{-7}, 10^{+3}, 10^0, 10^0]$	1.0013(8298.2)	0.7641(4177.7)	0.3879(481.6)	0.6166(525.7)	

**Table 1.** A systematic model simulation approach to capture the true solution. Abbreviations: L the number of layers; N the number of neurons; AF activation function.

Number of iterations	5000	10000	100000	300000	1000000
$L_2$ error	1.12	1.23	0.0012	8.44e-04	6.19e-04
Computational time (s)	5.17	10.05	114.7	338.05	1178.00

Table 2.  $L_2$ error and the computational time over iterations.

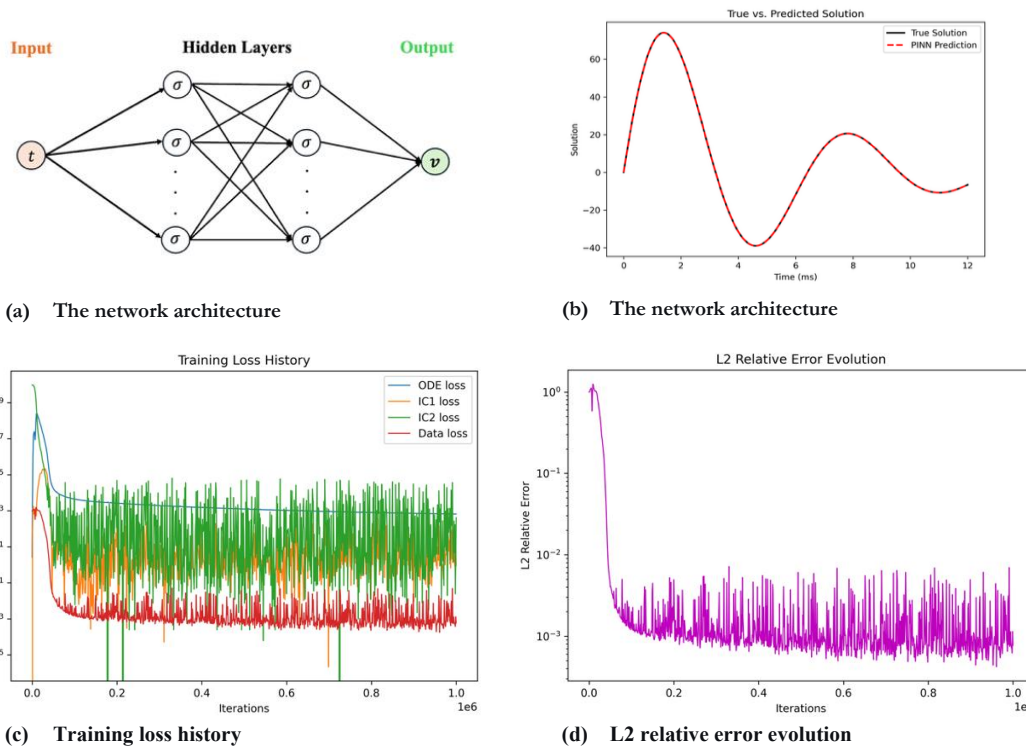


Figure 6. Graphical illustration of the network, solution and the loss and error histories of RLC Circuit Equation.

**Example 2.** Enhance learning by training with additional features

The main objective of this example is to demonstrate how incorporating feature layers can enhance the performance of PINNs. Using the Lotka–Volterra Equation as an example, it is shown that constructing feature layers based on characteristic patterns observed in dynamical systems can significantly improve the network’s training efficiency and accuracy. To this end, the Lotka–Volterra Equation is defined as follows:

$$\frac{dx}{dt} = 10x(1.5 - 9.5y) \quad \frac{dy}{dt} = 10y(5.7x - 1.05)$$

with initial conditions  $x(0) = 0.5$        $y(0) = 0.075$

As shown in **Figure 7(a)**, in addition to the standard layers, extra layers are added to make the network training easier. Since periodic behavior is expected in the *Lotka–Volterra Equation*, a feature layer with  $f_k(t) = \sin(kt)$  where  $k = 1, 2, \dots, n$  is added. This approach enforces periodicity in predictions, thereby enhancing accuracy. Additionally, hard constraints are applied to the boundary conditions, ensuring that the network does not need to infer them solely through training. As shown in **Table 3**, under identical conditions, the neural network achieves improved learning performance when feature layers are incorporated. **Figure 7(b)** presents the predicted solution without feature layers alongside the exact solution, whereas **Figure 7(c)** shows the predicted solution with feature layers compared to the exact solution. Both **Table 3** and **Figure 7** demonstrate that the inclusion of feature layers can significantly enhance the performance of PINNs.

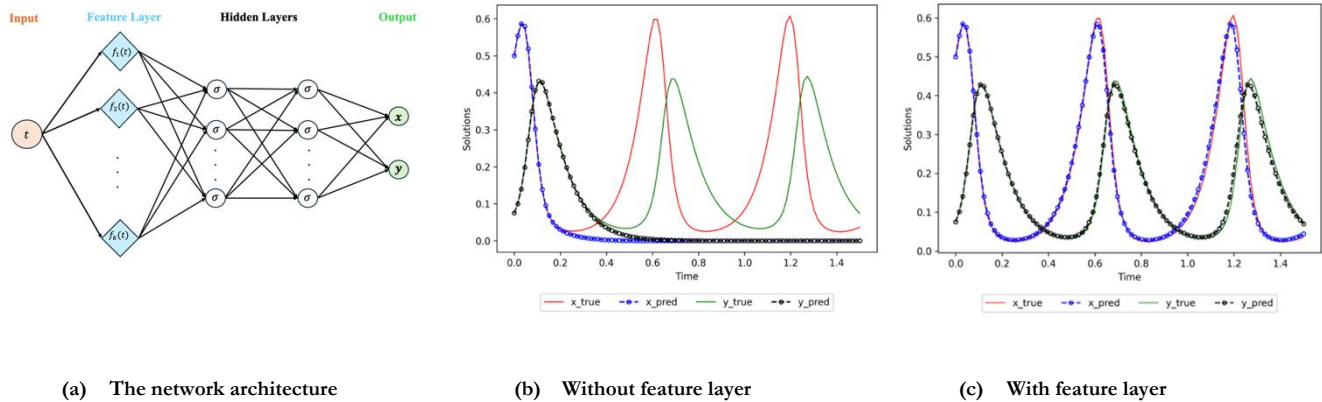


Figure 7. PINN solution to Lorenz system with noise data.

	L	N	AF	Alg	R	C	E	LW	$\mathcal{L}(\theta)_{ode}$	CT
Without features	6	64	sine	Adams, L-BFGS	0.001	400	20000	[1;1;1]	3.48e-03	599.81
With features	6	64	sine	Adams, L-BFGS	0.001	400	20000	[1;1;1]	1.10e-06	656.76

Table 3. The optimal network parameters NN Architecture. Abbreviations: L the number of layers; N the number of neurons; AF activation function; Alg optimization algorithm; R learning rate; C number of collocation points; E number of epochs; LW loss weights; CT computational time in seconds.

**Example 3. Model Performance in the Presence of Noisy Data**

This example, investigate the predictive capability of PINNs for the Lorenz system under noisy observations. The Lorenz system is defined as:

$$\frac{dx}{dt} = \delta(y - x) \quad \frac{dy}{dt} = x(\rho - z) - y \quad \frac{dz}{dt} = xy - \beta z \quad t \in [a, b]$$

with initial conditions  $x(0) = x_0, y(0) = y_0, z(0) = z_0$ . A reference solution was obtained using the classical numerical solver Tsit5() in Julia. To assess robustness to measurement noise, the reference solutions were contaminated with additive Gaussian noise, simulating observational errors. For each state variable  $j \in \{x, y, z\}$  and each time point  $t_i$ , the noisy observations were generated as  $\hat{x}_j(t_i) = x_j(t_i) + \sigma \cdot \epsilon_j(t_i)$  where  $\epsilon_j(t_i) \sim \mathcal{N}(0,1)$  are independent standard normal random variables. Three noise levels ( $\sigma = 0.2, 1.0, 3.0$ ) were considered, corresponding to low, medium, and high noise conditions, respectively. Random seed fixation ensured reproducible noise realizations. The system parameters were set to  $\delta = 10, \rho = 15$ , and  $\beta = 8/3$  with initial conditions  $x_0 = -8, y_0 = 7$ , and  $z_0 = 27$ . Optimal PINN hyperparameters for each noise level are summarized in Table 4. As noise levels increase, it may be necessary to adjust the network architecture and training parameters to preserve prediction accuracy. To this end, the number of layers, neurons, and collocation points is increased to enhance performance; however, this also results in an increase in both total loss and computational time, as shown in Table 4. Figure 8 demonstrates that PINNs can accurately recover the system dynamics from noisy observations. Each simulation employed a two-stage training procedure: an initial optimization using the Adams algorithm, followed by refinement with the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm. Figure 9 visualizes the resulting phase-space trajectories, illustrating the coupled evolution of  $x, y$ , and  $z$  and providing a comprehensive depiction of the system’s dynamics.

$\sigma$	L	N	AF	Alg	R	C	E	LW	Loss	CT
Low	3	40	tanh	Adams, L-BFGS	0.001	400	50000	[1;1;1]	1.24e-01	111.59
Medium	4	50	tanh	Adams, L-BFGS	0.001	500	75000	[1;1;1]	2.67e+00	227.64
High	5	60	tanh	Adams, L-BFGS	0.001	600	100000	[1;1;1]	2.40e+01	580.80

Table 4. The network settings for each noise level. Abbreviations:  $\sigma$  the noise level; L the number of layers; N the number of neurons; AF activation function; Alg optimization algorithm; R learning rate; C number of collocation points; E number of epochs; LW loss weights; CT computational time in seconds.

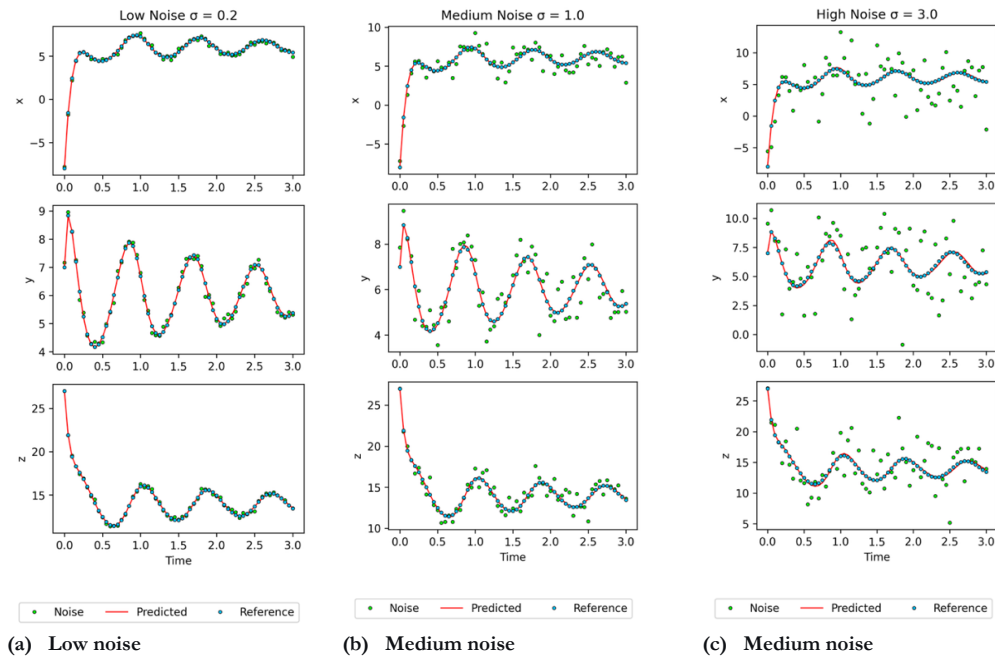


Figure 8. PINN solution to Lorenz system with noisy data.

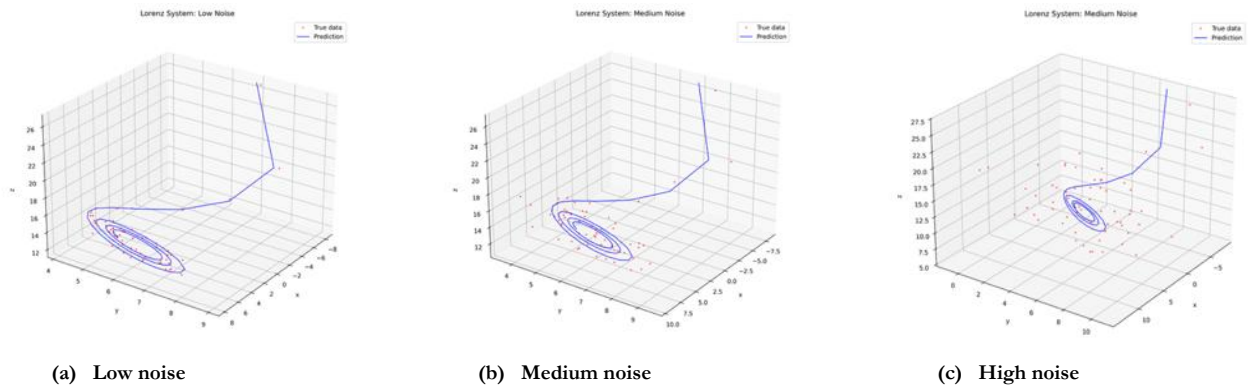


Figure 9. Phase plane of PINNs solutions to Lorenz system with noisy data.

**Example 4.** Performance in the Absence of Experimental Data

This example, demonstrate the performance of the neural network in the absence of experimental data. Consequently, the loss function comprises only two components, as described below:

$$\mathcal{L}(\theta) = \lambda_{ode}\mathcal{L}(\theta)_{ode} + \lambda_{ic}\mathcal{L}(\theta)_{ic}$$

Consider a coupled mass–spring system with parameters  $m_1 = 2, m_2 = 1, k_1 = 1,$  and  $k_2 = 1,$  initial conditions  $x(0) = 3, x'(0) = 0$   $y(0) = 3,$  and  $y'(0) = 0$  as described in the *Coupled Mass–Spring System*. Under these conditions, the general form of the *Coupled Mass–Spring System* reduces to:

$$2x'' + 2x - y = 0 \qquad y'' + (y - 1)x = 0.$$

The analytical solution to this system is  $x(t) = 2 \cos t + \cos 2t$  and  $y(t) = 4 \cos t - \cos 2t.$  **Figure 10(a)** illustrates the neural network architecture designed to solve the coupled mass–spring system, while **Figure 10(b)** compares the network predictions with the

exact solution. **Figure 10(c)** shows the evolution of the total  $L_2$  error (including the  $L_2$  errors of  $x$  and  $y$ ) with respect to the training iterations, where the  $L_2$  error is recorded every 500 iterations. The results demonstrate that, with appropriate hyperparameter tuning, the PINNs can accurately predict the system's behavior even in the absence of experimental data. The network configuration, hyperparameters,  $L_2$  error, and computational time are summarized in **Table 5**.

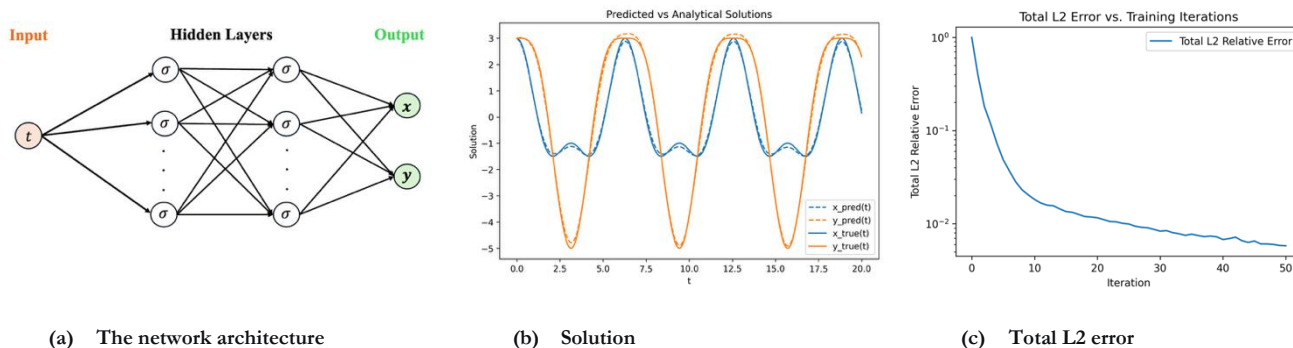


Figure 10. Graphical illustration of the network, solution and the L2 error.

L	N	AF	Alg	R	C	E	LW	$L_2$ error	CT
3	40	tanh	Adams, L-BFGS	0.001	400	50000	[1;1]	2.96e-03	220.51

Table 5. The network setting for the coupled mass–spring system. Abbreviations: L the number of layers; N the number of neurons; AF activation function; Alg optimization algorithm; R learning rate; C number of collocation points; E number of epochs; LW loss weights; CT computational time in seconds.

### CONCLUSIONS

In this paper, the Physics-Informed Neural Networks methodology for solving dynamical systems governed by ordinary differential equations is presented. Several strategies were introduced to enhance prediction accuracy, including residual weighting, the addition of feature layers, and the imposition of hard boundary constraints. The predictive capability of the PINNs approach was validated through four benchmark problems from engineering and biology, demonstrating that successful convergence for complex systems requires careful tuning of hyperparameters. The results highlight the strong predictive power of PINNs, establishing them as a promising tool for solving dynamical systems. A notable advantage of the PINNs framework is its flexibility, allowing inverse problems to be addressed with only minor adjustments to the implementation used for forward problems. Thus, future work will focus on extending the methodology to inverse problems, particularly for systems of nonlinear ODEs.

### ACKNOWLEDGEMENTS

The student authors gratefully acknowledge their research mentors, Charuka Wickramasinghe from Wayne State University and Pradeep Ranaweera from Siena Heights University, for leading and mentoring the experimental research.

### REFERENCES

- Atkinson, K., Han, W., & Stewart, D. E. (2009). Numerical solution of ordinary differential equations. *John Wiley & Sons*.
- Cheney, W., & Kincaid, D. (1998). Numerical mathematics and computing. *International Thomson Publishing*.
- Teschl, G. (2012). Ordinary differential equations and dynamical systems (Vol. 140). *American Mathematical Soc*.
- Li, H., Wickramasinghe, C. D., & Yin, P. (2026). Analysis of a  $C^0$  finite element method for the biharmonic problem with Dirichlet boundary conditions. *Numerical Algorithms*, 101(3), 1739-1784. <https://doi.org/10.1007/s11075-025-02062-4>
- Wickramasinghe, C. D. (2024). A Numerical Scheme for Solving Fourth-Order Convection–Reaction–Diffusion Problems with Boundary Layers. *Communications in Mathematics and Applications*, 15(3), 1153. <https://doi.org/10.26713/cma.v15i3.2778>
- Wickramasinghe, C. D., & Ahire, P. (2024). A Graded Mesh Refinement Method For 2d Poisson's Equation on NonConvex Polygonal Domains. *Asia Pac. J. Math*. 2024 11:63. <https://doi.org/10.28924/APJM/11-63>
- Wickramasinghe, C. D. (2022). A  $C^0$  Finite Element Method For The Biharmonic Problem In A Polygonal Domain. *Wayne State University*. [https://digitalcommons.wayne.edu/oa\\_dissertations/3704/](https://digitalcommons.wayne.edu/oa_dissertations/3704/)
- Li, J., Wickramasinghe, C., Jiang, J., Wu, A., Jiang, Y., Tovmasyan, A., ... & Sanai, N. (2025). Mechanistic modeling of spatial heterogeneity of drug penetration and exposure in the human central nervous system and brain tumors. *Clinical Pharmacology & Therapeutics*, 117(3), 690-703. <https://doi.org/10.1002/cpt.3505>

9. Wickramasinghe, C. D., & Hapuhinna, N. S. (2025). A Data-Driven Estimation of Drug-Specific Parameters in a System of Differential Equations Used in a Four-Compartment Brain Model for Brain Cancer Therapy. *Advanced Mathematical Models & Applications*, 10(2). <https://doi.org/10.62476/amma.102277>
10. Wickramasinghe, C. D., Kim, S., & Li, J. (2025). SpatialCNS-PBPK: An R/Shiny Web-Based Application for Physiologically Based Pharmacokinetic Modeling of Spatial Pharmacokinetics in the Human Central Nervous System and Brain Tumors. *CPT: Pharmacometrics & Systems Pharmacology*, 14(5), 864-880. <https://doi.org/10.1002/psp4.70026>
11. Wickramasinghe, C., Kim, S., Jiang, Y., Bao, X., Yue, Y., Jiang, J., ... & Li, J. (2025). Population Pharmacokinetic Modeling of Total and Unbound Pamiparib in Glioblastoma Patients: Insights into Drug Disposition and Dosing Optimization. *Pharmaceutics*, 17(4), 524. <https://doi.org/10.3390/pharmaceutics17040524>
12. Wickramasinghe, C. D., & Hapuhinna, N. S. (2026). TumorPred: A computational framework implemented via an R/Shiny web application for parameter estimation and sensitivity analysis in compartmental brain modeling. *Software Impacts*, 100831. <https://doi.org/10.1016/j.simpa.2026.100831>
13. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*
14. Wickramasinghe, C. D., Weerasinghe, K. C., Ranaweera, P. K., & Hapuhinna, N. S. (2025). PBPK-iPINNs: Inverse Physics-Informed Neural Networks for Physiologically Based Pharmacokinetic Brain Models. *arXiv preprint arXiv:2509.12666*
15. Banerjee, C., Nguyen, K., Fookes, C., & George, K. (2024). Physics-informed computer vision: A review and perspectives. *ACM Computing Surveys*, 57(1), 1-38. <https://doi.org/10.1145/3689037>
16. Ryu, I., Park, G. B., Lee, Y., & Choi, D. H. (2024). Physics-informed neural network for engineers: a review from an implementation aspect. *Journal of Mechanical Science and Technology*, 38(7), 3499-3519. <https://doi.org/10.1007/s12206-024-0624-9>
17. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). *Automatic differentiation in pytorch*.
18. Grohs, P., Hornung, F., Jentzen, A., & Von Wurstemberger, P. (2023). A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations (Vol. 284, No. 1410). *American Mathematical Society*.
19. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378, 686-707.
20. Pang, G., Lu, L., & Karniadakis, G. E. (2019). fPINNs: Fractional physics-informed neural networks. *SIAM Journal on Scientific Computing*, 41(4), A2603-A2626. <https://doi.org/10.1137/18M1229845>
21. Nabian, M. A., & Meidani, H. (2018). A deep neural network surrogate for high-dimensional random partial differential equations. *arXiv preprint arXiv:1806.02957*.
22. Lu, L., Meng, X., Mao, Z., & Karniadakis, G. E. (2021). DeepXDE: A deep learning library for solving differential equations. *SIAM review*, 63(1), 208-228. <https://doi.org/10.1137/19M1274067>
23. Nagle, R. K., Saff, E. B., & Snider, A. D. (2014). *Fundamentals of differential equations*. Pearson.
24. Kinga, D., & Adam, J. B. (2015, May). A method for stochastic optimization. *In International conference on learning representations (ICLR)* (Vol. 5, No. 6).
25. Moritz, P., Nishihara, R., & Jordan, M. (2016, May). A linearly-convergent stochastic L-BFGS algorithm. *In Artificial intelligence and statistics* (pp. 249-258). PMLR.

## ABOUT THE STUDENT AUTHORS

Andrew Particka is an undergraduate senior studying environmental engineering at Siena Heights University who will graduate in the spring of 2026.

Tyrus Whitman is an undergraduate junior studying mechanical engineering at Siena Heights University who will graduate in the spring of 2027.

Christopher Diers is an undergraduate senior studying cybersecurity at Siena Heights University who will graduate in the spring of 2026.

Ian Griffin is an undergraduate sophomore studying mechanical engineering at Siena Heights University who will graduate in the spring of 2028.

## PRESS SUMMARY

This study demonstrates and validates the predictive capability of Physics Informed Neural Networks for solving engineering and biological dynamical systems governed by ordinary differential equations. While traditional numerical methods often struggle with stiff, high-dimensional, or irregular problems, PINNs handle these challenges by embedding physical laws directly into the learning process. Balancing losses, tuning hyperparameters, and adding constraints without losing the generality of the ODE system are key to improving PINNs' accuracy and predictive power.

