2-7. Internship Researchers

No.	Name	Project Title
		Period
2-7-1	**Mohan BAJAJ** Indian Institute of Technology Roorkee (IIT Roorkee), India	Comparative numerical study on the effect of various combined metallic dampers with integrated yielding mechanisms on the behavior of ordinary moment resisting RC frame in the soft story
		Online: November 7, 2024 Onsite: November 28, 2024 – December 27, 2024
2-7-2	**Chari Sudhakar KAVYA** Centre for Cosmology and Science Popularization, India	Study of pulsar population statistics with machine learning.
		Online: November 1, 2024 – December 27, 2024
2-7-3	**Rohmah HIDAYAH** Institut Teknologi Sepuluh Nopember, Indonesia	Research of Biomedical Engineering
		Online: December 16, 2024 – December 20, 2024 Onsite: January 6, 2025 – February 28, 2025
2-7-4	**Pimao LI** Shandong University of Science and Technology, China	Evaluation of microscopic stress state in asperities on a rock discontinuity
		Online: September 23, 2024 - September 27, 2024 Onsite: October 7, 2024 - November 29, 2024
2-7-5	**Shofiyul Anam Ai MUBAROK** Institut Teknologi Sepuluh Nopember, Indonesia	Deep Learning Approaches for Rainfall-runoff Modeling and Flood Mitigation Strategies
		Online: November 18, 2024 – December 13, 2024 Onsite: December 17, 2024 - January 10, 2025
2-7-6	**Yoonjoo NAM** Sungkyunkwan University, Korea	The role of extrachromosomal DNA in cancer drug resistance acquisition
		Online: July 1, 2024 - July 5, 2024 Onsite: July 16, 2024 – July 29, 2024
2-7-7	**Zhichun WU** Zhengzhou University, China	Environment-friendly technology / Strengthening resilience
		Online: October 15, 2024 Onsite: November 1, 2024 - March 24, 2025
2-7-8	**Dan YUAN** Xi'an University of Technology, China	Post-earthquake damage assessment of RC structures based on deep learning
		Online: June 15, 2024 – July 31, 2024 Onsite: August 1, 2024 – September 6, 2024

No. 2-7-1	Comparative numerical study on the effect of various combined metallic dampers with integrated yielding mechanisms on the behavior of ordinary moment resisting RC frame in the soft story			
Name	Mohan BAJAJ	Title	Ph.D. Student	
Affiliation	Indian Institute of Technology Roorkee (IIT Roorkee), India Email: mbajaj@eq.iitr.ac.in			
Research Field	Strengthening resilience			
Period of Internship	Online: November 7, 2024 Onsite: November 28, 2024 - December 27, 2024			
Host Professor	Gaochuang CAI	Title	Associate Professor	
Affiliation	IROAST Email: cai@kumamoto-u.ac.jp			

Table of content

1   Introduction

Metallic dampers are energy-dissipating devices that dissipate through the mechanism of metal yield. In the last few decades, metallic dampers have received considerable attention in the field of earthquake engineering due to their stable hysteresis, long-term reliability, easy installation and replacement, and relative insensitivity to the environment. During moderate to severe earthquakes, the energy is concentrated in these dampers, and these dampers dissipate this energy through the yielding of metal. X-shaped added damping and stiffening device (X-ADAS) is the most commonly used metallic damper, which yields under flexure. Under shear loading at one end of the damper, these plates experience double curvature bending, and their tapered X-shape ensures uniform bending stress throughout the height of the plate. Similarly, the shear plate and shear link are the commonly used shear damper that yields under shear when experiencing in-plane shear loading. Shear dampers show more considerable initial stiffness and resistance than flexure-yielding dampers, but their hysteresis is less stable. In order to utilize the merits of both types of metallic dampers, new combinations of both have been proposed in the past [1-2].

The present study is focused on the numerical simulation of the integrated metallic plate dampers using the FEM-based software ABAQUS. The numerical results obtained are verified from the experimental data from the previously conducted tests on the Integrated metallic plate dampers (IMPDs) [3]. A theoretical formulation is also developed and verified for the IMPD subjected to a moment in addition to the moment at the free end.

2   Integrated Metallic Plate Dampers: Theoretical calculations

Integrated metallic dampers integrate flexure and shear-yielding components in a single damper unit. In the present study, two configurations of the IMPDs are considered. Both configurations utilize a square shear plate as a shear-yielding component. The number of X-ADAS plates is varied in the configuration as two and four and are placed on either end of a square shear plate.
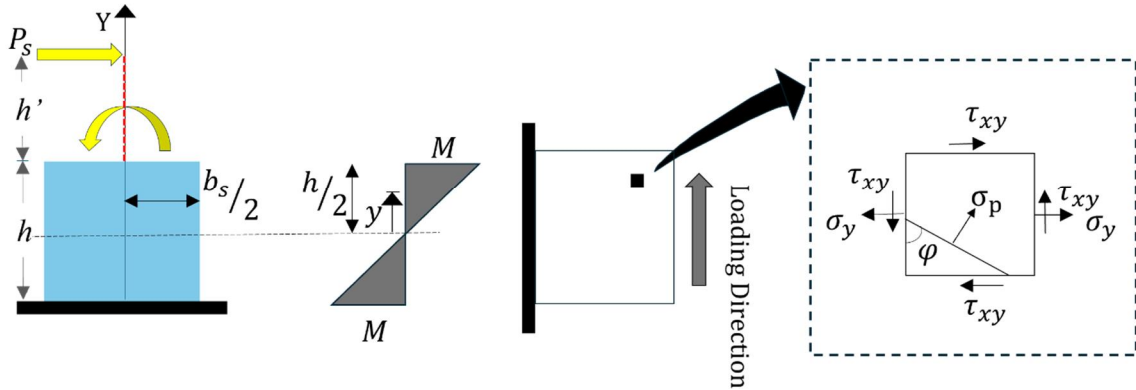

Figure 1. Formulations for shear component.

Fig. 1 shows the shear plate subjected to the moment and shear load at the free end. The stress generated due to small displacements with the plane section assumption is also expressed, where σ_y and τ_{xy} are the normal and shear stress. σ_p and Φ are the major principal stress and angle of major principal stress with the X axis. The values of principal stress can be calculated as follows:

$$\frac{\sigma_1}{\sigma_2} = \frac{\sigma_x + \sigma_y}{2} \pm \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2} \; ; \; \sigma_x = 0 \tag{1}$$

Following the distortional strain energy theory, the equivalent stress can be written as follows:

$$\sigma_e = \sqrt{\frac{1}{2}[(\sigma_1 - \sigma_2)^2 + (\sigma_2)^2 + (\sigma_1)^2]}$$

$$\sigma_e = \sqrt{[\sigma_y^2 + 3\tau_{xy}^2]} \tag{2}$$

1

The maximum value of bending and shear stress at y distance can be calculated as follows:

$$\sigma_y = \frac{M_{(y)}}{Z} \; ; \; Z = \frac{t_s b_s^2}{6} \; ;$$

$$\sigma_y = \frac{6(P_{ys} \, y + P_{ys} \, h')}{t_s b_s^2} \; ; \; \tau_{xy} = \frac{P_{ys}}{b_s t_s}$$

$$\sigma_e = \frac{P_{ys}}{t_s b_s} \sqrt{\frac{36(y + h')^2}{b_s^2} + 3} \tag{3}$$

Assuming the uniform yielding of shear plate edges along the different heights, the yield strength of the shear plate can be calculated as follows:

$$P_{ys} = \frac{f_y t_s b_s}{\sqrt{36\left(\frac{y + h'}{b_s}\right)^2 + 3}} \; ; \; y = h/2 \tag{4}$$

During relative motion between the upper and lower plates of the metallic X-Adas plate, it undergoes double curvature bending, resembling a fixed-fixed beam settled at one end (Fig. 2). If the moment generated on both the fixed ends is M, the corresponding reaction in the horizontal direction (Shear) can be calculated as:
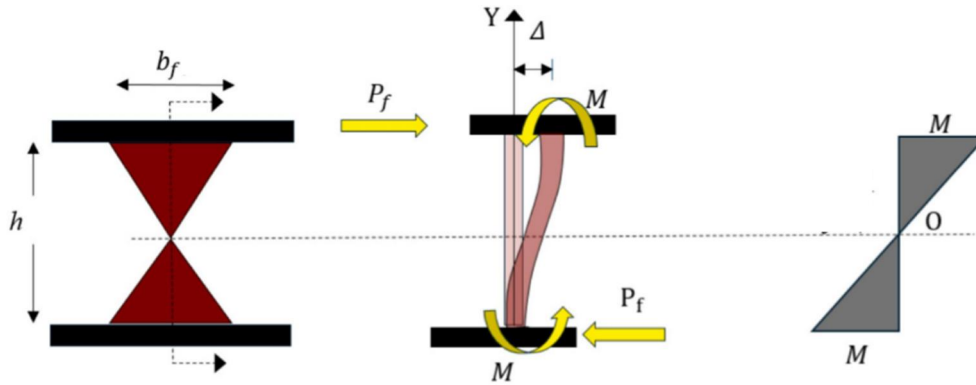

Figure 2. Formulations for flexure component.

$$P = \frac{2M}{h} \tag{5}$$

where, h is the height of the plate. By the theory of simple bending, the bending equation can be written as:

$$M = f_y \frac{bt^2}{6} \tag{6}$$

From the equation (5) and (6), yield parameters for the plate subjected to double curvature bending can be represented as follows:

$$P_y = n.f_y \frac{bt^2}{3h} \tag{7}$$

$$K = n \frac{12EI}{h^3} \tag{8}$$

$$K = \frac{12E}{h^3} \frac{b_{avg}t^3}{12} = \frac{Eb_{avg}t^3}{h^3} \; ; \; ; \; b_{avg} = \frac{b}{2} \tag{9}$$

$$\Delta_y = \frac{P_y}{K} = \frac{2h^2 f_y}{3Et_f} \tag{10}$$

Where, P_y Δ_y, and K represent the yield strength and displacement for the X-ADAS plate. Notations b, t,

and h represent the width, thickness, and height of the damper. n is the number of flexure plates. The resultant load resisted by the combined damper is the algebraic sum of the load resisted by each component. Since both shear and flexure plates have different yielding strengths and stiffness, simultaneous yielding of both components may not be possible. Thus, Assuming the elastic-perfectly plastic behavior of both the components, the overall load-displacement behavior of the combined damper can be assumed as trilinear, having three threshold boundaries where both flexural and shear components remain elastic in the first part. The yielding of the shear plate can be observed in the second part, and in the third part, the yielding of both shear and flexural plates and buckling of the shear plate can be observed.

3 Data collection

3.1 Hysteretic behavior and Failure mechanism

Fig. 3 and Fig. 4 show the hysteretic behavior of the IMPD-A dampers. Fig. 3 indicates that IMPD-A1 consists of two ADAS plates with a central square shear plate, whereas IMPD-A2 comprises four ADAS plates. The hysteresis curve shows that the damper IMPD-A1 and IMPD-A2 exhibit elastic behavior till 2.75% drift.
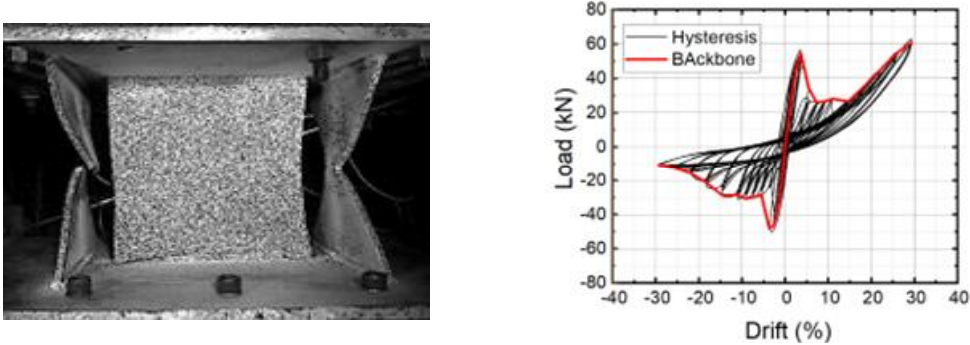

Figure 3. Failure mechanism and hysteretic behavior of IMPD-A1

IMPD-A1 and IMPD-A2 show maximum resistance of 54.54 kN and 73.388 kN, respectively. After the initiation of out-of-plane buckling of the square shear plate, a drop in the hysteresis curve is observed. The increasing branch in IMPD-A1 after 15 % drift is due to the collision of the broken ADAS plates together. The flexure component in IMPD-A1 and A2 undergoes double curvature bending, but one of the ADAS plates in IMPD-A1 experiences early failure due to stress concentration. Compared to IMPD-A1, the shear component in IMPD-A2 shows relatively improved behavior in terms of diagonal tension yielding and compression buckling.
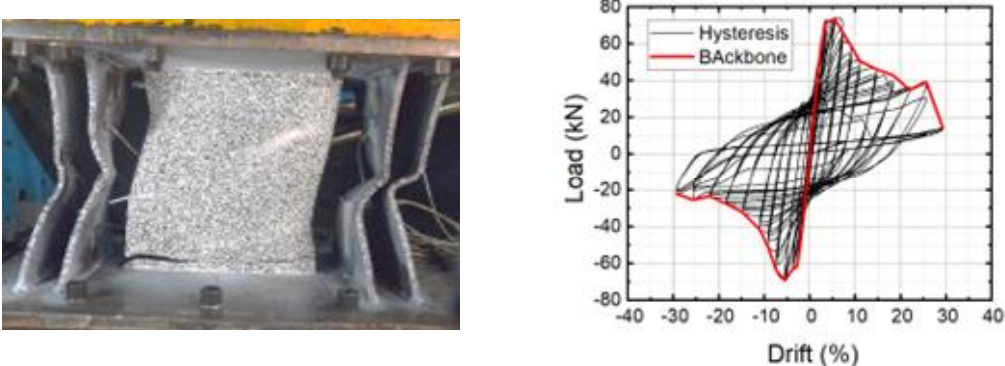

Figure 4. Failure mechanism and hysteretic behavior of IMPD-A2

4   Numerical Modeling

An analytical study has been conducted to predict the cyclic behavior of the IMPD samples using finite element analysis software ABAQUS. Quadrilateral four-node double-curved shell (S4R) elements have been used to model all the components. To consider the effect of loading assembly, the whole assembly was modeled using the S4R element. The connection between the loading assembly and damper is not modeled and is assumed to be rigidly connected. The material properties of the damper components have been assigned as per the experimental study. The initial imperfection of the damper plates has been incorporated before the testing using linear *perturbation*. The bottom plate has been assumed to be fixed, and the loading has been applied to the upper loading assembly. The effect of heat has not been considered, and the connection between the members is assumed to be rigid. A finite element mesh has been generated by seeding the edges. Shell elements (S4R) of 10mm size have been used to mesh the entire model using free mesh using the medial axis theorem.

4.1   Detailed Modelling:

Part: A 3D deformable shell element (S4R) is used to design the damper. Fig. 5 shows the dimensions of the flexure and shear components. The height and end widths of the ADAS plates are kept at 273 mm, and a reduced section of 40 mm is kept at the center. A square plate of 273 mm x 273 mm is used as a shear component. The dimensions of the bottom and upper base plate are maintained as 655 mm x 375 mm x 20mm. The upper loading beam and rigid bracket (loading beam assembly) are also designed as the experimental test.
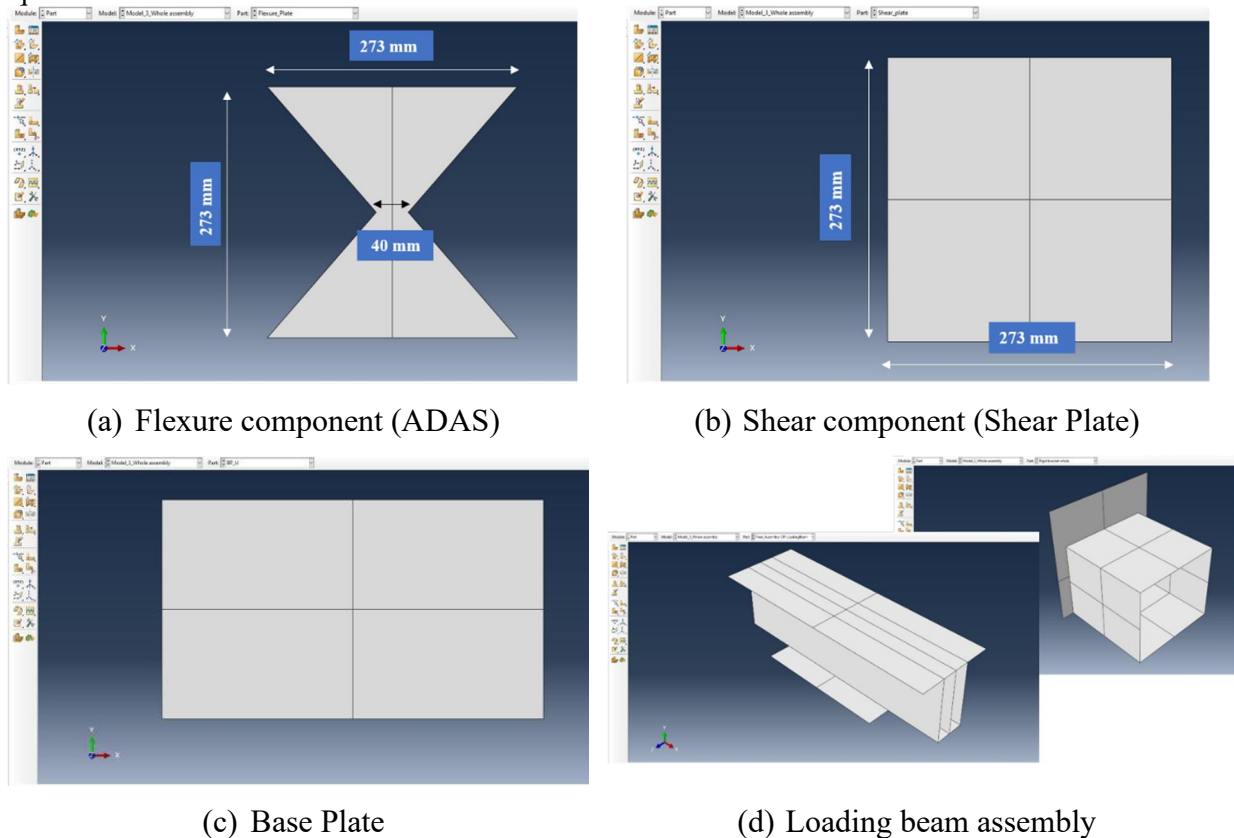

(a) Flexure component (ADAS)	(b) Shear component (Shear Plate)
(c) Base Plate	(d) Loading beam assembly

Figure 5. Part module of Damper assembly

Material and section assigned: Elastic properties are assigned to the base plates and loading assembly. Fig. 6 &7 show the material properties assigned to the shear and flexure components. Similar elastic properties are assigned to both components. The elastic modulus and Poisson's ratio for these components are assigned as 1.8×10^5 and 0.3, respectively. Plastic properties of the material are provided using combined hardening, and the calculated parameters for the material are assigned. The thickness of the flexure components is taken as 8 mm as per the experimental program; similarly, the thickness of the shear plates
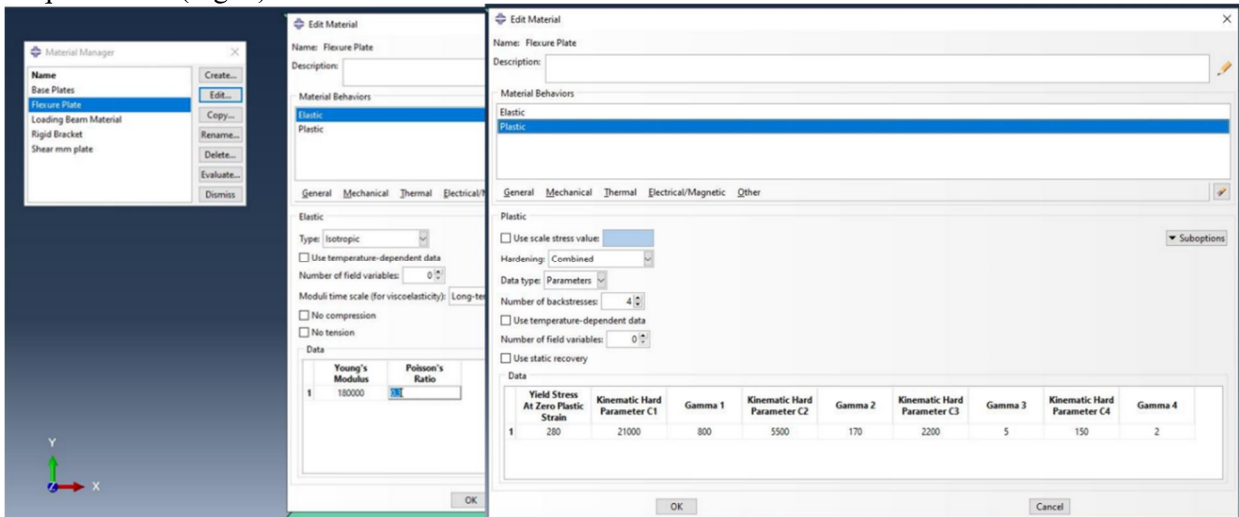
is kept at 5 mm (Fig. 8).

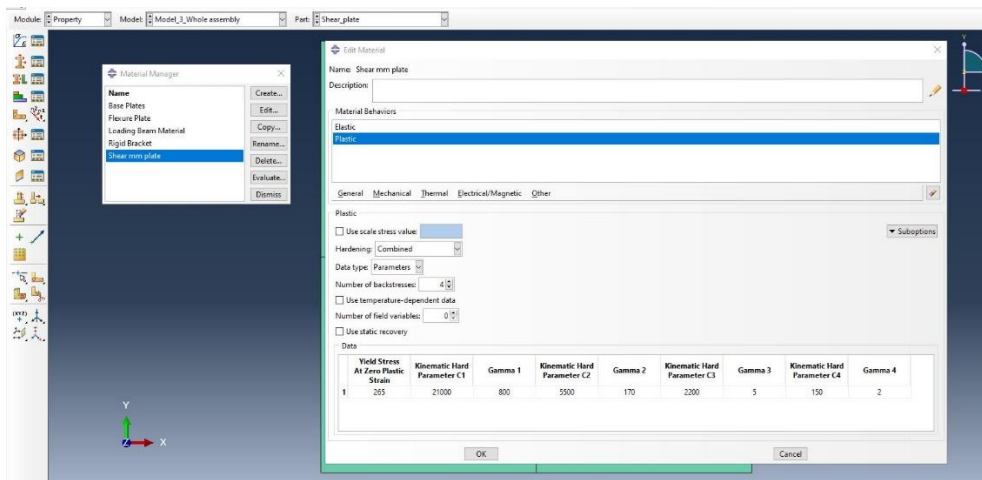Figure 6. Material assigned to the flexure component.
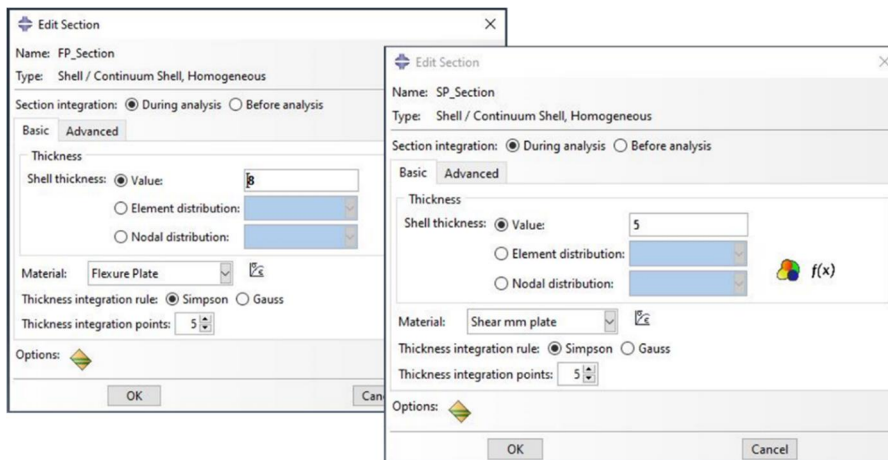

Figure 7. Material assigned to the shear component.

Figure 8. Sections assigned to the flexure and shear components.

Assembly: All the components are placed at their appropriate locations, and the final assembly can be seen

in Fig. 9. The whole assembly is assumed to be connected rigidly and is made a single unit using the merge instances icon.
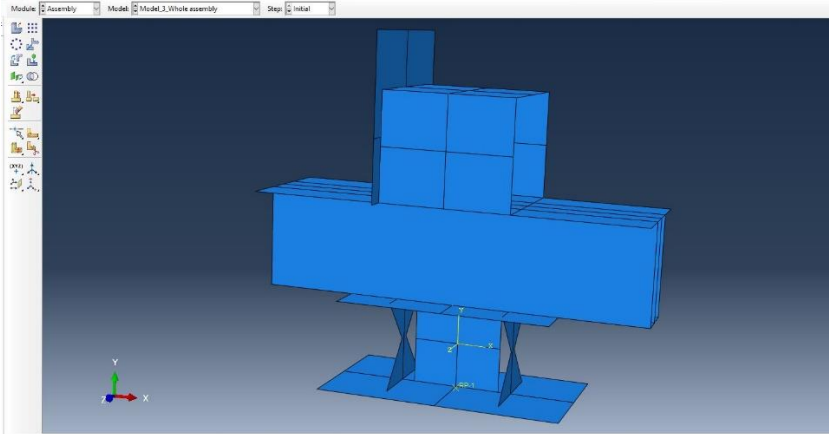

Figure 9. Final assembly of the Damper testing

Interaction, loading and Mesh: The bottom face of the base plate is coupled to a reference point just below the plate face and the fixed. The loading is applied through the rigid bracket provided at the top of the loading beam assembly. The whole part is meshed with a free meshing technique using the medial axis algorithm available in the Abaqus. The mesh size is kept at 10 mm (Fig. 10).
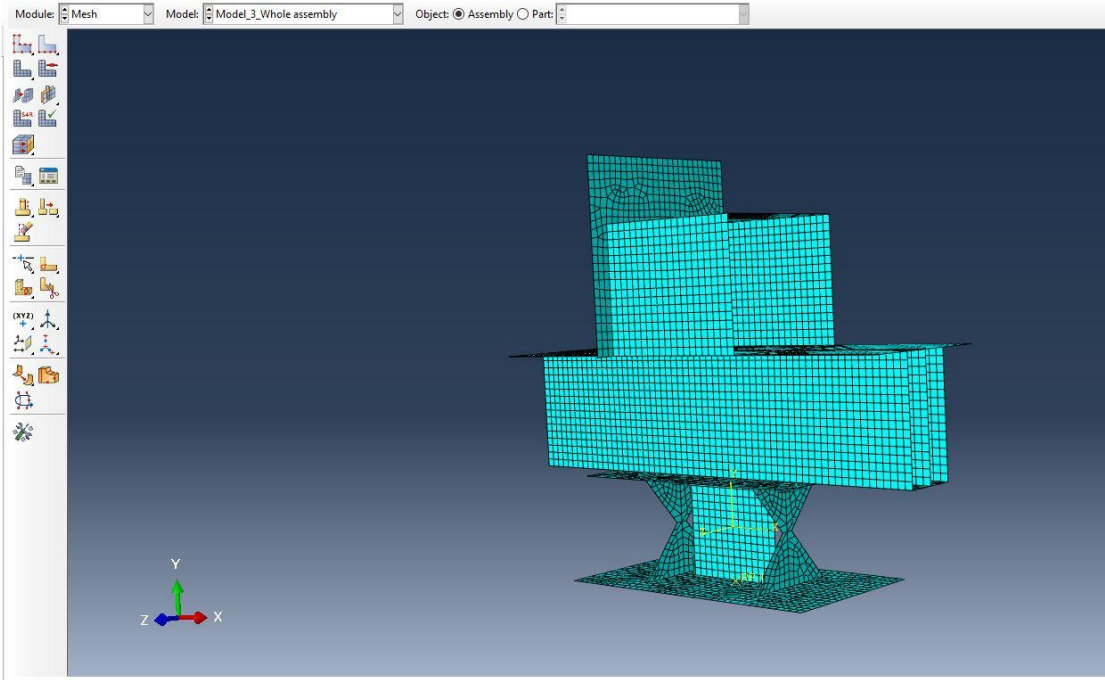

Figure 10. Meshing of the whole setup

Initial Imperfection: The initial imperfections in the shear plates available in the shear plate before the testing are simulated using the linear perturbation option in the step icon of the ABAQUS. A total of 15 modes of the dampers are monitored, and the desired imperfections are incorporated into the damper (Fig. 11).
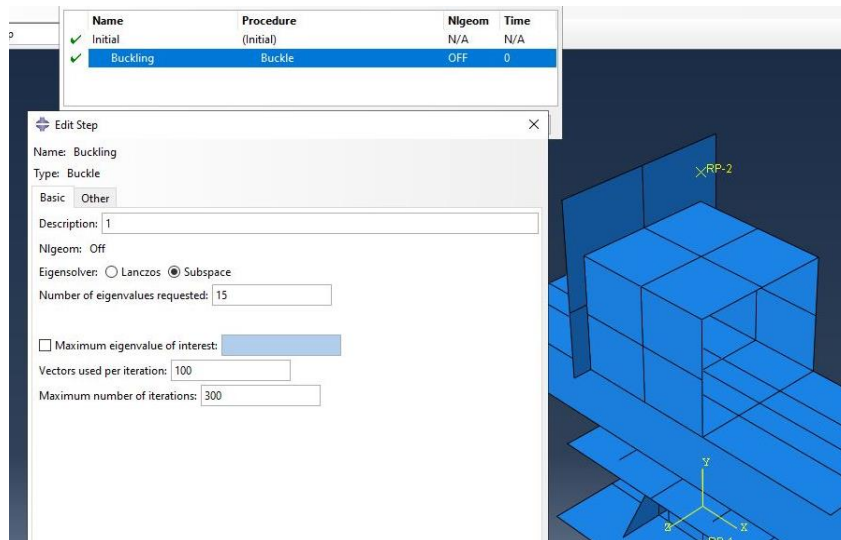
Figure 11. Consideration of the initial imperfection of the damper components

5 Comparison of numerical and experimental results

The results obtained from the numerical modeling are verified with the experimental results.
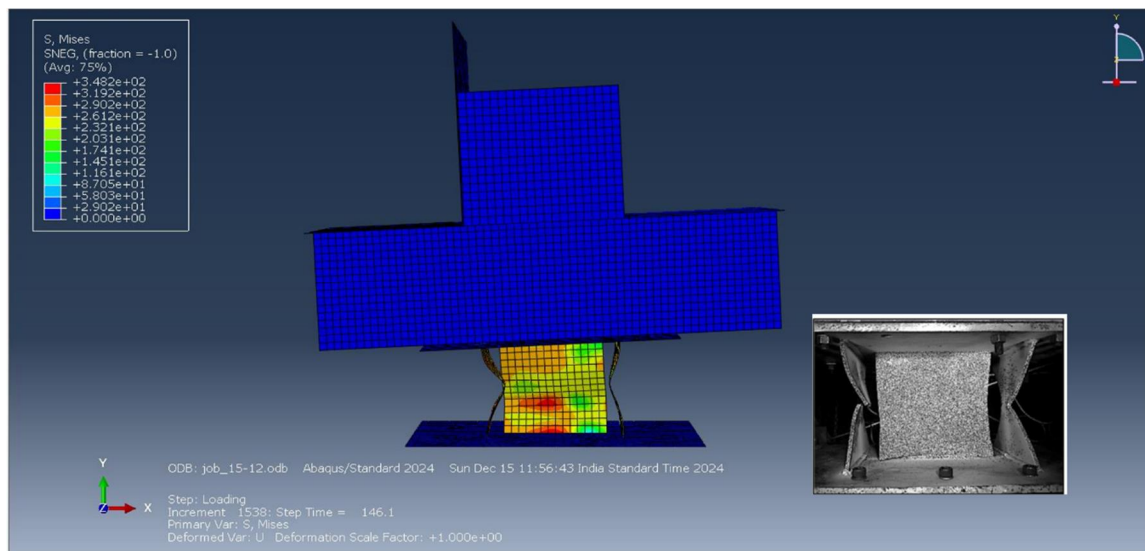

Figure 12a. Comparison: Deformation behavior of numerical and experimental model
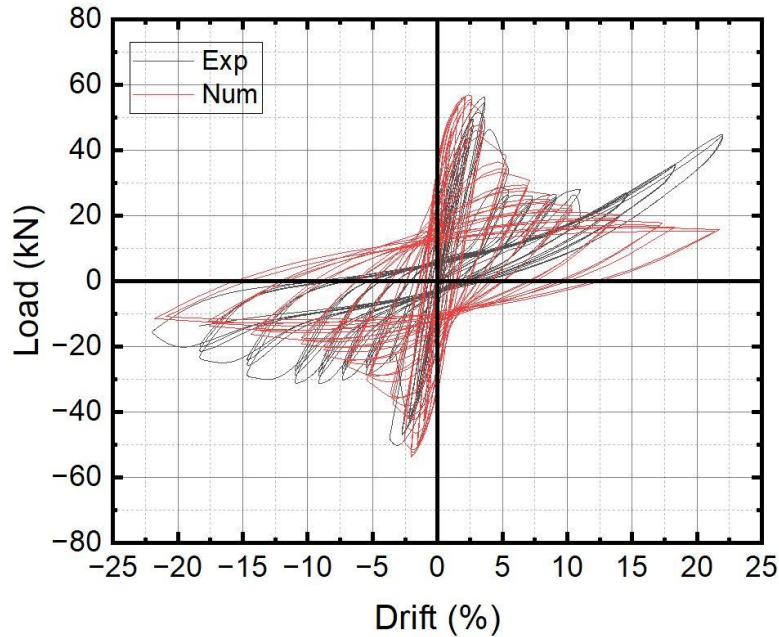
7

Figure 12b. Comparison: hysteretic behavior of numerical and experimental results.

Fig. 12 shows that both results are in good agreement except after the 15 % drift, after which broken ADAS plate counterparts collide. Additionally, the area enclosed by the numerical results are slightly higher than the experimental results. That may be due to the slippage effect in the bolted connection between the loading beam assembly and damper, which is not explicitly designed in the present modeling.

6   Limitations and future scope

As discussed in the previous section, despite a good match, there are some deviations due to setup limitations and connections. The imperfection is modeled in the present study, but the bolted connections need to be modeled. Early failure due to crack also needs to be incorporated to get a perfect match. Moreover, the other damper IMPD-A2 with four ADAS plates also needs to be modelled and simulated.

As a future scope, both configurations of the dampers will be studied under pure shear cyclic loading. They will be assigned to a moment-resisting RC frame to study the behavior of the RC frame under both types of dampers.

7   Lab Visit and Practical Exposure

During my internship, I had the opportunity to visit the state-of-the-art laboratory and actively work alongside the research team and colleagues. The enriching experience offered hands-on exposure to advanced testing setups and instrumentation techniques.

The lab houses an impressive setup for testing **Concrete-Filled Tubes (CFTs)**. I observed the meticulous preparation process, where the team independently handled all aspects of the setup. This included welding, cutting, grinding, and all instrumentation. The use of heavy machinery, such as cranes, to move large components further highlighted the technical expertise and self-sufficiency of the team.

In addition to the CFT testing setup, I explored the ongoing research projects of other PhD students. Their work focuses on resilient structures, employing innovative materials such as higher-strength reinforcement steel bars and Fiber-Reinforced Polymers (FRPs). This exposure broadened my understanding of cutting-edge techniques and materials used in structural engineering.

The collaborative environment in the lab, combined with the opportunity to observe and discuss these projects, significantly enhanced my technical knowledge and practical skills. It was inspiring to see how the team integrates theoretical knowledge with hands-on experimentation to advance the field of structural engineering.

8

8   Summary and Future Collaboration:

Work Done: Numerical modeling of one of the integrated metallic dampers with combined yielding mechanism IMPD-A1 is completed in the limited time span. The modeling of another integrated damper with four ADAS plates is in progress and is supposed to be completed by the end of the internship period.

Work to be Done: The Numerical Simulation of the IMPD-A2 damper needs to be completed and verified with the experimental results as a future scope. Additionally, the cyclic performance of these dampers under pure shear loading needs to be calculated, and the obtained results are supposed to be utilized in the calculation of the cyclic performance of RC frames mounted with these dampers.

Short-term goals: To complete the remaining work and *convert the work into a publication in a well-established journal*.

Long-term goals: To continue future collaborations with Prof. Cai and IROAST and explore the various aspects of the Structural dynamics.
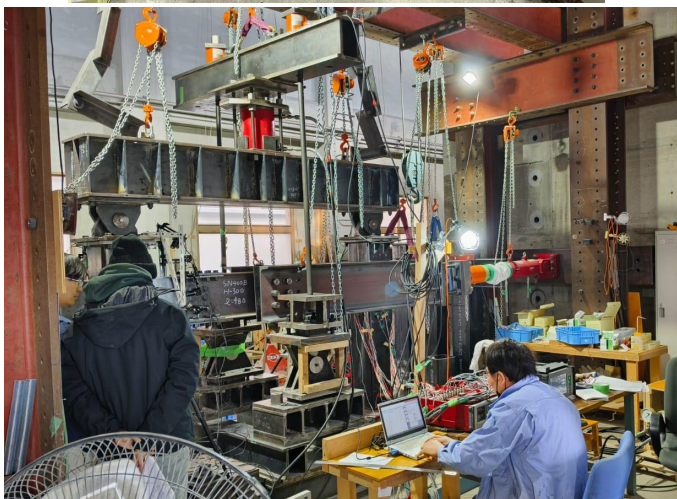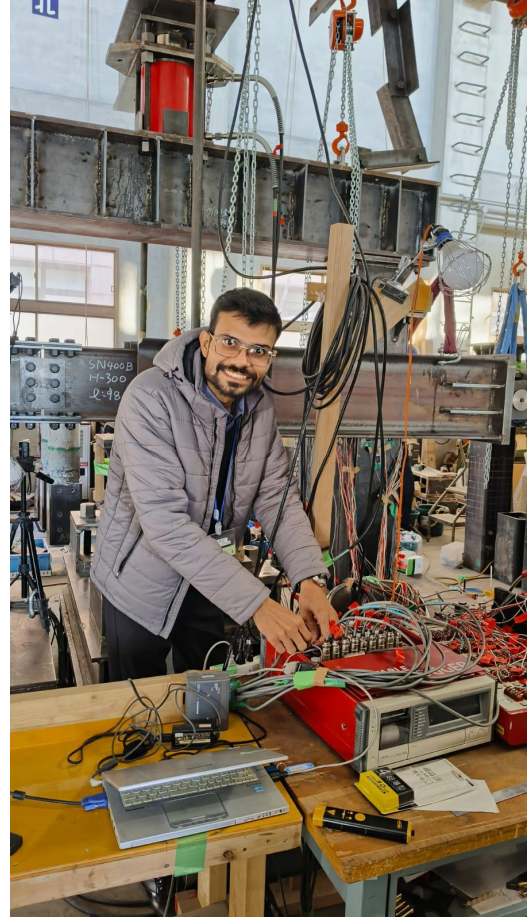
References:
[1] Sahoo, D. R., Singhal, T., Taraithia, S. S., & Saini, A. (2015). Cyclic behavior of shear-and-flexural yielding metallic dampers. Journal of Constructional Steel Research, 114, 247-257.
[2] Li, Z., Shu, G., & Huang, Z. (2019). Development and cyclic testing of an innovative shear-bending combined metallic damper. *Journal of Constructional Steel Research*, *158*, 28-40.
[3] Bajaj, M., Agarwal, P. (2025) Configuration and scale effect on cyclic performance of integrated plate dampers. Journal of Constructional Steel Research, 224(A), 109111.

| No. 2-7-2	Study of pulsar population statistics with machine learning			
Name	Chari Sudhakar KAVYA	Title	Project Intern	
Affiliation	Center for Cosmology and Science Popularization (CCSP), India Email: kavya.chari5@gmail.com			
Research Field	Data science and AI			
Period of Internship	Online: November 01, 2024 – December 27, 2024			
Host Professor	Keitaro TAKAHASHI	Title	Professor	
Affiliation	Faculty of Advanced Science and Technology Email: keitaro@kumamoto-u.ac.jp			

Details of activities

Welcome to the Pulsar Astronomy!

What are pulsars?

Pulsars are a special type of neutron star. Neutron stars are incredibly dense remnants formed when a massive star (typically around 1-3 solar masses) undergoes a supernova explosion. This explosion causes the core of the star to collapse, compressing all of its material into a tiny, dense "carcass" that's only about 10-15 km in diameter, roughly the size of a small city. This intense density results in a star supported not by normal gas pressure, but by neutron degeneracy pressure — a quantum mechanical effect that prevents neutrons from being forced into the same state.
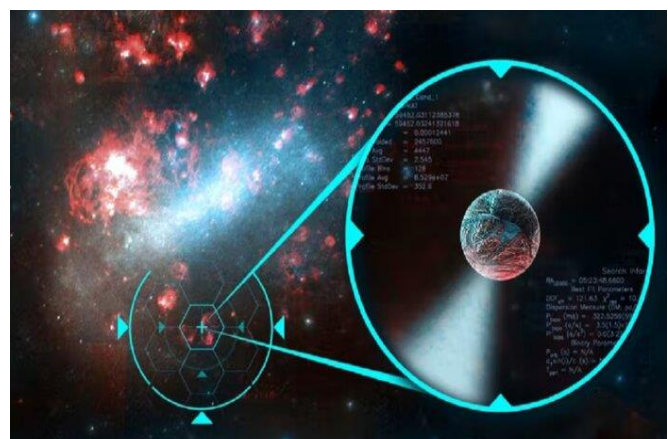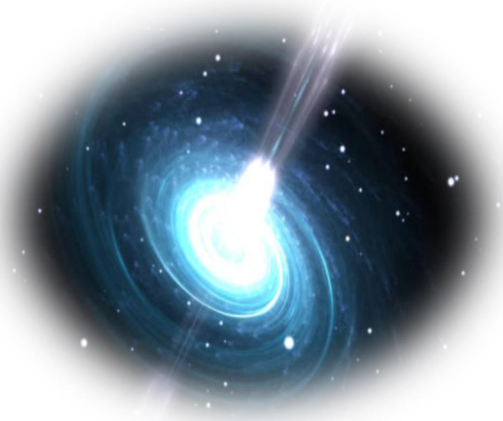
Credit: singularityhub.com

➢ **Discovery of pulsars**

- The first pulsar was discovered in 1967 by 'Jocelyn Bell Burnell' (a PhD in radio astronomy at Cambridge University) and her professor 'Antony Hewish'. This was a groundbreaking discovery that opened the up the new ways to explore astrophysics.
- Bell noticed a small strange scruff in the data. They initially thought it might be a signal from aliens, but it was an evidence of pulsar which is separated by 1.33 sec from same position in the sky.
- Second pulsar was discovered in the Crab Nebula with a period of 33 ms confirmed that it was rotating neutron star. At present, nearly 3000 known millisecond pulsars are there.

Prof.Antony Hewish

By: S. Jocelyn Bell Burnell

Credit:www.bigear.org
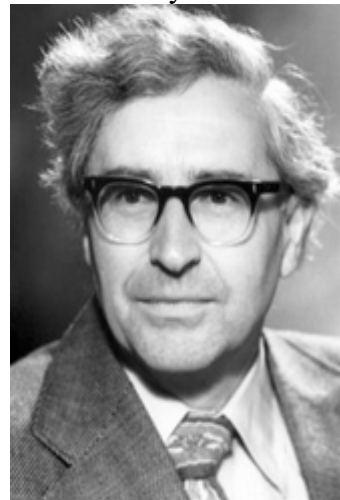
Credit: Wikipedia

➢ **Importance of pulsars as comic lighthouses:**

- Pulsars act as natural laboratories for extreme physics.
- Pulsars often exist in extremely strong gravitational fields which help researchers test aspects of Einstein's theory of General Relativity.
- For example, some pulsars are in binary systems with other stars or black holes, allowing scientists to observe the effects of intense gravitational interactions.
- Also the arrival time of pulses will hint the presence of sensitive gravitational waves.
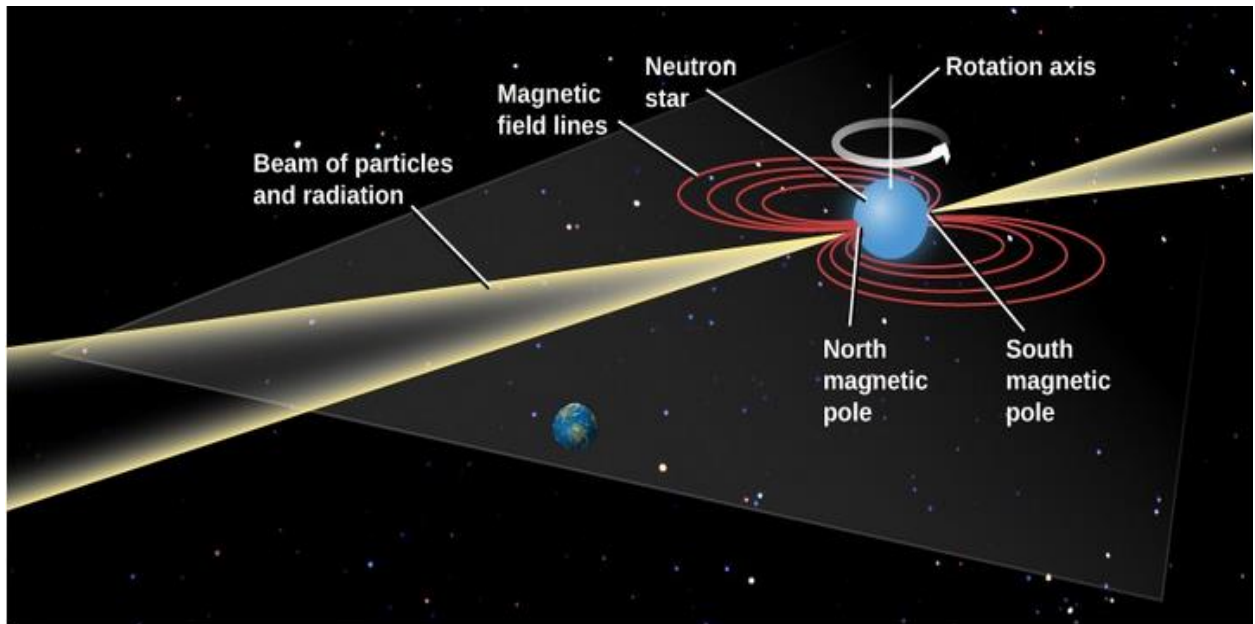
Figure 3: Model of a Pulsar. A diagram showing how beams of radiation at the magnetic poles of a neutron star can give rise to pulses of emission as the star rotates. As each beam sweeps over Earth, like a lighthouse beam sweeping over a distant ship, we see a short pulse of radiation. This model requires that the magnetic poles be located in different places from the rotation poles. (credit "stars": modification of work by Tony Hisgett)

➢ Light house model

- The lighthouse model explains how we see pulsars as flashing or pulsing objects.
- Imagine a lighthouse on Earth: as its light rotates, we only see the beam when it points directly at us.
- A pulsar works similarly, sending out beams of radio waves in specific directions. We observe a "pulse" every time one of these beams points toward us**.**

Credit: Adobe stock

Credit: http://www.seasky.org/celestial-objects/pulsars.html
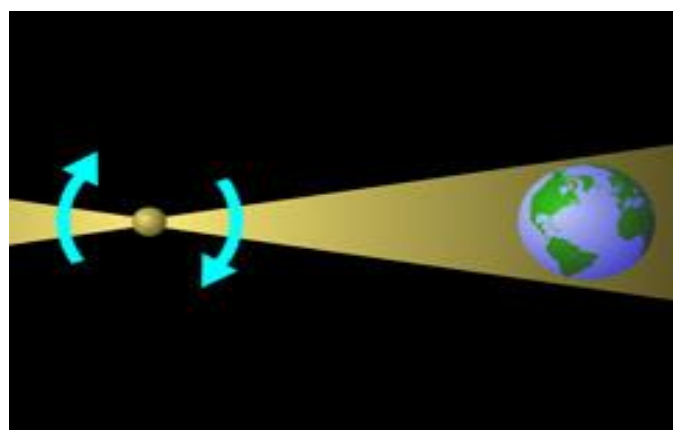
➢ Characteristics of pulsars

- **Rotation Period**: Normal pulsars have rotation periods ranging from 1.4 milliseconds to several seconds. The most typical pulsars rotate with periods of about 1 to 10 seconds**.**
- **Emission**: Pulsars emit beams of electromagnetic radiation, primarily in the form of radio waves**,**

but they can also emit X-rays and gamma rays. This emission is highly regular, appearing as pulses.

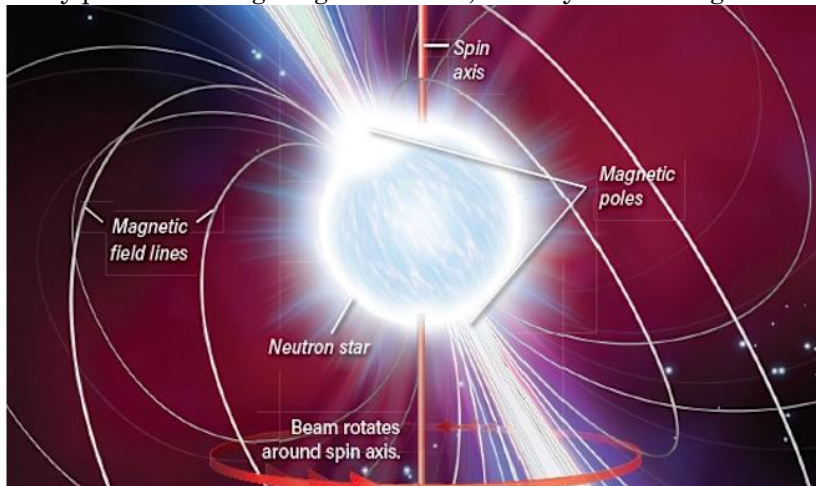- **Neutron Stars**: Pulsars are a type of neutron star, which means they are incredibly dense remnants of massive stars that have undergone supernova explosions. They are about 20 kilometers in diameter but contain more mass than our Sun.
- **Magnetic Field**: They possess strong magnetic fields, usually in the range of 10^8 to 10^{15} gauss.

(Credit: Roen Kelly/Discover)

➢ Classification of pulsars

The two main classes of pulsars and they are treated differently in gravitational wave (GW) research:

1. Canonical Pulsars:

- These are young pulsars with typical rotation periods around 1 second and are often newly formed from recent supernova explosions.
- Canonical pulsars aren't ideal for precise timing (which is crucial for GW searches) because their rotation can be unstable over time.
- They may also experience glitches—sudden spin-up events thought to happen when the inner neutron super fluid momentarily reconnects with the outer crust, causing a jump in spin rate.

2. Millisecond Pulsars (MSPs):

- MSPs are older pulsars with much faster rotation rates, typically spinning in milliseconds. They started as slower rotators but were recycled: they gained speed by accumulating material and angular momentum from a nearby binary companion.
- MSPs are highly stable in their rotation, making them reliable for long-term timing. This stability makes them excellent candidates for detecting gravitational waves since they don't experience the unpredictable spin changes seen in younger pulsars.

➢ Formation of pulsars

When a massive star collapses, if its core is between about 1.4 and 3 solar masses, it can form a neutron star. During the collapse of a massive star's core, gravity pulls the material inward intensely. As the core shrinks and density rises, neutron degeneracy pressure kicks in, countering the inward pull of gravity. If the core's mass is within the neutron star range (1.4 - 3 solar masses), this pressure halts the collapse, forming a neutron star. However, this rapid core collapse triggers a rebound effect in the outer layers, causing the star's outer layers to violently explode outward in a supernova

explosion.

Neutron degeneracy pressure:
Neutron Degeneracy Pressure is a quantum mechanical force that arises when neutrons are packed extremely closely, as in the core of a collapsing massive star. This pressure occurs due to the Pauli Exclusion Principle, which states that no two neutrons (or other fermions like electrons) can occupy the same quantum state. When neutrons are forced into very close proximity, they resist further compression, creating a pressure that acts against gravitational collapse.

➢ InterStellar Medium(ISM) effects on pulsars

1. **Dipersion**: The pulses from a pulsar represent its rotation, and as these radio waves travel through the interstellar medium, which contains free electrons (plasma), they experience dispersion. This effect causes lower frequency pulses to arrive later than higher frequency ones, resulting in a diagonal shift in the frequency versus phase plot. Analyzing this data is essential for understanding the pulsar's characteristics and the influence of the interstellar medium

2. **Pulse broadning**: When waves scatter, some of them travel longer distances because they are bent or refracted by irregularities in the ISM. These bent waves arrive later than direct waves, causing a "tail" effect in the pulse signal. This pulse broadening looks like a stretched or smeared-out pulse and depends strongly on frequency (generally proportional to v^{-4}), with lower frequencies being affected more.
The broadening decreases the overall strength of the pulse signal and can reduce the precision of timing measurements, since it adds noise or blurring to the arrival times. So, the ISM can cause multiple issues for pulsar signals, it not only delays lower frequencies (dispersion) but also causes fluctuations in signal strength (scintillation) and stretches the pulse due to multipath scattering.

3. **Scintillation**: Inhomogeneities in the turbulent ISM lead to diffractive and refractive scintillation, which cause variations in the flux density of the pulsar signal, similar to how stars appear to twinkle due to the Earth's atmosphere.
Diffractive Scintillations: Occur over timescales of minutes to hours and can involve radio bandwidths ranging from kilohertz to hundreds of megahertz, resulting in significant fluctuations in flux density (more than an order of magnitude).
Refractive Scintillations: Generally produce smaller amplitude changes (less than a factor of 2) and occur over longer timescales, such as weeks.

- **Reason:**
  1. The ISM consists of free charged electron density forming plasma. It possess a refractive index which determines how radio waves pass through it. Mathematically it can be expressed-

  $$\mu = [1 - (v_p/v)^2]^{1/2}$$

  where, v_p= plasma frequency, v= frequency of radio waves.

  Here plasma frequency depends on density of electrons in the ISM. Here the refractive index of ISM is less than 1 which affects the propagation of radio waves passing.

  2. Plasma frequency is the threshold frequency below which radio waves cannot travel through plasma. Mathematically expressed as-

  $$v_p = (e^2 n_e/\pi m_e)^{1/2}$$ = approx. 8.97kHz (constant multiplier).

  where, e = electron charge, n_e= 0.03 cm^{-3} = electron number density in per cm cube, m_e= mass of electron. So we get plasma frequency around 1.5kHz.

  3. Hence, If the radio frequency is lower than plasma frequency then waves can't propagate through ISM. However, for most radio frequencies used in observations the radio frequency would be greater, so waves can travel through it.

Why lower frequencies slower down?

Since refractive index depends on v_p/v the lower frequency waves(closer to plasma frequency) experience a greater reduction in group velocity. Mathematically it is expressed as

$$v_g = c \times 1 - (v_p/2v)^2$$

Hence this reduction in velocity is more at lower frequencies, causing these lower frequencies travel slower and experience a delay compared to higher frequencies.

➢ **Pulse phase and broadening effect**
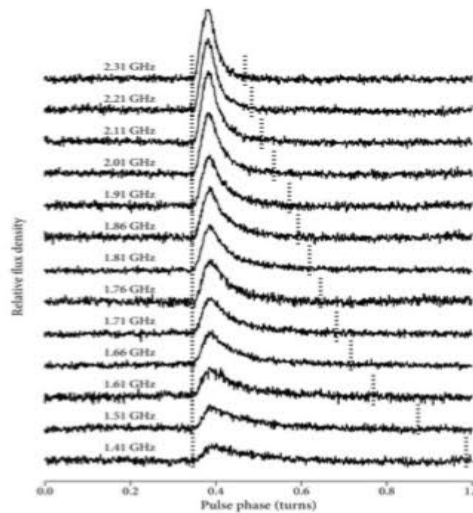
This is plot of Galactic center magnetar J1745-2900 shows the relative flux density of a pulsar signal on the y-axis versus the pulse phase on the x-axis,with each line representing observations at different frequencies.

1. **Pulse profile:** Each peak in the plot represents a single pulse from the pulsar. The x-axis (pulse phase) indicates the progress through one full rotation of the pulsar, from 0 to 1.0 (or 0 to 360

degrees in phase).

2. **Frequency dependence of pulse delay:** We notice how the pulse shifts to the right (higher pulse phase) at lower frequencies. This is due to dispersion in the interstellar medium (ISM), which delays lower frequencies more than higher ones as they travel through the ionized gas and electrons in the ISM. At lower frequencies the range is wider because the scattering causes the pulse to spread more making it broadening.

3. **Relative flux density:** The y-axis shows the strength of the pulse signal at each frequency. The peak shape indicates how the intensity of the pulse varies over one rotation of the pulsar.

4. **Dispersion smearing:** At lower frequencies, you can see that the pulse profiles are slightly broader or "smeared." This happens because dispersion effects cause a spread of arrival times across the pulse, especially at lower frequencies, which can lead to a more extended pulse profile.

5. **Aligning pulses with De-dispersion:** In practical analysis, researchers use de-dispersion techniques to shift these pulses back to a common phase, aligning them to study the intrinsic pulse shape more accurately without the delay caused by the ISM.

➢ Pulsar Timing Array(PTA)
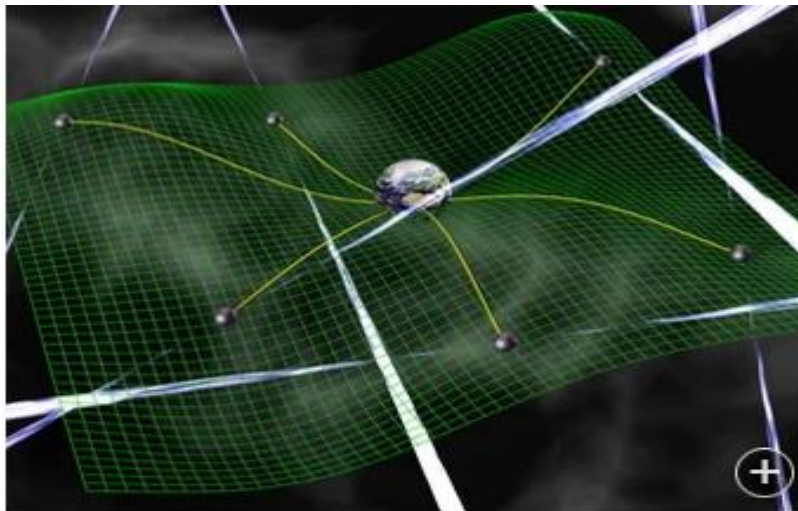
Pulsar Timing Arrays are searching for low-frequency gravitational waves by regularly observing many millisecond… [more]
picture: David Champion/Max Planck Institute for Radio Astronomy

Pulsar timing is a powerful observational technique that utilizes pulsars—rapidly rotating neutron stars that emit beams of radio waves.

1. Pulsar timing involves closely monitoring the arrival times of these radio pulses. By recording when each pulse arrives, astronomers can track the pulsar's rotation very accurately over long periods (years to decades).

2. Precision and Accuracy: This precise timing allows researchers to account for every rotation of the neutron star. This level of detail is critical because it helps in understanding the physical conditions inside neutron stars and how they behave over time.

3. Although individual pulses may vary slightly due to random changes in the emission region, the average pulse shape over many rotations is stable and consistent, sometimes even for decades which is a unique "signature" for each pulsar, allowing us to measure each rotation precisely.

4. Applications of PTA: Used to study the interior Physics of Neutron Stars, Detecting Gravitational Waves at low frequencies as it allows us to track the exact arrival times of radio pulses with high accuracy.

➤ Pulsar Timing Array process

1. Starting with Raw Data in the Time Domain: The raw data are indeed initially in the time domain, showing the signal as it arrives over time.

2. Conversion to Frequency Domain using FFT: You convert the time-domain data to the frequency domain using a Fast Fourier Transform (FFT). This step helps reveal how the signal varies with frequency, which is essential for addressing dispersion.

3. Observing the Dispersion Curve: lower frequencies experience more dispersion (they arrive later), while higher frequencies arrive sooner. In the dispersion curve resulting pulse phase shifts more for lower frequencies, causing a "spread-out" effect in the signal.

4. Phase Correction through De-dispersion: There are two methods as follows-
- Incoherent Dedispersion: Splits the signal into sub-bands, adjusts for delays individually, and then combines them. It's simpler but might not achieve the highest precision.
- Coherent Dedispersion: Processes the entire signal as a whole single wave, applying a correction that accounts for all frequencies simultaneously. It's more complex and computationally intensive but typically offers greater accuracy. We use coherent dispersion.

5. Exact Time of Arrival (TOA): Finally, by adding this time offset to a reference point (like the start of your observation), you calculate the Time of Arrival (TOA) for each pulse. This TOA tells you exactly when each pulse from the pulsar reaches Earth.

6. Inverse Fourier Transform: After de-dispersion, apply the inverse Fourier Transform to bring the data back from the frequency domain to the time domain. Now, the pulses are "compressed" and aligned in time, resulting clearer peaks.

7. Folding and Averaging: With the data back in the time domain, by folding or average over many pulses. This process gives a sharper, more stable pulse profile by reducing random noise.

8. Multiplying by the Pulse Period: After obtaining a stable average pulse profile, we calculate the

phase shift by comparing this profile with a template profile. Once we have the phase shift, we multiply it by the pulse period. This gives a time offset that helps us to align the observed profile accurately with the expected profile.

9. Convert TOA from Topocentric to Barycentric time: To make measurements consistent and independent of Earth's motion, we adjust all TOAs to the Solar System Barycenter (SSB), a stable reference frame centered near the Sun.
- Einstein Delay: Time dilation and gravitational redshift caused by the gravitational influence of the Sun and other solar system bodies.
- Shapiro Delay: Delay caused by the pulse passing through the gravitational field of the Sun, making the signal take slightly longer to reach us.
- Roemer Delay: The light travel time between Earth and the SSB, which varies as Earth orbits the Sun.
- Clock Corrections: Adjustments to sync the observatory's local time with global timing standards.

Timing equation –

$$t = t_t - t_0 + \Delta_{clock} - \Delta_{DM} + \Delta_{R\odot} + \Delta_{E\odot} + \Delta_{S\odot} + \Delta_R + \Delta_E + \Delta_S$$

By applying these corrections, we can accurately pinpoint the exact moment a pulse left the pulsar. This high-precision timing model allows us to study the pulsar's rotation and orbital characteristics, detect subtle changes in its behavior, and potentially observe gravitational wave effects. Accurate pulsar timing relies on accounting for every influence on the TOA, from Earth's position to the interstellar medium, and even the pulsar's companion star if it has one.

10. Template Comparison for Shape Verification: Finally, comparing with the template we get the accuracy and stability of the pulse shape. If there's any discrepancy, we can adjust and verify until the observed pulse shape aligns well with the template.

11. Timing residual: This is the difference between corrected TOA and predicted TOA. Predicted TOA is when we use pulsar's known spin frequency and slowdown rate to predict each pulse should arrive at SSB. These residuals serves as indicators of uncertainties or potential opening of new insights. Minimizing them is key to achieve high precision pulsar timing.

➢ **Various sources of timing residuals**
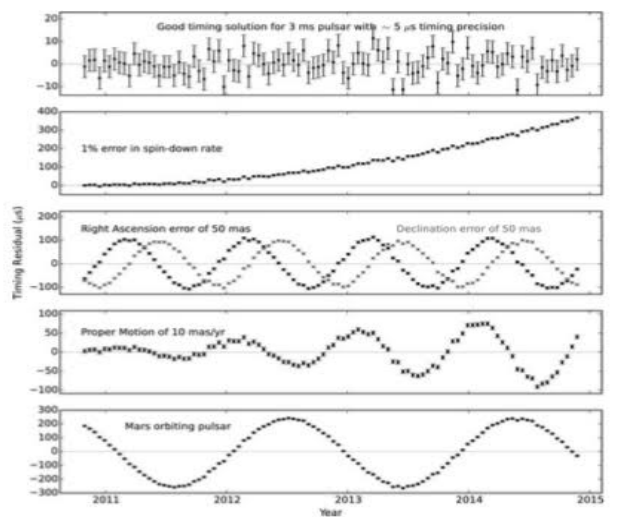
1. *Top Panel*: Timing residuals close to zero indicate a good model fit with minimal uncertainties in TOA predictions—meaning the model accurately represents the pulsar's timing behavior.

2. *Second Panel*: The increasing residuals indicate a 1% error in the spin-down rate. Since the pulsar's rotation slows slightly over time, this error accumulates, causing TOAs to appear increasingly delayed, forming a parabolic curve.

3. *Third Panel*: Incorrect Right Ascension (RA) and Declination (Dec) positions show up as sinusoidal waveforms in the residuals, as Earth's orbit changes the line-of-sight to the pulsar. RA and Dec errors result in two different waveforms due to their directional effects on timing.

4. *Fourth Panel*: Proper motion error of the pulsar—due to assuming an incorrect rate of movement across the sky—creates a mix of linear and sinusoidal residuals. This is because the pulsar appears to "drift" over time, altering the timing model in a more complex way.

5. *Bottom Panel:* The pattern in the Mars-orbiting example is due to an assumed periodic motion of the pulsar around Mars, which would create regular, cyclic timing deviations in the residuals.

➢ **Significance of PTA**

1. Purpose: PTAs detect nanohertz gravitational waves (GWs) by observing correlated timing residuals in millisecond pulsars (MSPs).

2. Key GW Sources: Supermassive black hole binaries (SMBHBs) with masses –, merging after galaxy collisions. Distant systems could form a "stochastic background" of GWs.

3. Detection Mechanism: GWs introduce spatially and temporally correlated systematic in MSP timing residuals. Achieving tens-of-nanosecond precision enables detection at nanohertz frequencies.

4. Global Collaboration: Three PTA experiments—NANOGrav (North America), Parkes PTA (Australia), and European PTA—collaborate as the International Pulsar Timing Array (IPTA).

5. Scientific Progress: Current PTA results challenge models of galaxy mergers and signal significant advancements in sensitivity.

6. Future Outlook: With ongoing improvements in timing precision and new MSP discoveries, direct GW detection via PTAs is increasingly possible, potentially within the next five years**.**

(05/11/2024-12/11/2024)
1. <u>Task</u>: Understand pulsar parameters: pick up some physically-important parameters. Make plots within the website: various combination of parameters and find correlation between parameters. Download pulsar data from the catalog site. Make plots with programming: reproduce the plots.

- **Database**: Australia Telescope National Facility (ATNF) Pulsar catalogue.
  This is a comprehensive database of all published pulsars. Below is the page of the website.
- **Predefined selected pulsar parameters**:
  From the catalogue database we selected important parameters required for our project.
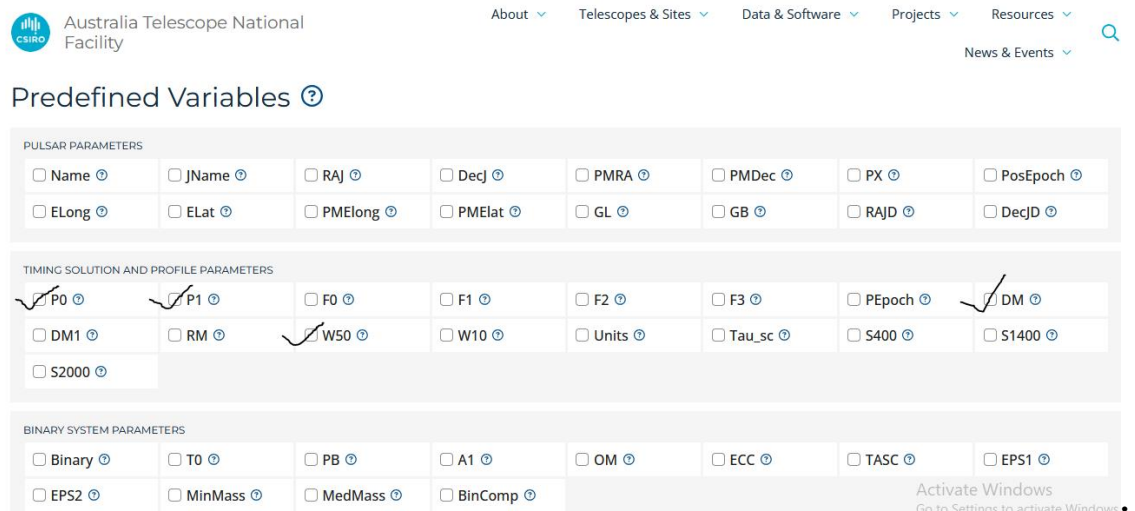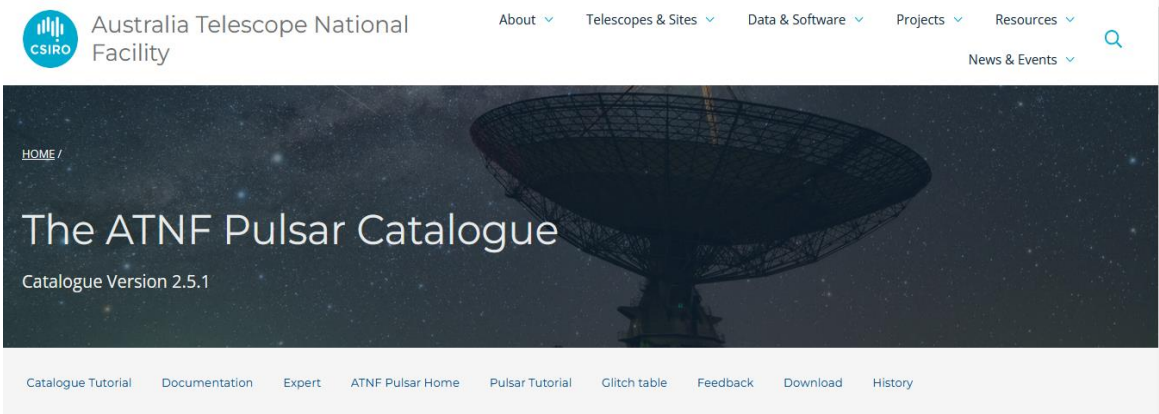
1. **P0** (*period of the pulsar*) -
   pulsar period is the time taken by pulsar to complete one full rotation, typically measured in seconds or milliseconds.

2. **P1** (*Time derivative of period*) –
   It's a dimensionless parameter that indicates how the period of the pulsar is changing over time.

3. **DM** (*Dispersive Measure*)–
   It is the quantity that represents the integrated column density of free electrons along the line of sight from pulsar to us is called DM. It is in pc per cubic centimeter.

4. **W50** (*width of pulse at 50% of peak or pulse width*) –
   It's measured in milliseconds (ms) and provides an indication of the pulse's duration, pulse width.

- **Generated all the plots of different combinations within the website. By selecting parameters as shown in the below fig.**

Plot-1   **P0 versus P1**
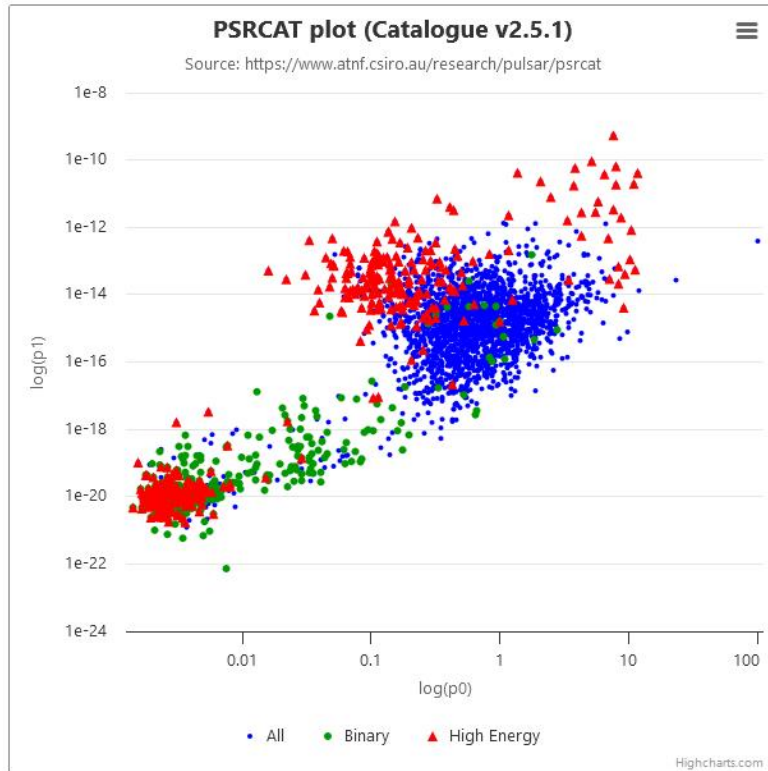
Plot-2   **P1 versus DM**
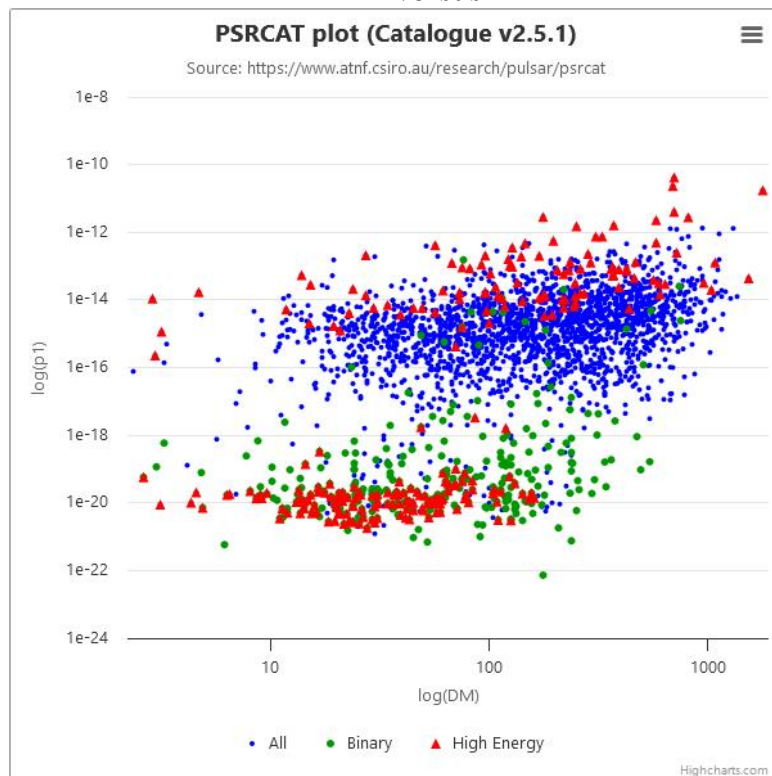
Plot-3   **P1 versus W50**
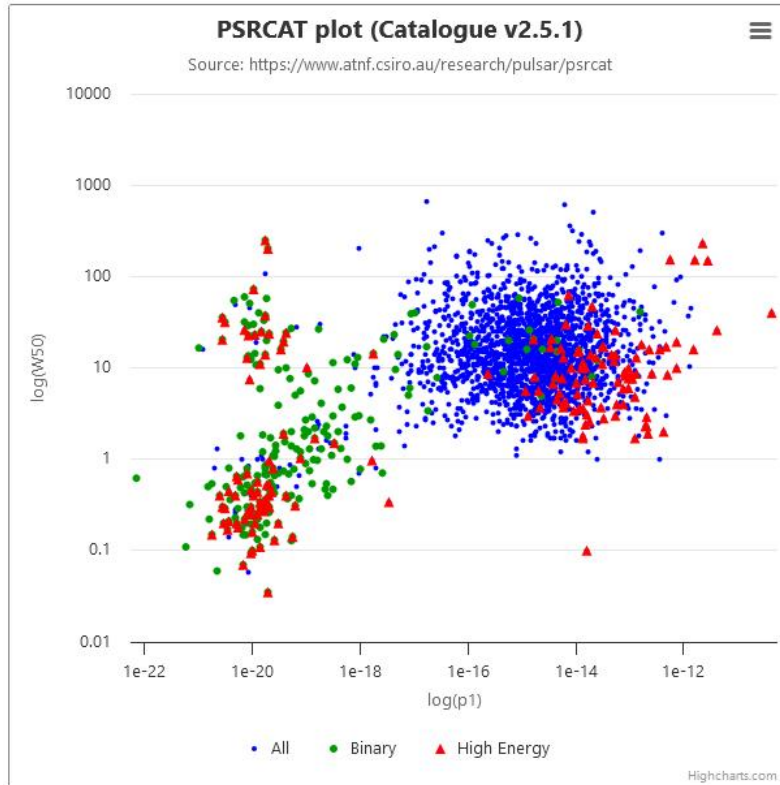
Plot-4   **W50 versus P0**
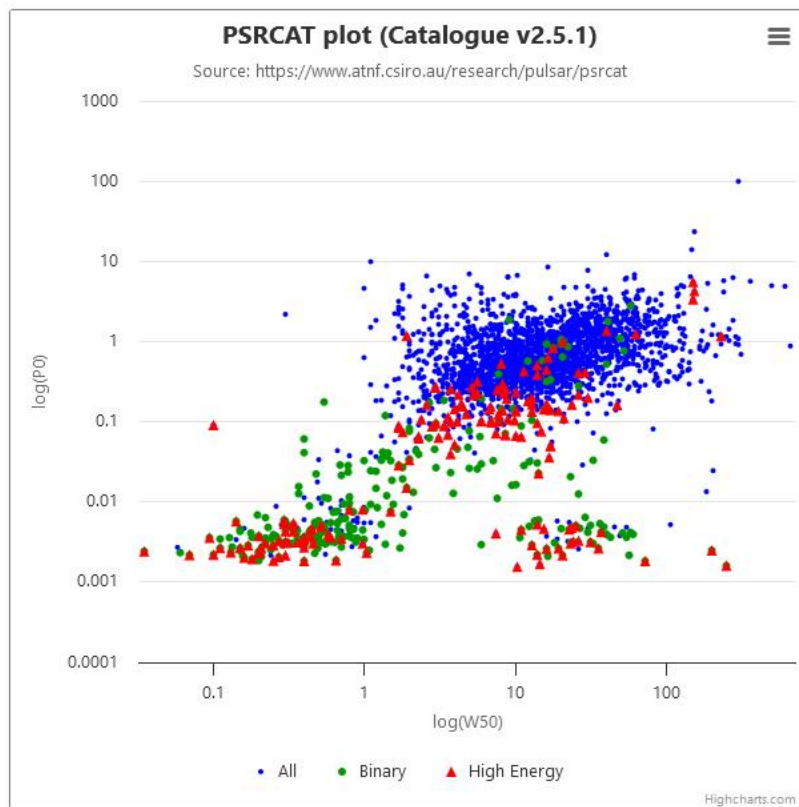
- **Reproduced the plots using python programming.**

Plot 1-   log(P0) versus log(P1)

Plot-2   log(DM) versus log(P1)

Plot-3   log(P1) versus log(W50)

Scatter Plot of P1 vs. W50

Plot 4 -   log(W50) versus log(P0)
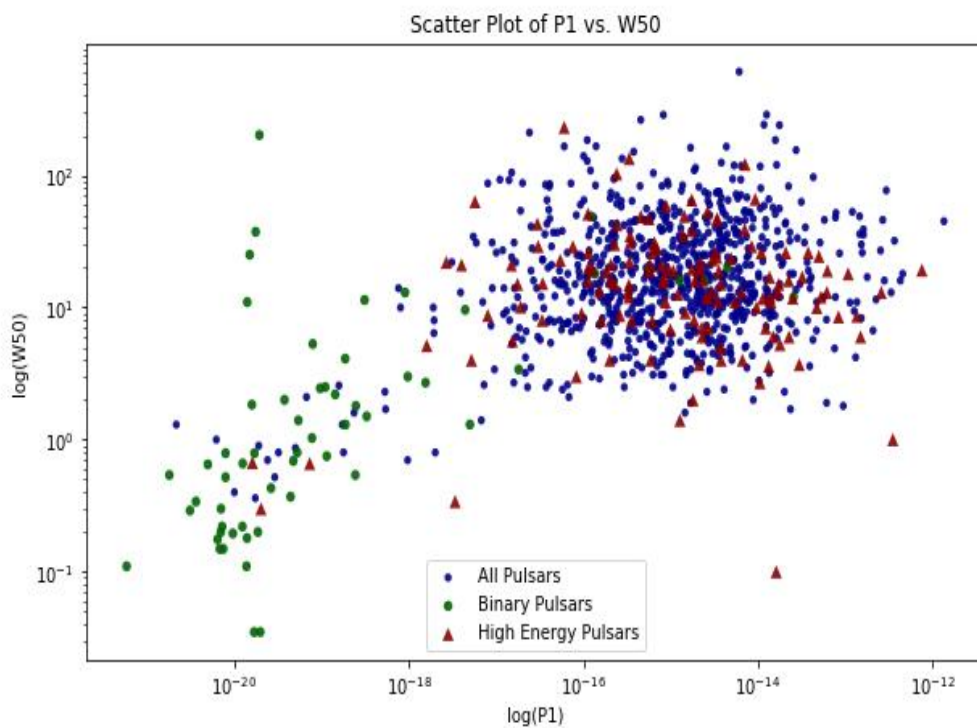
Scatter Plot of W50 vs. P0.
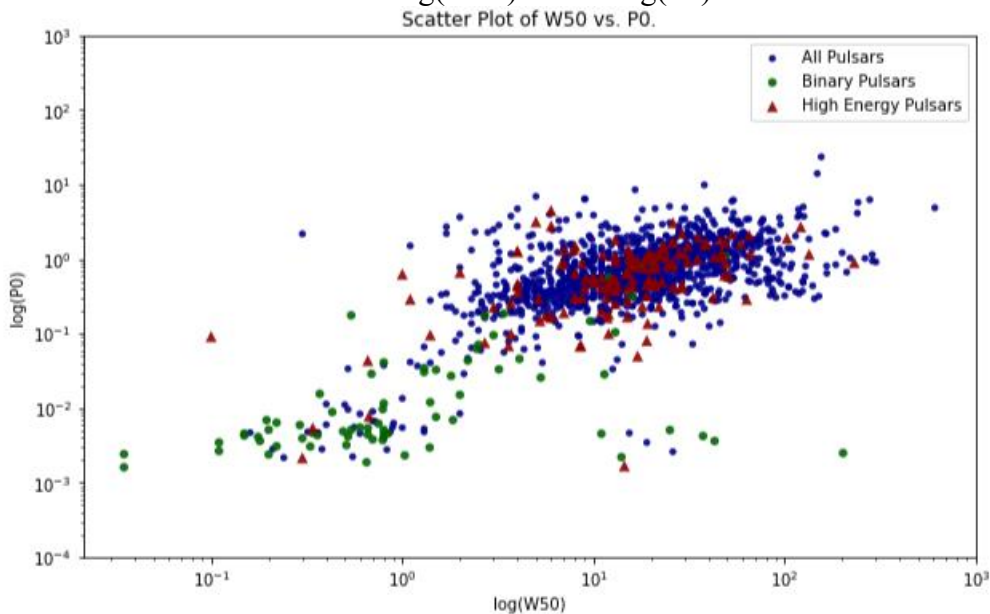
【12/11/24】to 【26/11/24】.

Tasks:Learn machine learning methods for clustering to classify objects according to the parameter values

<u>Application of Machine Learning for Pulsar population</u>

Introduction to Machine Learning(ML):
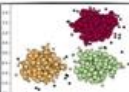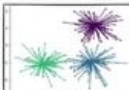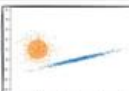
In today's data-driven world, every aspect of our lives is digitally recorded and interconnected. Effective data management tools and techniques are essential for extracting valuable insights and knowledge promptly. Artificial Intelligence (AI), particularly Machine Learning (ML), has significantly advanced in recent years, enhancing data analysis and enabling intelligent application functionality.

Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed. By analyzing patterns in data, ML algorithms can make predictions, identify trends, and drive decision-making processes in various applications.

Project focused learning:

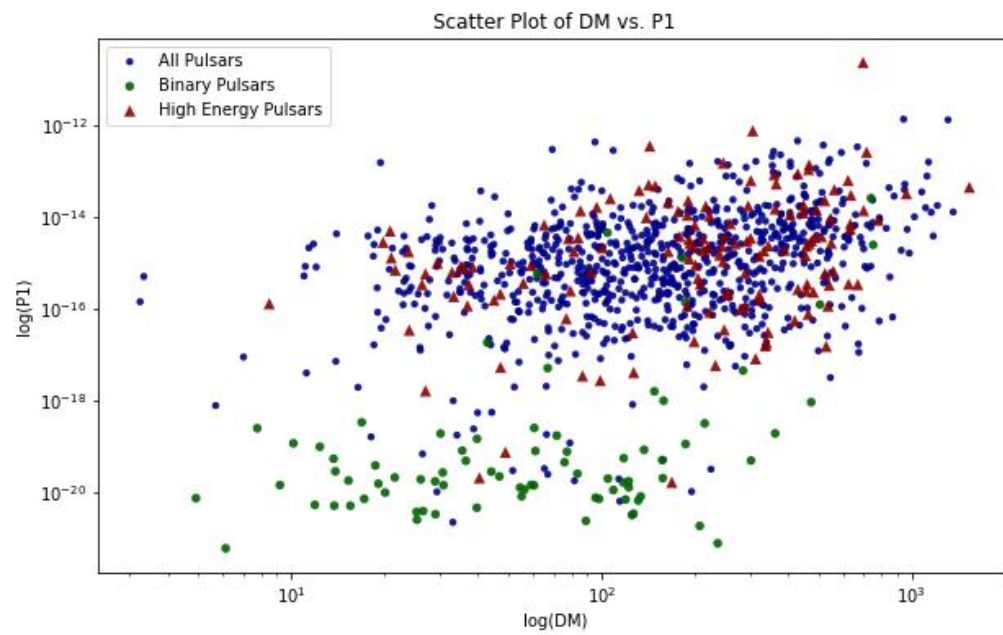Our project goal is to perform clustering, which requires approaching the data using unsupervised learning techniques. Our data type is numerical and non-uniformly distributed density of data points. So appropriate would be DBSCAN clustering. But we shall compare it with other clustering technique to confirm the appropriate method.

- Unsupervised learning: *Clustering method*
  1.Unsupervised learning involves training algorithms on data without predefined labels. The goal is to identify patterns and group similar data points together based on their characteristics.

Clustering Algorithm Type		Clustering Methodology	Algorithm(s)
	Centroid-based	Cluster points based on proximity to centroid	KMeans KMeans++ KMedoids
	Connectivity-based	Cluster points based on proximity between clusters	Hierarchical Clustering (Agglomerative and Divisive)
	Density-based	Cluster points based on their density instead of proximity	DBSCAN OPTICS HDBSCAN
	Graph-based	Cluster points based on graph distance	Affinitiy Propagation Spectral Clustering
	Distribution-based	Cluster points based on their likelihood of belonging to the same distribution.	Gaussian Mixture Models (GMMs)
	Compression-based	Transform data to a lower dimensional space and then perform clustering	BIRCH

➢ *Appropriate Clustering technique*

2. **DBSCAN Clustering**
(Density-Based Spatial Clustering of Applications with Noise)

It is a popular density-based clustering algorithm widely used in data mining and machine learning. It separates high-density clusters from low-density noise, making it robust to outliers and able to find clusters of various shapes and sizes without needing to specify the number of clusters in advance. Unlike K-Means, which is faster, DBSCAN excels at finding high-density regions and handling noise in the data.

1. Choose hyper- parameters:
Epsilon (ε)**:** Defines the radius of the neighborhood around a point.
Minpoints**:** Minimum number of points required to form a dense region.

2. Classify Points:
Core Points**:** Points with at least Minpoints within ε.
Border Points: Points within ε of a core point but with fewer than Minpoints within ε.
Noise Points**:** Points that are neither core nor border points.
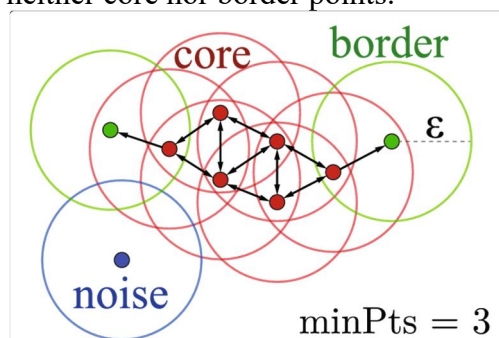
3. Form Clusters:
For each core point, form a cluster including all its reachable points (directly or indirectly within ε).Border points are assigned to the nearest core point's cluster.Noise points remain unassigned.

4. Repeat:
Continue until all points are either assigned to a cluster or marked as noise.

- Key points of DBSCAN clustering:
  1. Assumption**:** Does not assume any specific shape for clusters.
  2. Method**:** Density-based, identifies clusters based on the density of data points.
  3. Parameter**:** Uses epsilon (ε) and minimum points (minPts) to define clusters.
  4. Outliers**:** Can handle outliers by marking them as noise.
  5. Best for**:** Clusters of arbitrary shapes and datasets with noise.

- ## **Cluster validation:**

  1. Silhouette Score*:*
    The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. It is a measure of cluster cohesion and separation.

Calculation: For each point:

- Cohesion (a): Calculate the average distance between the point and all other points in the same cluster.
- Separation (b): Calculate the average distance between the point and all points in the nearest neighboring cluster.
- Silhouette Coefficient (s): For each point,

$$s = \frac{b - a}{\max(a, b)}$$

Where, a: The mean distance between a point and all other points in the same cluster (intra-cluster distance).

b: The mean distance between a point and all points in the nearest cluster (inter-cluster distance).

The Silhouette Score for the entire dataset is the mean of the Silhouette Coefficients of all points.

- High Silhouette Score:
  1. Indicates that data points are well-clustered.
  2. Data points are closer to their own cluster center and far from other clusters.
  3. Values close to **1** suggest well-separated clusters.

- Low or Negative Silhouette Score:
  1. Indicates that data points might be assigned to the wrong clusters.
  2. Values close to **0** suggest overlapping clusters.
  3. Negative values indicate points may be in the wrong cluster.

2. Davies-Bouldin Index (DBI):

The Davies-Bouldin Index is a metric that evaluates the average similarity ratio of each cluster with its most similar cluster. It measures cluster dispersion and separation.

Calculation: For each cluster:

- Cluster Dispersion (S): Calculate the average distance between each point in the cluster and cluster centroid.
- Cluster Similarity (Rij): For each pair of clusters (i, j), calculate the similarity

$$\mathbf{R_{ij}} = \frac{\mathbf{S_i + S_j}}{\mathbf{D_{ij}}}$$

Where,
S_i: Cluster dispersion of cluster i.
S_j: Cluster dispersion of cluster j.
D_{ij}: Distance between the centroids of clusters i and j.

The DB index is average of the maximum R values for each.

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} R_{ij}$$

where, k: Total number of clusters and R_{ij}: Cluster similarity between clusters i and j.

- Range: The DBI values range from 0 to ∞.
- Lower values: Indicate better clustering, with more distinct and compact clusters.

- Higher values**:** Indicate poorer clustering, with less distinct and more dispersed clusters.

These metrics help assess the quality of clustering results by evaluating the cohesion and separation of clusters.

【26/11/24】to【17/12/24】

Tasks: Apply clustering methods to pulsar population data: find groups in pulsar population and make a physical interpretation if possible

I applied DBSCAN clustering technique to our dataset. Our dataset is of numerical type with non-uniform distribution of data points. Hence DBSCAN is a density based method which is applied for such kind of data, also it will handles the noise/outliers which doesn't belong to any cluster. On the other hand GMM is a probabilistic likelihood based clustering which will also handles the outliers but it won't differentiate those outliers as separate cluster as in the case of DBSCAN. I applied silhouette score and Davies Bouldin score to validate the models which best suitable for our dataset. The reason for not choosing K-Means clustering is, it assumes the dataset to be in spherical shape which is not the characteristics of our dataset and Hierarchial clustering is computational expensive.

- **DBSCAN algorithm workflow:**

  1. Import libraries and load the data file.
  2. Preprocessing the data by dropping NAN or missing values.
  3. Encoding & extracting the features of raw data without scaling. Also generate k-plot to choose epsilon value. Fit the DBSCAN model.
  4. Apply DBSCAN clustering using parameters eps & min_points.
  5. Identify and print noise data points.
  6. Mapping colors and shapes for encoded categories.
  7. Plot the results with log scale on both axes.
  8. Display the 2D scatter plot.
  9. Validation process using Silhouette and Davies-Bouldin scores. Choose best scores.
  10. Repeat the process fine tuning the parameters epsilon and adjust minimum sample points to get good quality of clustering patterns.

- **DBSCAN K-plots:**
  k-distance plots represents the sorted data points by their distance to the k-th nearest neighbor.

  What the k-Distance Plot Shows?
  X-axis: This is just the order of your data points. Think of it as a list of points from left to right.
  Y-axis: This shows how far away each data point's 3rd or 4th or 5th nearest neighbor is (or whichever "k" we can chose).

  Why It's Useful?
  The plot helps you find the best "eps" (epsilon) value for DBSCAN clustering.

  How to Use It?
  Look for the 'Elbow': Find the point on the plot where the curve makes a sharp turn. This is where the distance to the 3rd nearest neighbor starts to increase quickly.
  Choose Eps: The Y-value at this elbow is your "eps".

- **Generated 2D K-plots and their epsilon values-**

Fig-1

K-Distance Graph for log(P0) vs. log(P1)

eps = 0.5198

Fig-2

K-distance plot for log(DM) versus log(P1)
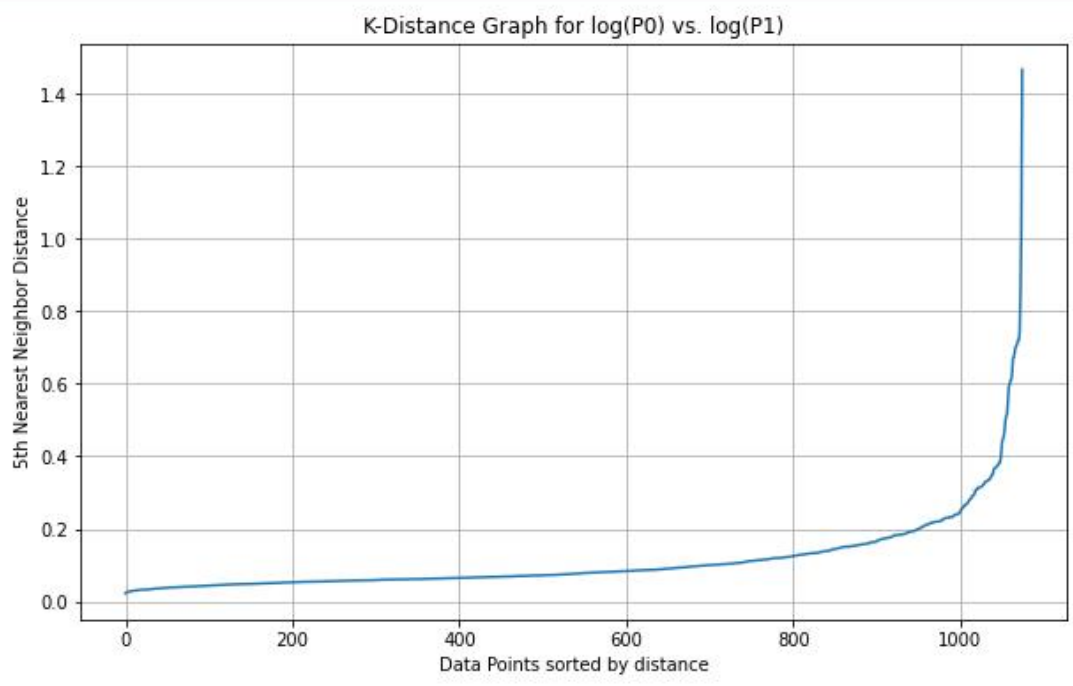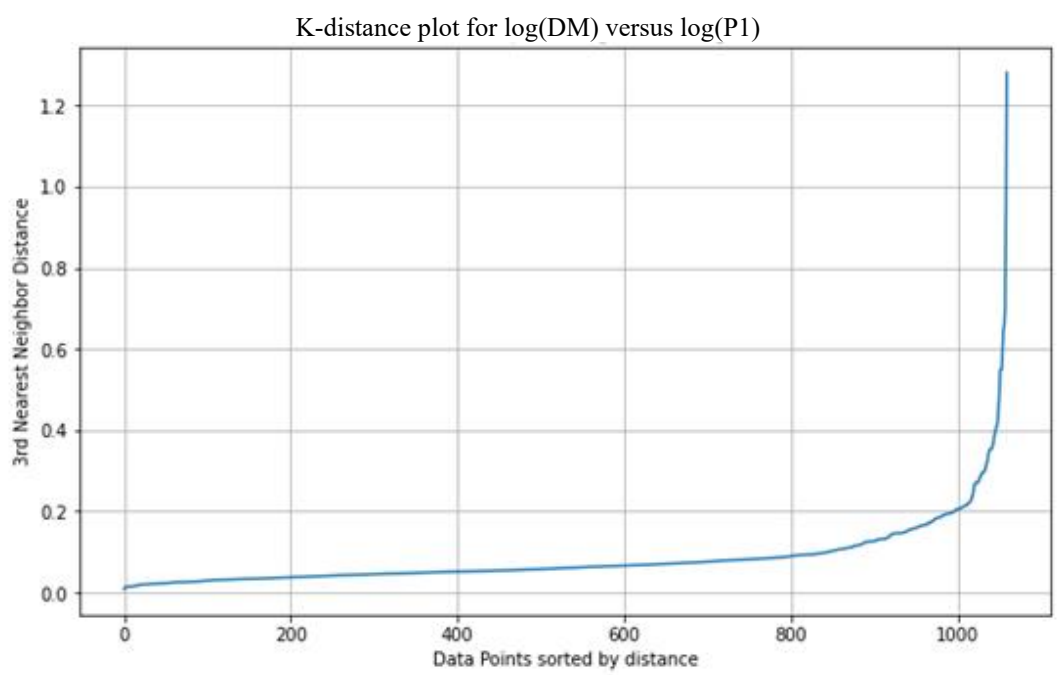
eps = 0.426

Fig-3

K-Distance Graph for log(P1) vs. log(W50)
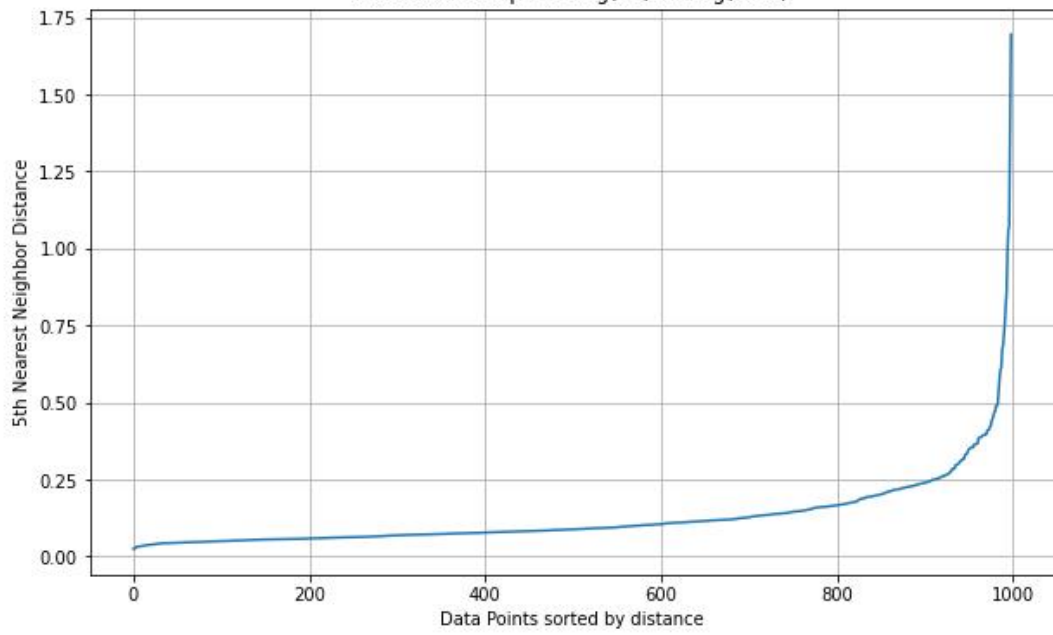
eps = 0.438867

Fig-4

K-Distance Graph for log(W50) vs. log(P0)
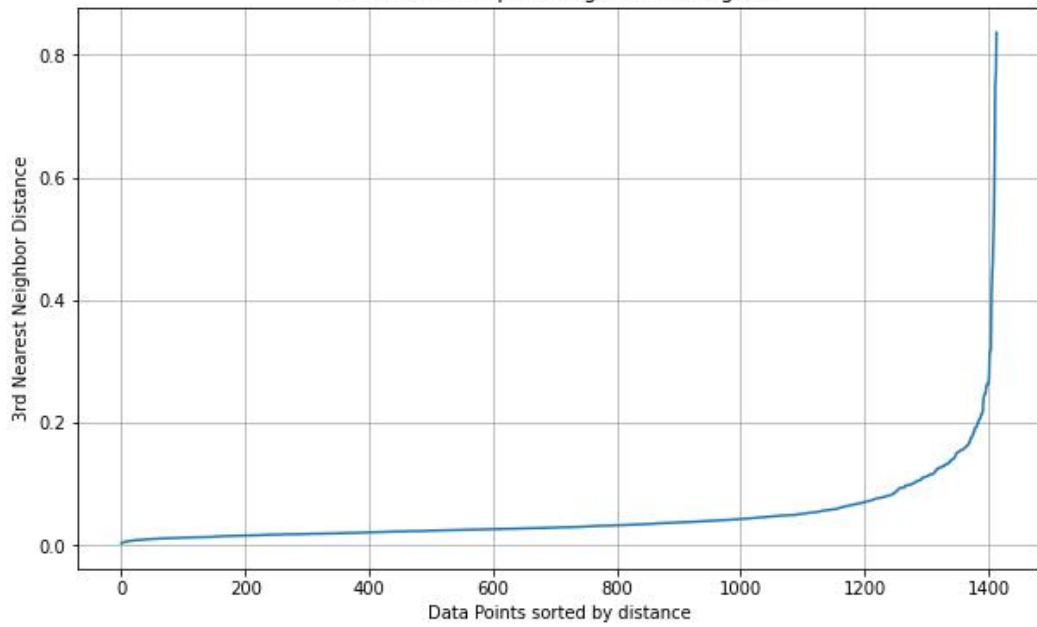
eps = 0.32

- **Generated DBSCAN 2D scatter plots**.

Fig-1

noise points: 12
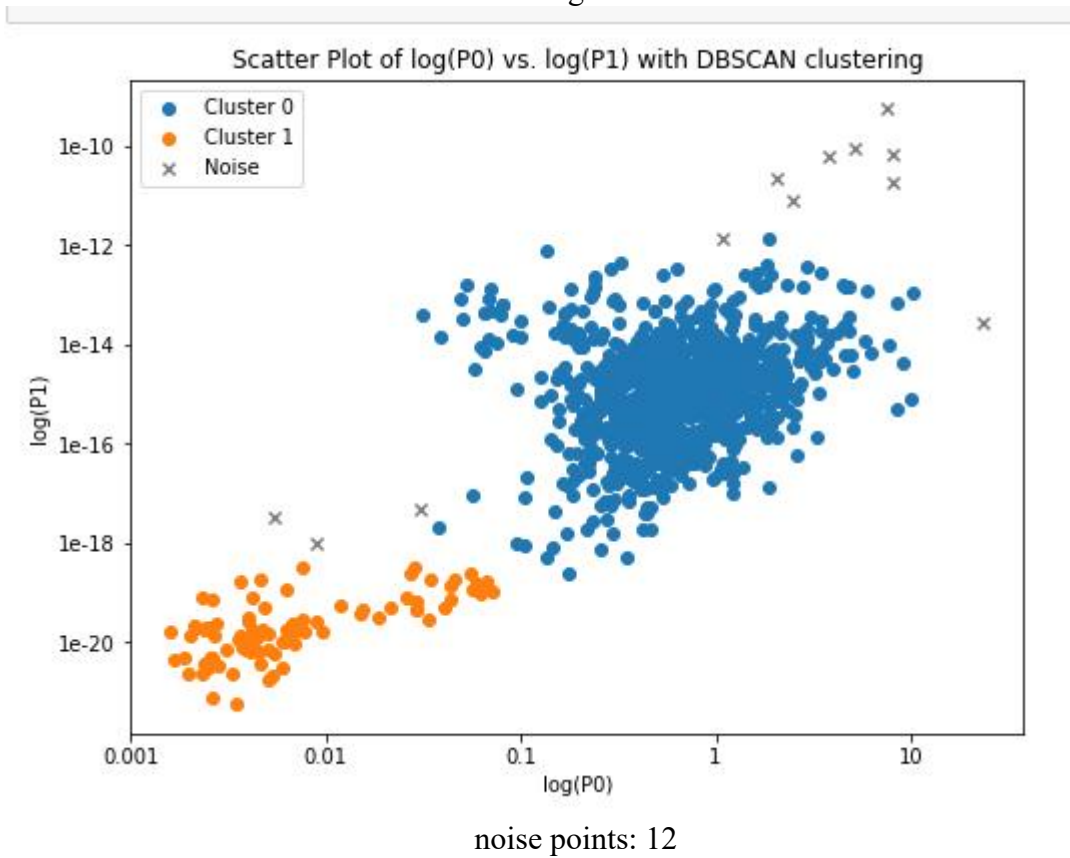
Plot analysis:
- **Cluster-1**
  1. Ranges between 10^{-3} to 10^{-1} in log(p0) and P1 approx from 10^{-22} to 10^{-18}..
  2. These are likely millisecond pulsars.
  3. Having short spin periods and low period derivatives due to binary spun-up.Due to less period Derivative they remain stable.
- **Cluster-0**
  1. Ranges between 10^{-1} to 10^{1} in log(p0) and P1 approx from 10^{-18} to 10^{-12}. .
  2. These are likely young or old with wide diversity of pulsars population distribution.
  3. Having moderate to large spin periods and includes high spin down rates due to high energy loss due to presence of high energy pulsars (magnetars) in this cluster.
  4. They have strong magnetic field.
- **Noise**: 12 noise points which doesn't belong to any cluster.They might possess unique properties.
- **Cluster validation**
  Silhouette score: 0.73111 and Davies-Bouldin index: 0.3114
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

Fig-2

Scatter Plot of log(DM) vs. log(P1) with DBSCAN clustering
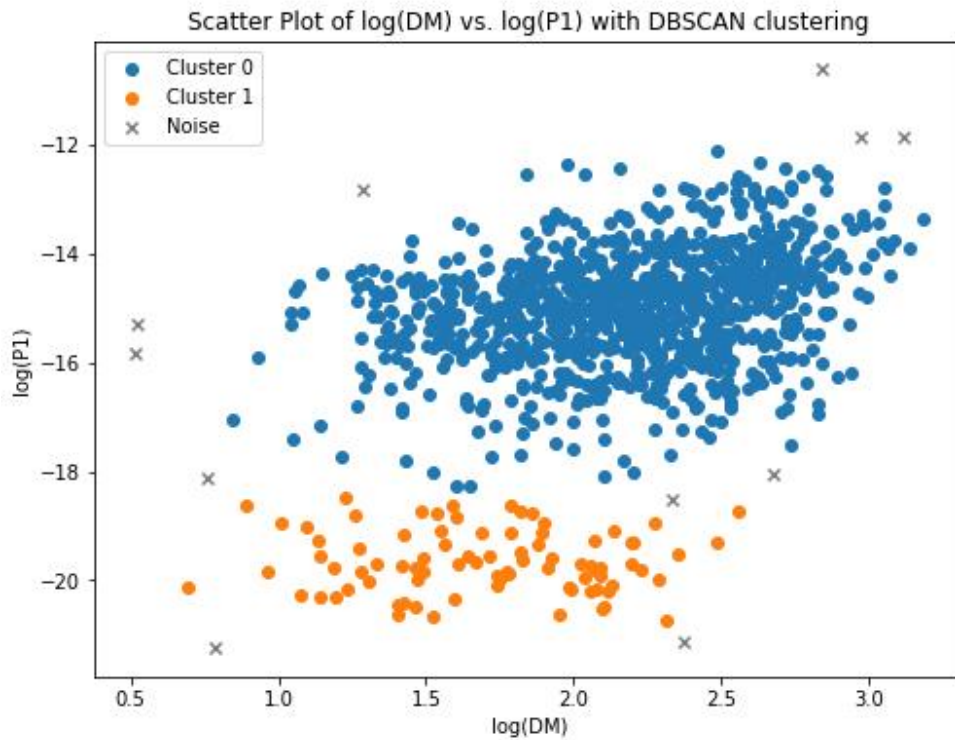
noise points: 11,.

Plot analysis:

- **Cluster-1**
  1. Ranges between 10^1 to approx 200 in log(DM) and log(P1) approx from 10^{-22} to 10^{-18}..
  2. These are likely millisecond pulsars.
  3. Lesser DM values indicate that these pulsars are located near to earth and in lower ISM Density regions and lesser period derivative indicates they are stable candidates.
- **Cluster-0**
  1. Ranges between 10^1 to 10^3 in log(DM) and log(P1) approx from 10^{-18} to 10^{-12}. .
  2. These are likely young or old with wide diversity of pulsars population distribution.
  3. Having moderate to large spin periods and includes high spin down rates due to high energy loss due to presence of high energy pulsars (magnetars) in this cluster.
  4. DM values are ranging from moderate to high indicates most pulsars are located in moderate High ISM density regions.
- **Noise**: 11 noise points which doesn't belong to any cluster.They might possess unique properties.
- **Cluster validation**
  Silhouette score: 0.6993044, Davies-Bouldin score: 0.33600516
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

Fig-3

Scatter Plot of log(P1) vs. log(W50) with DBSCAN clustering
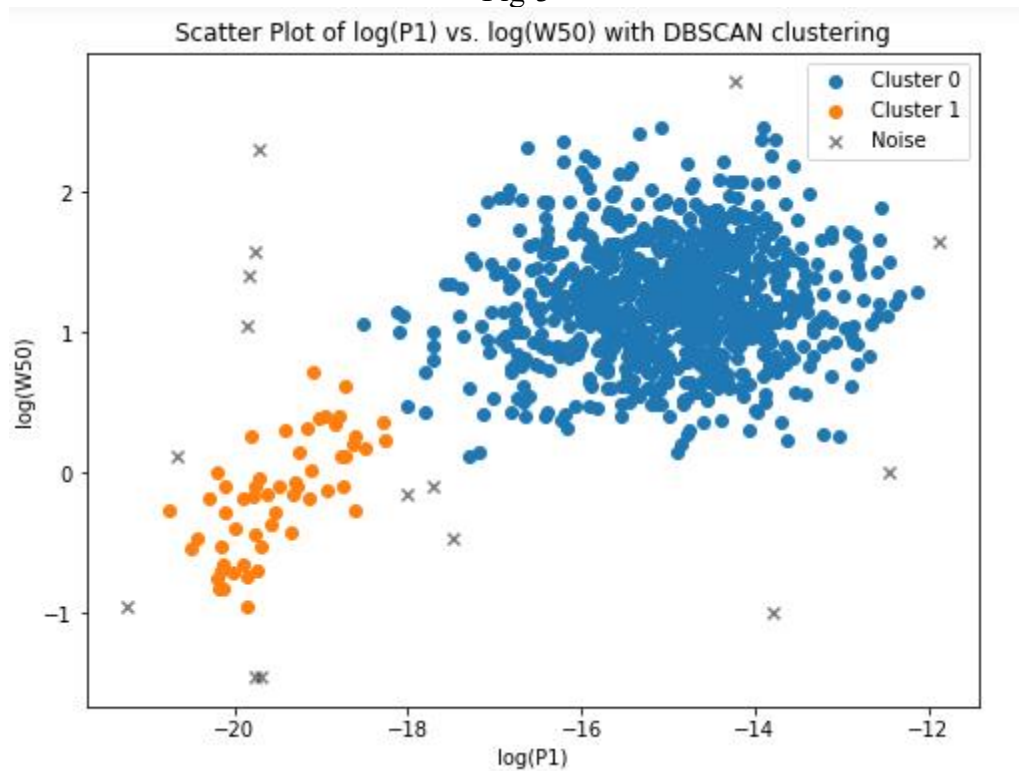
noise points: 15

Plot analysis:

- **Cluster-1**
  1. Ranges between 10^{-1} to approx 10^{+1} in log(W50) and log(P1) approx from 10^{-22} to 10^{-18}..
  2. An extremely small period derivatives tells that these are likely millisecond pulsars
  3. Lesser W50 values indicate that these pulsars experience less pulse broadening effect in lesser ISM density regions or located nearer to earth. And lesser period derivative indicates these pulsars are stable.

- **Cluster-0**
  1. Ranges between 10^{+1} to approx 10^{+2} in log(W50) and log(P1) approx from 10^{-18} to 10^{-12}..
  2. These are likely young or old with wide diversity of pulsars population distribution.
  3. Moderate to high period derivative and high W50 values indicate high pulse broadening.
  4. This tells they are located in high dense regions of ISM or faraway where electron density is more.

- **Noise**: 15 noise points which doesn't belong to any cluster.They might possess unique properties.

- **Cluster validation**
  Silhouette score: 0.7036218, Davies-Bouldin score: 0.337638567
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

Fig-4

Scatter Plot of log(W50) vs. log(P0) with DBSCAN Clustering

noise points: 9
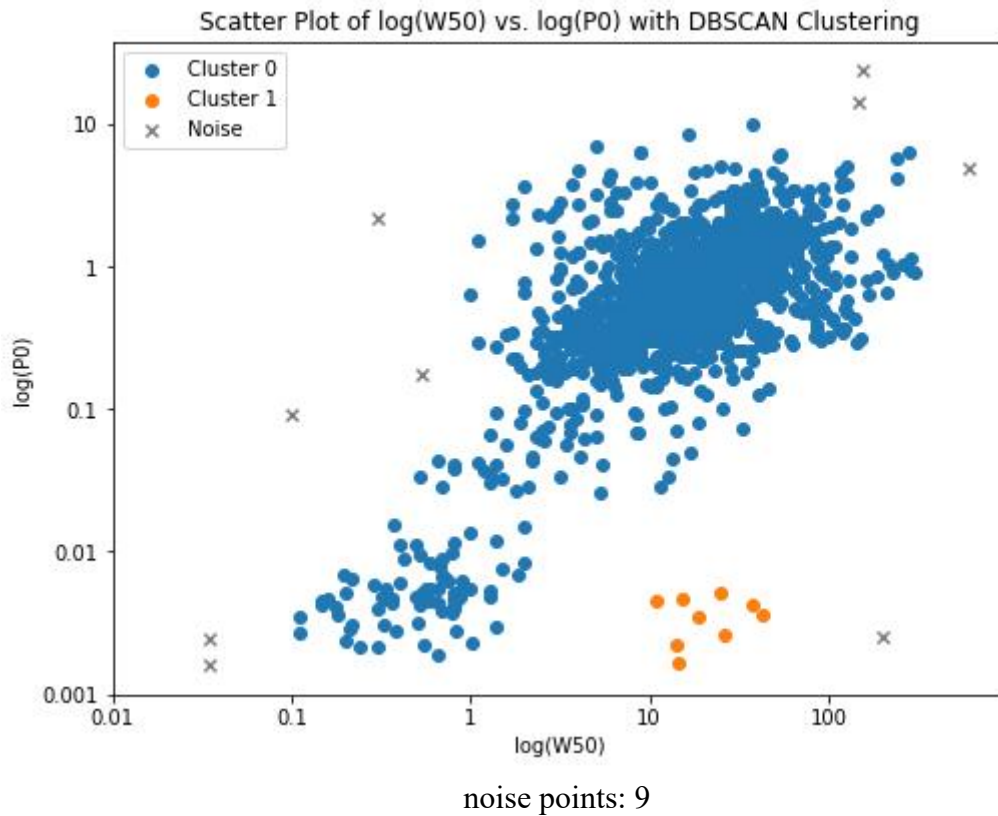
Plot analysis:

- **Cluster-0**
  1. Ranges between 10^{+1} to approx 200 in log(W50) and log(P0) approx from 10^{-3} to 10^{+1}..
  2. These are likely young or old(MSPs) with wide diversity of pulsars population distribution.
  3. Moderate to high period derivative and high W50 values indicate less to high pulse broadening.
  4. This tells the pulsar population shows that they are located in less dense region of ISM to high dense regions of ISM or faraway where electron density is more.
- **Cluster-1**
  1. High W50 value and low P0 values indicates factors like beam geometry,emission altitude Or magnetic inclination dominate over spin period in shaping pulse width.
  2. High pulse broadening at small P0 suggests that pulsar's emission beam is wide due to large Magnetic inclination or emissions originating from higher altitudes in magnetospheres.
- **Noise**: 9 noise points which doesn't belong to any cluster.They might possess unique properties.
- **Cluster validation**
  Silhouette score:0.6728116, Davies_Bouldin_index:0.38676487
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

 - **Generated 3D K-plots and their epsilon values**

Fig-1
3D k-plot for log(P0,),log(P1),log(DM)
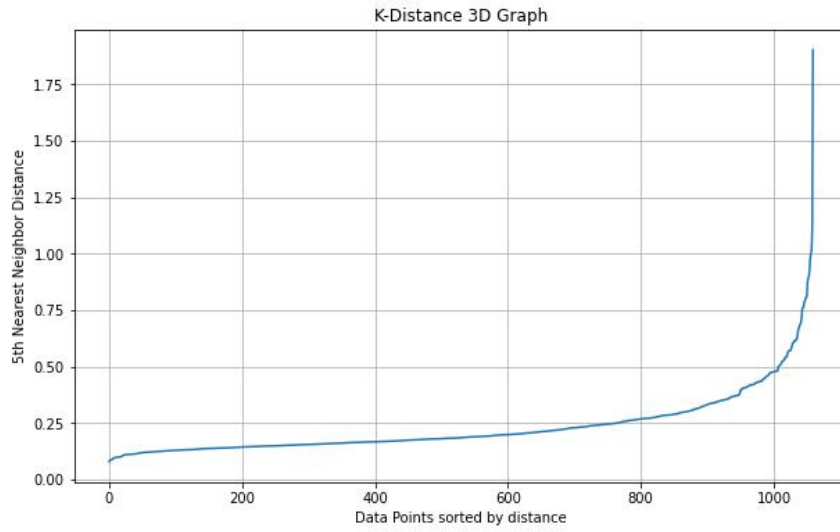
eps = 0.607

Fig-2
3D k-plot for log(P0,),log(P1),log(W50)

eps = 0.54

- **Generated DBSCAN 3D scatter plots.**

Fig-1
3D scatter plot of log(P0),log(P1),log(DM)
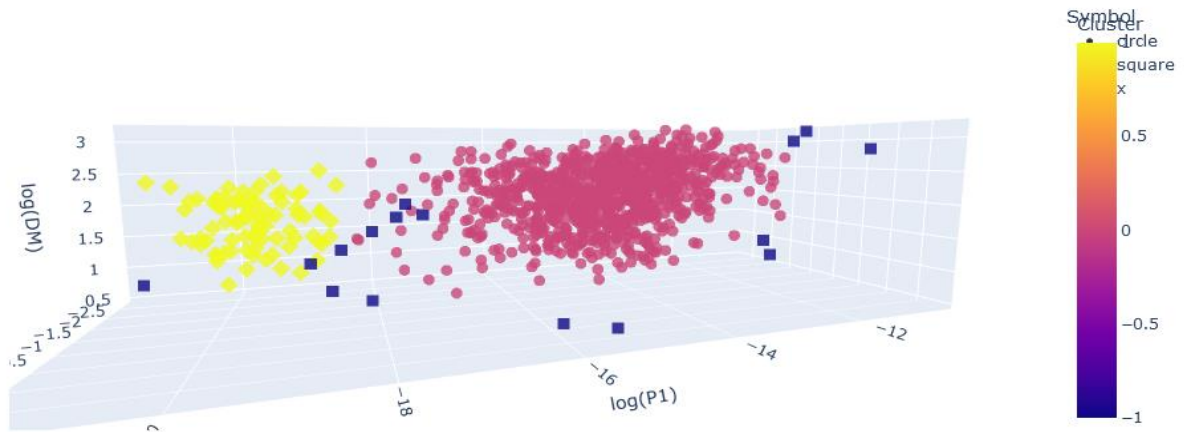
noise points: 16, Silhouette score: 0.710455, Davies-Bouldin score: 0.348528950

Plot analysis:
1. There are two clusters 0 and 1. Cluster 0 consisting pink color circle data points and cluster 1 has yellow square data points.
2. The values log(P0),log(P1),log(DM) of cluster 1 are smaller than the cluster 0. This indicates short period, low period derivatives and lesser DM values.
3. They are located in lesser inter stellar medium denser regions or nearer to earth.
3. Likely the cluster 1 has millisecond pulsars which are very stable and suitable for research in the gravitational waves.
4. We observe an increasing trend in the values of log(P0),log(P1),log(DM).such as moderate to high values of cluster 0 pulsar data points.
5. The increase in the parameter values suggest that there are younger and older diverse pulsar population. As we notice the cluster 0 data points towards right,it would suggest that pulsars which are high energetic with strong magnetic field are located far away and high dense inter stellar medium regions. Also these are less stable due to variations in the period derivative.

- **Noise**: 16 noise points which doesn't belong to any cluster.They might possess unique properties.

- **Cluster validation:**
  Silhouette score: 0.710455, Davies-Bouldin score: 0.348528950.
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
 3. Low Davies score represents, inter-cluster separation is high.

Fig-2
3D scatter plot of log(P0),log(P1),log(W50)
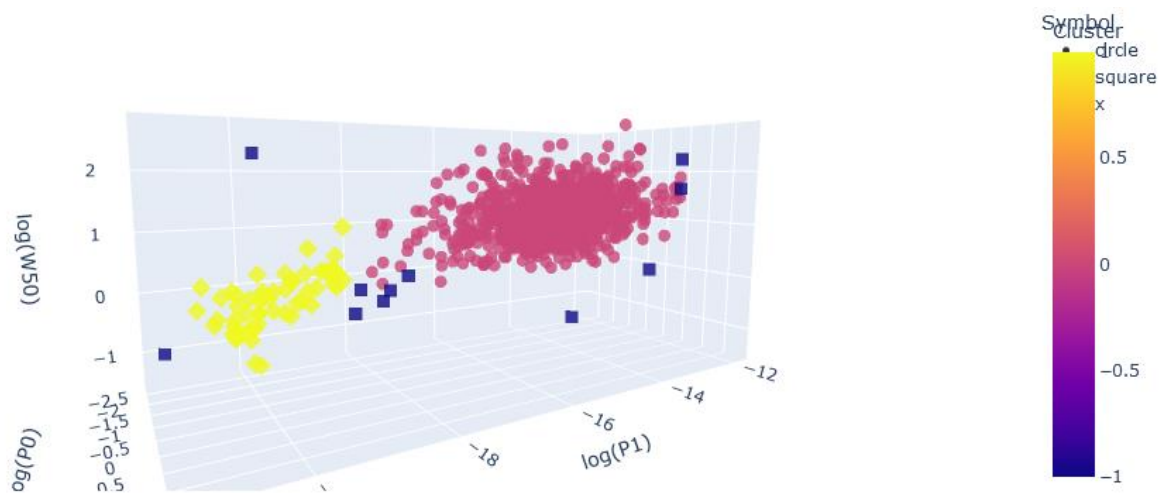
noise points: 11

Plot analysis:
1. There are two clusters 0 and 1. Cluster 0 consisting pink color circle data points and cluster 1 has yellow square data points.
2. The values log(P0),log(P1),log(W50) of cluster 1 are smaller than the cluster 0. This indicates short period, low period derivatives and lesser W50 values.
3. They are located in lesser inter stellar medium denser regions or nearer to earth. Where the pulse broadening effect is less. Hence, we get sharp pulse profile.
3. Likely the cluster 1 has millisecond pulsars which are very stable and suitable for research in the gravitational waves and testing relativity experiments.
4. We observe an increasing trend in the values of log(P0),log(P1),log(W50).such as moderate to high values of cluster 0 pulsar data points.
5. The increase in the parameter values suggest that there are younger and older diverse pulsar population. As we notice the cluster 0 data points towards right would suggest that pulsars are high energetic with strong magnetic field are located far away and at high dense inter stellar medium regions.
6. Due to this denser regions, there would be pulse broadening effect on these pulse signals. It would become difficult to receive narrow sharp pulse profile.

- **Noise**: 11 noise points which doesn't belong to any cluster.They might possess unique properties.

- **Cluster validation:**
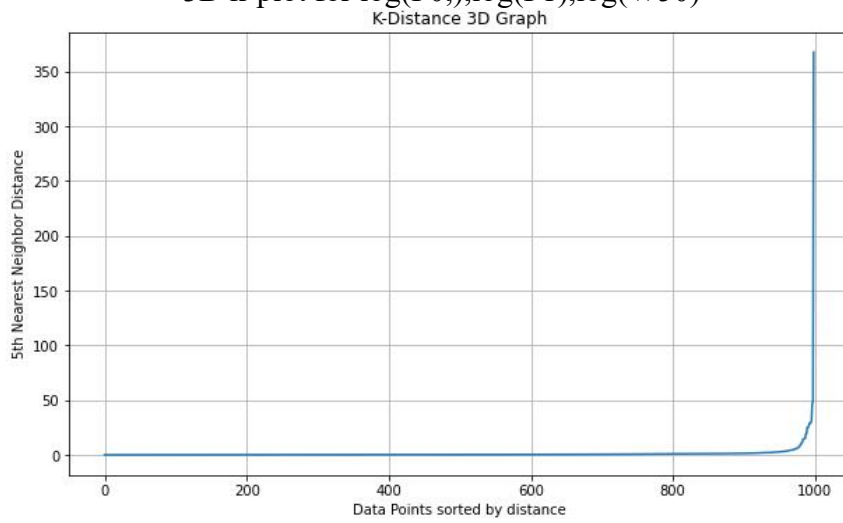 Silhouette score: 0.68693808, Davies-Bouldin score: 0.55070258.
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

- **Generated 4D K-plot and their epsilon value.**

K-plot For log(P0,),log(P1),log(DM),log(W50)
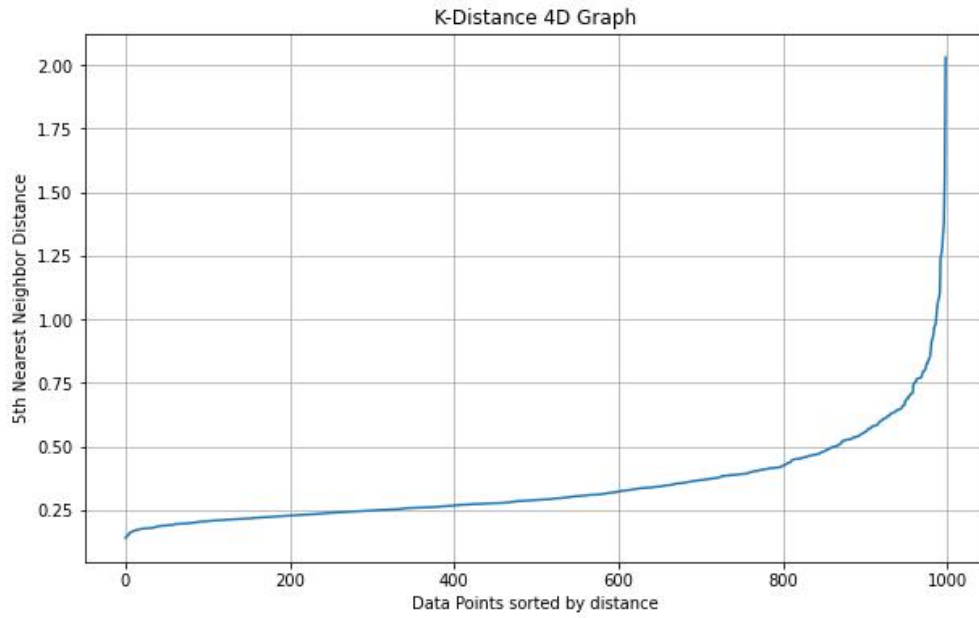
eps = 0.699

- **Generated DBSCAN 4D scatter plots**.

Fig-1
4D scatter plot of log(P0),log(P1),log(DM),log(W50)
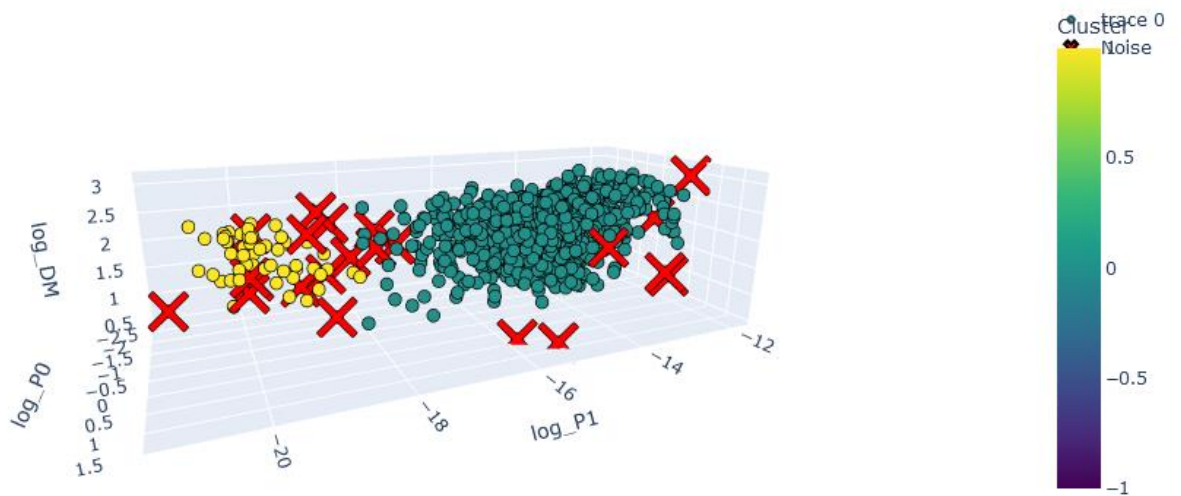
noise points: 23

Plot analysis:

1. There are two clusters 0 and 1. Cluster 0 consisting blue color circle data points and cluster 1 has yellow circle data points.
2. The values log(P0),log(P1),log(DM),log(W50) of cluster 1 are smaller than the cluster 0. This indicates short period, low period derivatives. And lesser W50,DM values shows they are located near the earth with low density ISM regions.
3. If the distance is less then there would be less electron density which the pulse would be affected by ISM resulting dispersion, pulse broadening. Hence, we get sharp pulse profile.
4. Likely the cluster 1 has millisecond pulsars which are very stable and suitable for research in the gravitational waves and testing relativity experiments.
5. We observe an increasing trend in the values of log(P0),log(P1), log(DM),log(W50).such as moderate to high values of cluster 0 pulsar data points.
5. The increase in the parameter values suggest that there are younger and older diverse pulsar population. As we notice the cluster 0 data points towards right would suggest that pulsars are high energetic with strong magnetic field are located far away and at high dense inter stellar medium regions.
6. This results moderate to high pulse broadening effect on these pulse signals. Where the pulse pass more electron density regions. It would become difficult to receive narrow sharp pulse profile.

- **Noise**: 23 noise points which doesn't belong to any cluster. They might possess unique properties.

- **Cluster validation:**
  Silhouette score: 0.694185769, Davies-Bouldin score: 0.3841697358
  1. The data points are well-matched to their own clusters.
  2. Well-separated from other clusters, suggesting high-quality clustering.
  3. Low Davies score represents, inter-cluster separation is high.

➤ **Future Aspects of P0,P1,W50,DM in pulsar studies.**

- Understanding Evolution: Continuous observations of across a variety of pulsars can help refine pulsar evolutionary models (For example, how pulsars evolve from young, fast-spinning states to slower, older states.)
- Gravitational Wave Detection: Millisecond pulsars (P0~ 1-10 ms) are used in pulsar timing arrays (PTAs) to detect nanohertz-frequency gravitational waves.
- Millisecond Pulsar formation: Insights into accretion-induced spin-up of recycled pulsars in binary systems,providing more evidence about binary star evolution.
- Magnetic Field Evolution: Using P1 to estimate the decay of the magnetic field over time and relate it to the pulsar's age.
- Extreme Physics: Investigating pulsars with extremely high or low values to study their unique mechanisms of energy loss or spin-up.
- ISM Mapping: Pulsars act as probes to map the distribution of free electrons in the Milky Way and other galaxies, helping to build precise models of the interstellar medium (ISM).
- Scattering Studies: Observing how broadens with increasing DM can quantify scattering effects caused by the ISM,leading to improved models of pulse propagation. Also we can detect high redshift pulsars in distant galaxies.
- Pulsar Timing Precision: Narrower pulses improve timing precision, crucial for PTAs and detecting gravitational waves.

➢ **Machine Learning meets Pulsars: The future of Pulsar studies.**

Machine learning (ML) holds immense potential for future pulsar studies due to its ability to handle complex, high-dimensional data and uncover patterns that might not be apparent through traditional methods.

- ML can distinguish real pulsars from artifacts with high accuracy, reducing the burden on manual inspection.

- ML can help model the relationships between intrinsic properties and environmental factors, shedding light on how pulsars evolve in different ISM conditions.

- ML can optimize pulsar selection for PTAs, identifying the most stable pulsars or improving data noise modeling, enhancing the sensitivity of gravitational wave searches.

- Future telescopes like the Square Kilometre Array (SKA) will generate petabytes of data. ML models, especially deep learning algorithms (e.g., CNNs), can automate pulsar candidate detection from radio survey data, filtering noise and improving discovery rates.

➢ **Future research Plans**

Research Plan for PhD – Pulsar Timing Array Analysis with Machine Learning and AI for Gravitational Wave Detection.

➢ **Background**

The detection of gravitational waves (GWs) has revolutionized our understanding of the universe, providing insights into phenomena like black hole mergers and neutron star collisions. Pulsar Timing Arrays (PTAs) are an emerging observational tool to detect low-frequency GWs by monitoring the precise timing of millisecond pulsars. However, systematic errors, including irregularities in pulse arrival times and interstellar medium effects, remain significant challenges.

Recent advancements in machine learning (ML) and artificial intelligence (AI) offer new possibilities for improving data analysis techniques. By leveraging these technologies, it is possible to address systematic errors and enhance the sensitivity of PTAs, enabling the detection of GWs with unprecedented precision.

➢ **Objectives**

The primary goal of this research is to develop novel ML-based methods to analyze pulsar timing data, reduce systematic errors, and enhance the accuracy of GW detection using PTAs. Specifically, the research will focus on:

1. Reduction of Systematic Errors: Identify, model, and mitigate sources of systematic errors in pulsar timing measurements.

2. Analysis of Pulsar Statistical Properties: Explore the intrinsic and extrinsic statistical behaviors of pulsar pulses to enhance timing models.

3. Development of ML Frameworks: Innovate data analysis methods using ML and AI to extract faint GW signals embedded in noisy and complex datasets.

➢ **Research Methodology**

1. Data Collection and Preparation:

- Utilize publicly available PTA datasets, such as ATNF, NANOGrav, EPTA, or PPTA.
- Preprocess the data to remove noise and standardize formats for ML applications.

2. Systematic Error Reduction:

- Investigate and model sources of systematic errors, such as pulse profile evolution, interstellar scattering, and clock inaccuracies.
- Use statistical tools like Bayesian inference and Principal Component Analysis (PCA) to identify error patterns.
- Apply statistical techniques and ML algorithms (e.g., Random Forest, Gradient Boosting) to identify these patterns and isolate & correct these errors.

3. Development of ML Algorithms:

- Supervised Learning: Train models using labeled datasets to detect GW signals and refine timing precision.
- Unsupervised Learning: Apply clustering techniques to identify hidden patterns in pulsar timing residuals.
- Neural Networks: Utilize deep learning models (e.g., Convolutional Neural Networks, Recurrent Neural Networks) to enhance sensitivity to subtle GW signatures.
- Explore transfer learning to adapt models trained on one dataset to another.

4. Validation and Testing:

- Validate the developed methods using synthetic datasets with known GW signals for model performance.
- Test developed algorithms on real PTA datasets to measure improvements in timing precision and GW signal detection capabilities.

➢ **Expected Outcomes**

1. A novel ML-based framework for analyzing pulsar timing data with reduced systematic errors.
2. Enhanced detection capabilities for low-frequency GWs using PTAs.
3.Contribution to the global effort in detecting and characterizing the nanohertz GW background.

➢ **Significance**

This research will address critical challenges in PTA-based GW detection and contribute to advancing astrophysical and cosmological studies. By integrating ML techniques into pulsar timing analysis, the study will pave the way for improved precision and novel discoveries in GW astronomy.

No. 2-7-3	Research of Biomedical Engineering			
Name	Rohmah HIDAYAH	Title	Master Student	
Affiliation	Institut Teknologi Sepuluh Nopember (ITS), Indonesia Email: rohmahhidayah418@gmail.com			
Research Field	Biotechnology & healthcare technology / Data science and AI			
Period of Internship	Online: December 16, 2024 - December 20, 2024 Onsite: January 6, 2025 - February 28, 2025			
Host Professor	Yuta NAKASHIMA	Title		Associate Professor
Affiliation	Faculty of Advanced Science and Technology Email: yuta-nmec@h.kumamoto-u.ac.jp			

RESEARCH OF BIOMEDICAL ENGINEERING
(Final Internship Report – IROAST Young Internship Researcher)

CHAPTER 1: INTRODUCTION

1.1 Background

The IROAST Young Internship Researcher program is designed for Master's students who intend to pursue a PhD at Kumamoto University, providing an opportunity for international collaboration in academic research. Through this program, I worked under the supervision of Prof. Yuta Nakashima in the Bioengineering Laboratory at Kumamoto University, with a focus on integrating computer science into biomedical engineering. As a researcher in computer science specializing in biomedical engineering, my role in this program was to develop artificial intelligence and programming-based solutions to improve efficiency in cancer and non-cancer cell identification and accelerate biomedical experiments. I was responsible for enhancing the computational process for cell analysis by developing algorithms and computational models that make the analysis of cell characteristics more efficient. Additionally, I worked on developing machine learning models using deep learning to classify and predict cancer and non-cancer cells with high accuracy. Beyond computational tasks, I was also involved in the fabrication of cell analysis tools, aiming to improve accuracy in biomedical experiments. I played a key role in monitoring the cell dataset used in research, ensuring proper dataset management, preprocessing, and cleaning to maintain high data quality for predictive modeling. Furthermore, I contributed to training laboratory members in Python programming, assisting them in building automated research systems to enhance efficiency and independence in data analysis. This collaboration serves as a bridge between bioengineering and artificial intelligence, utilizing deep learning technologies for medical image processing. The implementation of these methodologies allows research conducted in the laboratory to be more efficient, accurate, and well-integrated, supporting the development of AI-driven cancer diagnosis systems. Working alongside all laboratory members under Prof. Yuta Nakashima has deepened my interest in this research and highlighted its potential for further development. As a result, after completing this internship, I intend to continue this research as part of my PhD program at Kumamoto University. I hope that my future studies will make a significant contribution to the field of biomedical engineering, particularly in

advancing AI-based systems for medical applications.

1.2 Internship Objectives

This internship aimed to integrate computer science into biomedical engineering by applying artificial intelligence to enhance efficiency in cell analysis, classification, and prediction. Under the supervision of Prof. Yuta Nakashima at the Bioengineering Laboratory, Kumamoto University, I developed deep learning algorithms and models to accelerate cell analysis and improve the accuracy of cancer and non-cancer cell classification. Additionally, I contributed to the fabrication of cell analysis tools to enhance experimental precision and conducted dataset monitoring and preprocessing to ensure optimal data quality for research. Beyond technical contributions, I also trained laboratory members in Python programming to support the development of automated research systems. This internship not only focused on technological advancements in biomedical analysis but also encouraged knowledge exchange and collaboration between computer science and bioengineering, fostering more efficient and accurate research in medical diagnostics.

1.3 Scope of the Internship

The scope of my research in this internship was the application of computer science in biomedical engineering, particularly in developing AI-based methods for cell analysis, classification, and prediction. This research involved biomedical image processing using deep learning, optimizing algorithms to improve cell computation efficiency, and exploring data-driven prediction techniques. Beyond modelling and analysis, this research also included the fabrication of biomedical analysis tools to support laboratory experiments and the management and preprocessing of cell datasets to ensure optimal data quality for research. Additionally, the research scope involved collaboration with laboratory members in developing automation systems and utilizing computational methods to enhance bioengineering research.

CHAPTER 2: ACTIVITIES AND FINDINGS

2.1 Tasks and Activities Performed

In this section, I will detail the steps I undertook during the internship until I was able to successfully calculate cells and classify them using deep learning. Each stage of this process involved a series of steps, starting from understanding the laboratory system, processing cell datasets, developing deep learning models, and implementing the model. All activities were carried out under the guidance of Prof. Yuta Nakashima, following a structured approach to ensure that each aspect of the research progressed effectively. These stages included data preparation, computational model design, and parameter optimization, all of which played a crucial role in advancing research in biomedical engineering.

A. Microfluidic Device Fabrication

Microfluidic device fabrication refers to the design and manufacturing process of miniaturized systems used to control and manipulate fluids in extremely small volumes, typically at the microliter or nanoliter scale. These devices are widely applied in bioengineering research, diagnostics, and lab-on-a-chip technology, allowing precise analysis of biological samples. The fabrication process involves several key stages, including substrate preparation, patterning, casting, and bonding, to create microchannels and functional components that enable controlled fluid flow at the microscopic level.

The primary materials used in fabrication include polydimethylsiloxane (PDMS), glass, and silicon, chosen for their optical transparency, biocompatibility, and ease of molding. By integrating microfluidic technology into bioengineering, researchers can develop advanced platforms for cell analysis, drug testing, disease diagnostics, and organ-on-a-chip systems. This fabrication process requires high precision, as even minor variations in microchannel dimensions or bonding quality can significantly impact the device's performance. In this internship, I focused on microchannel patterning, ensuring high-quality device structures, and optimizing the fabrication process for bioengineering applications. Below are the step by step steps that I did to do the stages of Microfluidic Device Fabrication:

1. Conducted substrate cleaning using an ultrasonic bath with acetone, methanol, and ultrapure water.
2. Processed the substrate with piranha solution for deep cleaning and rinsed with ultrapure water.
3. Performed photoresist coating using SU-8 3050 and a spin-coating method to achieve a 100 μm film thickness.
4. Conducted prebake, exposure, post-exposure bake, and hard baking to develop the

microfluidic patterns.

B. PDMS (Polydimethylsiloxane) Preparation and Casting

PDMS (Polydimethylsiloxane) preparation and casting refers to the process of mixing, degassing, molding, and curing PDMS to create microfluidic structures. PDMS is a silicone-based elastomer widely used in bioengineering and microfluidic device fabrication due to its biocompatibility, flexibility, transparency, and ease of fabrication. This process involves mixing PDMS with a curing agent in a specific ratio, removing air bubbles through degassing, pouring the mixture into a mold, and curing it under controlled temperature conditions to solidify the structure. PDMS casting enables the formation of microchannels and other functional components essential for fluid manipulation in microfluidic systems. Proper preparation ensures a homogeneous and defect-free material, which is crucial for achieving accurate and efficient fluid flow in bioengineering applications. Additionally, PDMS can be bonded with glass or other surfaces using plasma treatment, making it a key material in the development of lab-on-a-chip devices, biosensors, and organ-on-a-chip systems. Below are the stages of making PDMS (Polydimethylsiloxane) Preparation and Casting:

1. Prepared PDMS mixture in a 10:1 ratio with a curing agent.
2. Used the Awa-tori Rentaro super mixer to mix and perform a defoaming process to remove air bubbles.
3. Coated the mold with a fluoropolymer spray to prevent adhesion before casting.
4. Degassed the PDMS mixture in a desiccator to eliminate trapped air before pouring it onto the mold.
5. Cured the PDMS at 70°C for one hour to solidify the structure.

C. Microfluidic Device Assembly

Microfluidic device assembly refers to the process of integrating different components of a microfluidic system, ensuring proper alignment, bonding, and functionality of the device. This process involves removing the fabricated microfluidic structure from the mold, creating access points for fluid flow, bonding the microchannels to a substrate (such as glass or PDMS), and conducting final quality checks to ensure the device operates correctly. Proper assembly is crucial to maintaining leak-free fluid flow, precise microchannel dimensions, and stable bonding between materials. A commonly used technique for bonding is oxygen plasma treatment, which enhances the adhesion between PDMS and glass by activating their surfaces at the molecular level. Other methods include thermal bonding, chemical bonding, and mechanical clamping, depending on the application and material compatibility. The final step in the assembly is the inspection and testing of the device, where microchannels are checked for blockages, leaks, and structural

integrity. Successfully assembled microfluidic devices are widely used in cell analysis, drug testing, and diagnostic applications, providing a powerful tool for biomedical and bioengineering research. below are the stages of making Microfluidic Device Assembly:

1. Carefully removed the cured PDMS from the mold and punched holes in the microchannels.
2. Performed oxygen plasma bonding to attach the PDMS layer to the glass substrate.
3. Ensured proper alignment and bonding to achieve a strong and durable connection.
4. Conducted final inspection and testing of the microfluidic device, verifying the integrity of microchannels and the bonding strength.

D. Developing and Printing 3D Models in the Laboratory

In the laboratory, I am experimenting with creating and printing 3D models as part of research exploration in bioengineering. This process involves digital modeling using 3D software, converting designs into a compatible format, and finally printing them using a 3D printer. By utilizing this technology, we can create prototype devices and microstructures that support various bioengineering experiments. As an initial step, I designed simple models that are used to aid research in bioengineering. These models are developed based on specific research requirements, such as microchannel structures for microfluidic devices, components for cell analysis, or parts of experimental equipment. Using SketchUp, I learned how to translate conceptual designs into digital models that can be printed, ensuring that each model has accurate dimensions, shapes, and functions suitable for laboratory experiments. Once the modeling was completed, I converted the design into a printer-compatible format, such as STL (Stereolithography), before going through the slicing process to generate precise printing instructions. This printing process allows us to produce highly accurate components, which are essential for bioengineering applications, particularly in microfluidic systems and medical diagnostic components. This experiment not only enhances my understanding of 3D printing technology and digital modeling but also opens up opportunities to optimize research equipment design in bioengineering. By having the capability to print and modify models directly in the laboratory, we can quickly test new designs, make adjustments based on experimental results, and improve efficiency in developing bioengineering-based research tools.
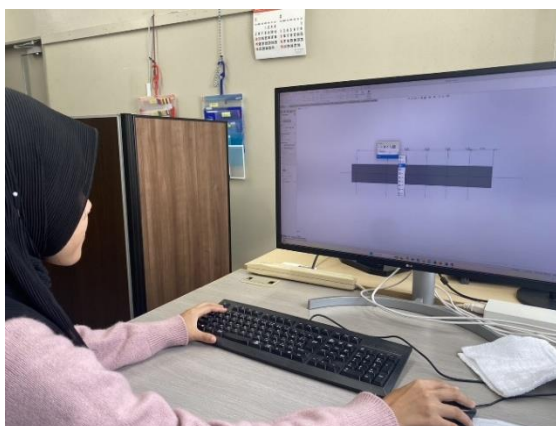
### E. Computational Cell Area Calculation Using Python
1. Overview of the Task

In this task, I collaborated with lab members in bioengineering research, who had long been trying to complete their study but faced challenges due to limited knowledge in computer science. This research aimed to develop a computational method for measuring cell areas using Python-based image processing techniques. Accurate cell area measurement is essential for monitoring growth, spatial distribution, and cell development. Without a proper calculation method, analyzing changes in cell size over time had been difficult, delaying progress in this research. The primary objective was to quantify the areas of two different cells (Cell 1 and Cell 2) as well as the surrounding empty space. By obtaining more precise data on cell size and empty regions, this research could provide a better understanding of growth patterns and spatial organization of cells. Since manual methods were inefficient and prone to errors, we decided to implement a more accurate and reproducible computational approach using Python.

2. Approach and Implementation

To achieve this, I developed a Python script that processes cell images and extracts area measurements through a structured computational workflow. The first step involved image preprocessing, including grayscale conversion, thresholding, and contour detection to clearly define cell boundaries. Once the contours were obtained, we applied pixel-based area calculation within the detected regions.

After successfully implementing this method, we obtained precise area values for both cells and the surrounding empty space, providing crucial data for the research. This computational approach not only improved accuracy but also ensured that the measurements were consistent and reproducible, allowing for further research development. This work demonstrates the effectiveness of Python-based computational techniques in bioengineering research, particularly in cell monitoring and spatial distribution analysis. By integrating image processing and computational analysis, we successfully addressed a key challenge and contributed to the study of cell growth and control mechanisms. The final results of our calculations are presented in the image attached below:
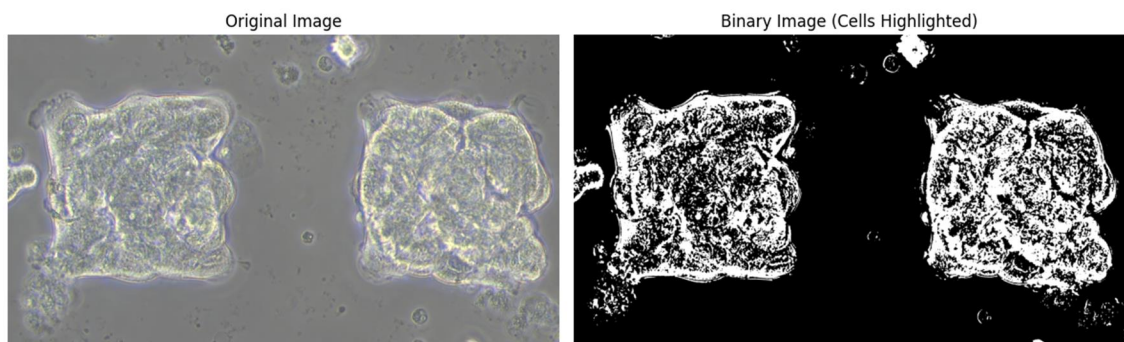
F. Bonding Process in Microfabrication for Bioengineering Applications

The bonding process in microfabrication is a critical step in ensuring the structural stability and functionality of patterned samples used in bioengineering research. In experiments involving microfluidic devices, biosensors, and other biomedical applications, proper bonding ensures that fabricated structures adhere securely to a substrate, maintaining their integrity during subsequent analysis and testing. Achieving strong,

uniform, and defect-free bonding is essential to prevent structural inconsistencies that could affect experimental results. In this research, I focused on optimizing the bonding process for microfabricated patterned structures, ensuring that the adhesion was strong and consistent. This step was crucial in preparing the samples for further evaluation and functional validation in bioengineering applications.
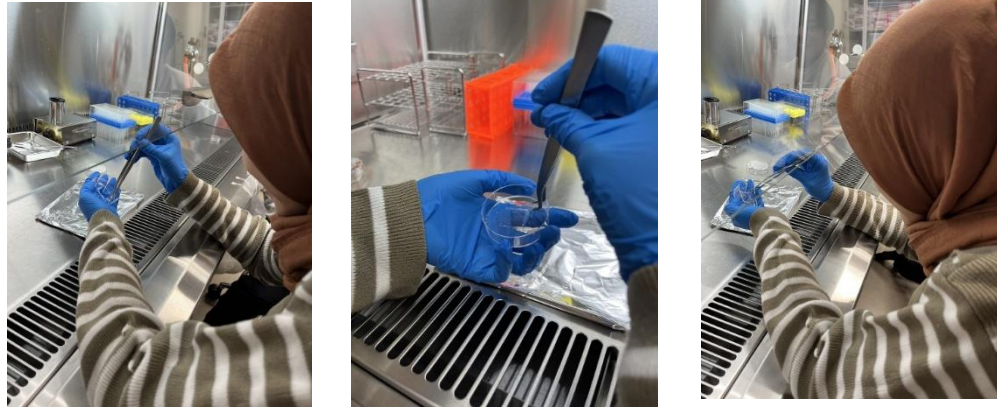
To ensure a reliable and high-quality bond, I followed a systematic bonding procedure, carefully optimizing key parameters to enhance adhesion strength and uniformity. The following steps were taken:

1. **Preparation of Patterned Structures and Substrate**
   - The patterned samples were carefully cleaned to remove contaminants that could interfere with bonding.
   - The substrate was also treated to enhance adhesion compatibility with the patterned structures.
2. **Alignment and Initial Contact**
   - The patterned structures were precisely aligned with the substrate to avoid misalignment.
   - Light pressure was applied to ensure initial contact and prevent air gaps between the layers.
3. **Optimization of Bonding Parameters**
   - **Pressure:** Adjusted to ensure sufficient contact without causing material deformation.
   - **Temperature:** Carefully controlled to promote material fusion without overheating.
   - **Bonding Duration:** Optimized to allow complete adhesion while preventing excessive exposure to heat.
4. **Final Bonding Process**
   - The bonded samples were subjected to controlled pressure and heat treatment to achieve uniform adhesion.
   - The structures were inspected to ensure no defects, delamination, or weak adhesion points.
5. **Post-Bonding Evaluation**
   - The bonded samples were tested for mechanical integrity and stability.
   - Any inconsistencies or weak points were analyzed to refine the bonding process further.

The successful completion of the bonding process ensures that the patterned structures remain intact and stable for further research. This step is essential for conducting fluid flow analysis, biological testing, and functional validation in bioengineering applications. By

optimizing the bonding parameters, this research contributes to the development of high-performance microfluidic systems, bioMEMS, and biomedical devices, which are essential for advancing diagnostic tools, tissue engineering, and lab-on-a-chip technologies.

G. Scaling Up Cell Area Measurement: Optimizing Image Processing Techniques for Multiple Cells

After successfully measuring the area of a single cell, the next challenge was to scale up the process and analyze multiple cells simultaneously. The goal was to develop a reliable and efficient method for measuring the areas of all collected cell images while maintaining accuracy. Given the increased complexity of the task, it was necessary to explore advanced image processing techniques to ensure precise segmentation and measurement of each cell. The primary challenge was that noise and inconsistencies in the images affected the accuracy of cell segmentation and area calculations.
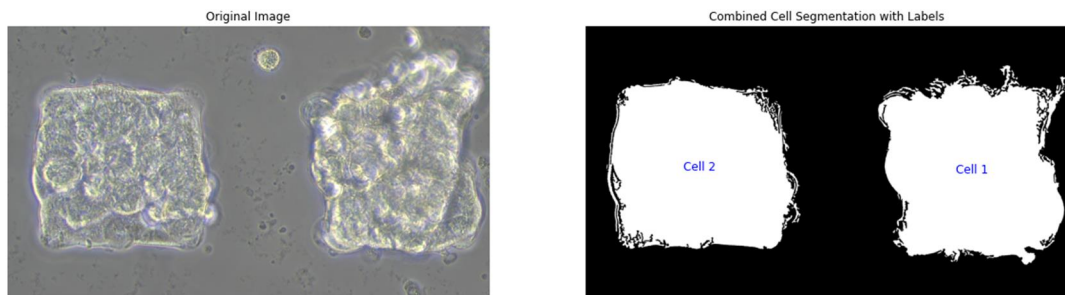
To address this issue, a preprocessing step was introduced to enhance the images, making cell boundaries more distinct and reducing segmentation errors. By optimizing this step, it was possible to improve the reliability of the calculations and obtain more accurate measurements. To achieve the most precise results, I tested different image preprocessing and segmentation techniques, focusing on:

Additionally, I visualized the differences between the Gaussian filtering method and the OpenCV-based approach, allowing Kazuki to evaluate their effectiveness. The comparison provided insights into how different preprocessing techniques affect segmentation accuracy and cell area calculation. By the end of the experiment, we successfully obtained area measurements for all 12 cells. The final results demonstrated variations in measurement accuracy based on different preprocessing techniques, helping Kazuki determine which approach best suited his research needs. This study provided valuable insights into optimizing cell area measurement for large-scale analysis, and the findings will contribute to further refinements in bioengineering and cellular research.
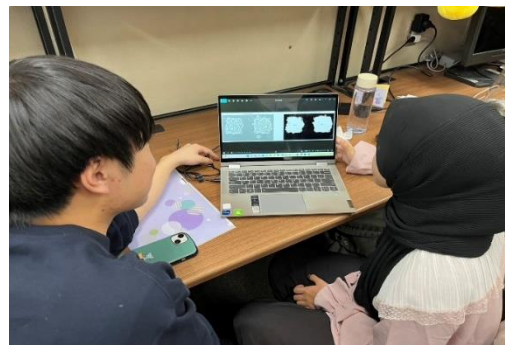
H. Dataset Collection for Cancerous and Non-Cancerous Cell Classification

The ability to accurately classify cancerous and non-cancerous cells is a crucial aspect of medical diagnostics and bioengineering research. To develop a reliable classification model, it is essential to create a high-quality dataset containing well-represented images of both types of cells. In this experiment, Kazuki and I worked on collecting a dataset using a microscope equipped with an imaging system, ensuring that the images captured provided clear morphological distinctions necessary for classification.
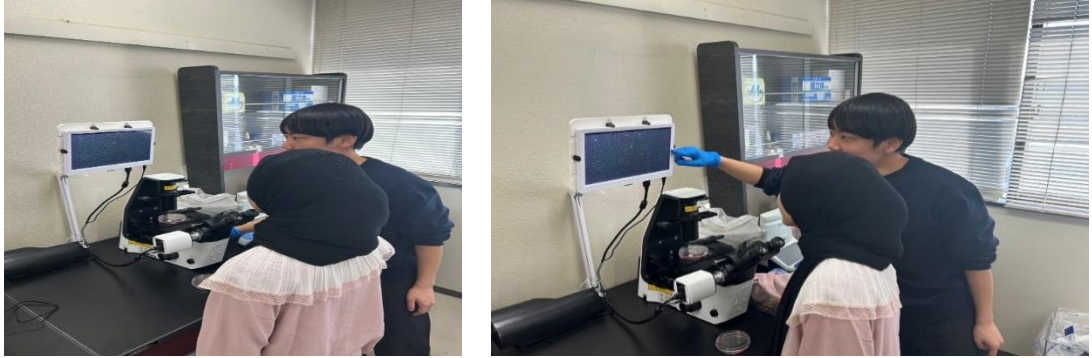
By building a structured and well-labeled dataset, we aimed to develop a machine learning-based model capable of identifying and classifying cells based on their visual characteristics. This research lays the foundation for future automated diagnostic systems, improving efficiency and accuracy in medical research and applications.

To ensure high-quality and reliable data, we followed a systematic approach to microscopic imaging and dataset organization:

1. Microscopic Imaging and Data Acquisition

   - Used a high-resolution microscope equipped with an imaging system to capture detailed cell images.
   - Adjusted focus, contrast, and lighting to enhance cell morphology and improve visibility.
   - Ensured consistency in imaging conditions to reduce variability in the dataset.

2. Balanced Representation of Cancerous and Non-Cancerous Cells

   - Captured equal numbers of cancerous and non-cancerous cell images to avoid bias in the dataset.
   - Carefully labeled each image, categorizing it as either cancerous or non-cancerous, which is essential for supervised learning in classification models.

3. Data Organization and Quality Control

   - Structured the dataset into organized folders for efficient data retrieval and training.
   - Performed initial quality inspection to ensure images were clear, well-focused, and free of distortions.
   - Removed images with artifacts or inconsistent lighting to maintain data integrity.

**I. Aspect Ratio Analysis of Cells Using Image Processing Techniques**

Accurate measurement of cell morphology is crucial in bioengineering research, particularly for analyzing cell shape, growth patterns, and classification. One of the key morphological descriptors is the aspect ratio, which is determined by measuring the major and minor axes of a cell. The aspect ratio provides insights into cell elongation,

deformation, and structural properties, which are important for applications in cancer cell analysis, tissue engineering, and diagnostic research. In this research, I applied image processing techniques to determine Axis A (major axis) and Axis B (minor axis) for calculating the aspect ratio of multiple cells. These measurements allow for a deeper understanding of cell morphology and spatial distribution in bioengineering studies.

J. Geometric Analysis of Macrophage Cells: Data Collection and Aspect Ratio Computation Under Varying Conditions

Understanding the geometric properties of macrophage cells is crucial in bioengineering and biomedical research, particularly in analyzing cell morphology, behavior, and responses to different experimental conditions. One key parameter used to quantify these properties is the aspect ratio, which provides insights into cell elongation, deformation, and structural changes under varying conditions.
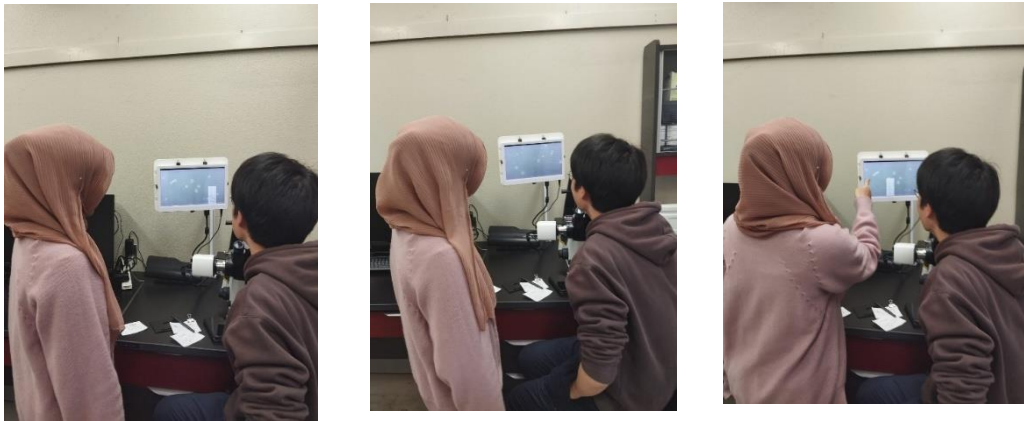
To investigate these morphological characteristics, Nakashima Sensei instructed me to collaborate with Bob in collecting new macrophage cell data and calculating their aspect ratios. The goal was to ensure a systematic analysis of macrophage cells by dividing each cell into 8 parts and capturing multiple images per section under controlled focus conditions. Additionally, we collected images of macrophage cells under 8 different conditions, allowing for a comparative analysis of their geometric properties across various experimental setups.

To ensure a structured and reliable dataset, we implemented the following systematic data collection process:

1. Dividing Macrophage Cells into 8 Sections
    *   Each cell was divided into 8 distinct parts for localized morphological analysis.
    *   For each part, we captured three images under identical focus conditions to maintain consistency in image acquisition.
2. Capturing Images Under 8 Experimental Conditions
    *   Beyond section-based imaging, we collected additional images representing 8 distinct conditions, with one image per condition to study the macrophage cell responses under different experimental setups.
3. Ensuring Image Quality and Consistency
    *   All images were taken using the same microscope settings, adjusting focus and contrast to minimize variability.
    *   The data was structured systematically to ensure accurate comparisons across different sections and experimental conditions.

K. Aspect Ratio Calculation Using Black-and-White and Color-Based Segmentation Methods

The aspect ratio is a crucial parameter in cell morphology analysis, providing insights into cell elongation, deformation, and structural characteristics. In this study, I calculated the aspect ratio using two different segmentation methods to ensure accurate and reliable results. The first method, black-and-white segmentation, is based on thresholding, which distinguishes cells from the background by analyzing grayscale intensity. The second method, color-based segmentation, utilizes color channels to more precisely identify different regions within the cell structure. The application of these two methods aims to evaluate the effectiveness of each approach in analyzing newly acquired macrophage cell data. By comparing these techniques, I can determine the optimal method for achieving more precise segmentation, enabling a better analysis of cell morphology under various experimental conditions.

L. Adaptive Analysis of Macrophage Cells: Aspect Ratio, Complexity, and Equivalent Diameter Calculation

In biological research, understanding cell morphology under different culture conditions is essential for evaluating growth patterns, structural adaptation, and functional changes. This study aimed to refine the aspect ratio calculation by adapting the segmentation process to cell orientation, ensuring a more precise representation of cellular structures. Unlike the initial dataset, where cells were freshly isolated, the new dataset was collected after 24 hours of culture, allowing for the observation of morphological changes over time. Additionally, to enhance the comprehensiveness of the analysis, two extra parameters—complexity and equivalent circle diameter—were introduced based on Nakashima Sensei's suggestion.

M. Aspect Ratio, Complexity, and Boundary Length Variations: Evaluating Macrophage Cell Morphology

Understanding the morphological characteristics of macrophage cells is essential in bioengineering and biomedical research, particularly in analyzing cell shape, structural complexity, and membrane dynamics under different experimental conditions. This study focuses on three critical parameters: Aspect Ratio, Complexity, and Boundary Length, which provide a comprehensive insight into how macrophage cells adapt, spread, and interact with their environment. Additionally, three key aspects were considered for each parameter: structural adaptation, functional implications, and analytical significance. The Aspect Ratio (AR) was calculated by comparing the major axis (Axis A) and minor axis (Axis B) of each cell. A higher aspect ratio indicates elongated cells, often associated with active migration or polarization, whereas a lower aspect ratio suggests more circular cells, indicating a resting or non-migratory state. This metric is crucial for evaluating cell movement patterns and morphological transformations during different biological processes. The three aspects of aspect ratio analysis include (1) cell elongation behavior, (2) shape stability over time, and (3) correlation with macrophage activation states.

Complexity describes the irregularity of cell shape, measured as the perimeter squared divided by the cell area. Higher complexity values indicate branched or irregular structures, commonly seen in activated macrophages, whereas lower values represent smooth and rounded cell morphologies. This measurement helps assess how macrophages respond to stimuli by altering their structural composition. The three key aspects of complexity evaluation include (1) cell

membrane roughness and extensions, (2) differentiation between activated and resting states, and (3) variability across different culture conditions.

The Boundary Length represents the total perimeter of a cell, providing insights into membrane expansion and cytoskeletal dynamics. Cells with longer boundary lengths often exhibit increased activity and motility, whereas shorter boundary lengths indicate more compact and less dynamic cells. This parameter is useful for studying how macrophages extend and retract their plasma membranes during immune responses. The three critical aspects of boundary length analysis are (1) extent of cytoplasmic protrusions (e.g., filopodia, lamellipodia), (2) membrane flexibility and deformation, and (3) relationship between boundary length and overall cell spreading efficiency. By integrating these morphological parameters, this study enhances our understanding of macrophage behavior, structural adaptation, and functional dynamics. These findings are essential for applications in biomedical engineering, immune system research, and AI-driven cell morphology classification. Future studies will focus on automated segmentation techniques and temporal morphological changes to further refine macrophage cell analysis.

1. **Analysis of Boundary Length Distribution in Macrophage Cells**

The analysis of macrophage cell boundary length provides essential insights into how cells adapt and respond to their surrounding environment.
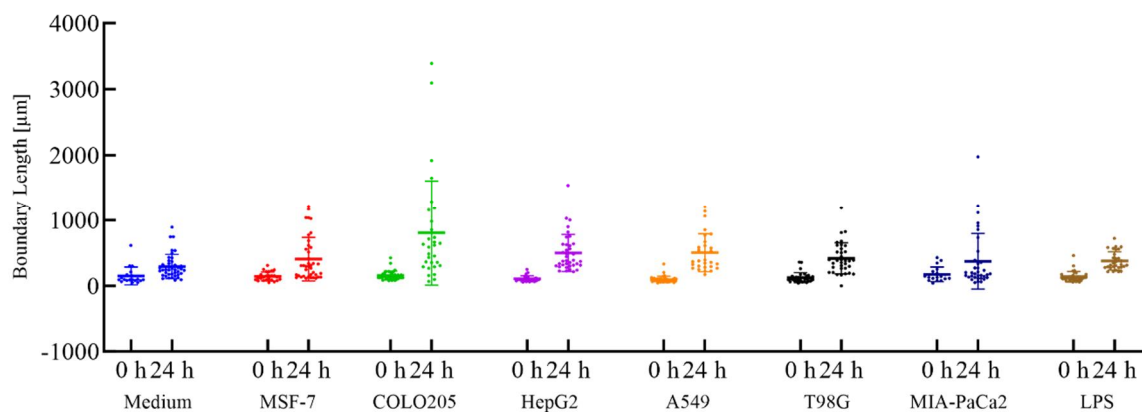
This graph illustrates the distribution of cell perimeter lengths across different experimental conditions, with each data group represented in different colors. From the image analysis results, significant variations in boundary length are observed, indicating that some cells exhibit greater membrane extensions than others. The mean intensity value of 0.89 with a standard deviation of 9.53 suggests high variability in the data distribution, likely influenced by differences in macrophage activation states.

Cells with longer boundary lengths are generally associated with more active macrophages, where the membrane undergoes increased expansion and the formation of cytoplasmic extensions such as filopodia and lamellipodia. In contrast, shorter boundary lengths indicate more compact and less dynamic cells, which may be in a resting state or exhibit minimal response to environmental stimuli. Some groups in the graph display a wide range of variations, which could suggest a high level of heterogeneity within the cell populations under different experimental conditions. This analysis serves as a foundation for further studies on the relationship between boundary length, complexity, and aspect ratio, helping to understand macrophage morphological changes. The next step is to conduct a more in-depth statistical analysis to determine whether the differences between groups are significant and how

environmental factors influence cell shape and membrane dynamics. With these insights, future research can focus on applying machine learning models to classify cells based on morphological parameters, enabling automated analysis in cellular studies and bioengineering.

2. Complexity Analysis of Macrophage Cell Morphology

The analysis of macrophage cell complexity provides insight into how cell shape varies under different experimental conditions. The graph visualizes complexity distributions, where each color represents a distinct condition, and the scatter plot with error bars shows mean complexity values along with their standard deviations. Complexity is measured based on the perimeter-to-area ratio, which quantifies how irregular or branched a cell structure is. A higher complexity value indicates cells with more membrane protrusions and irregularities, while lower complexity values suggest a smoother and more compact morphology. From the observed results, some groups exhibit significantly higher complexity, suggesting that macrophages in these conditions undergo morphological changes, potentially due to activation or environmental stimuli.
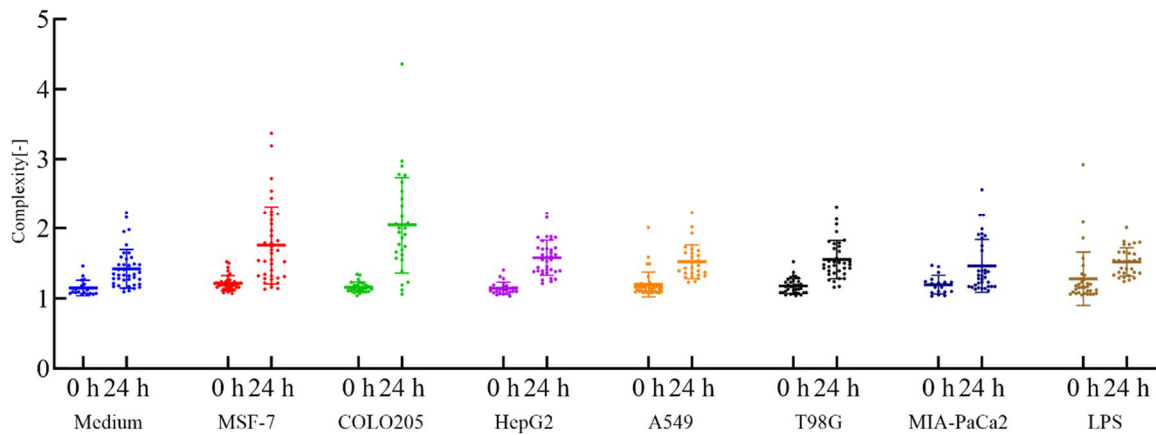
In contrast, groups with lower complexity values show a more rounded and uniform morphology, which may indicate that the cells are in a resting or less active state. The variability observed across different conditions suggests that external factors influence macrophage structure, potentially altering their functional behavior, such as their ability to migrate or interact with other cells. This analysis has important implications for understanding macrophage activation, structural adaptation, and immune responses. Higher complexity values could be linked to macrophage activation states, where cells extend their membrane for processes like migration or phagocytosis. Additionally, these morphological features could be integrated into machine learning models for automated cell classification, allowing for efficient identification of activated vs. non-activated cells. Future research will focus on performing statistical validation of these findings and correlating complexity with other morphological parameters, such as boundary length and aspect ratio, to provide a more detailed characterization of macrophage behavior.

3. Aspect Ratio Analysis of Macrophage Cells

The aspect ratio of macrophage cells is a crucial parameter in understanding cell elongation, shape variation, and structural adaptation under different experimental conditions. The graph represents the distribution of aspect ratio values, with each dataset corresponding to a specific condition. The aspect ratio is calculated by comparing the major axis (Axis A)

12

to the minor axis (Axis B), where higher values indicate elongated cells, typically associated with cell migration, polarization, or activation, while lower values correspond to more rounded and compact cells, suggesting a resting or non-motile state. From the observed results, some groups exhibit higher aspect ratios, indicating that macrophages in these conditions tend to elongate, possibly due to increased motility or an activated state. In contrast, other groups display lower aspect ratios, suggesting that the cells remain more rounded and stable, potentially representing a resting phase or a non-stimulated state. The presence of wide error bars in certain groups suggests high variability in cell elongation, indicating that different cells within the same condition may respond differently to environmental stimuli. Conversely, groups with narrower distributions reflect more uniform cell morphologies, suggesting that cells are adapting consistently under those conditions.
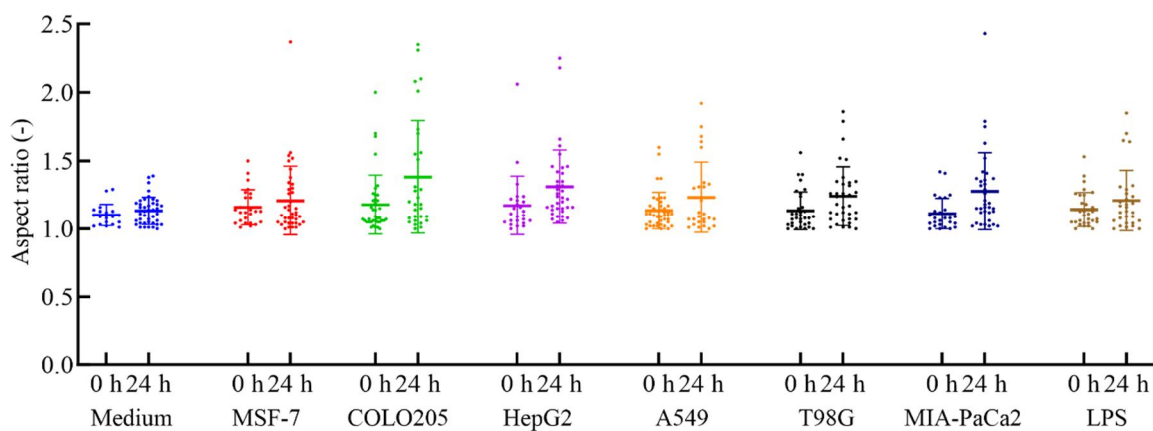
These findings provide valuable insights into macrophage behavior and structural adaptation, particularly in response to environmental changes. Cells with higher aspect ratios could be undergoing pro-inflammatory activation, extending their shape for migration or interaction with their surroundings. The variability in aspect ratio across different conditions suggests that external factors influence macrophage morphology, affecting their functional state and mechanical properties. Additionally, these measurements can be applied to automated classification models using AI, allowing for efficient and objective identification of activated vs. non-activated macrophages. Further statistical analysis will be conducted to confirm the significance of these variations, and future studies will explore correlations between aspect ratio, complexity, and boundary length to build a more comprehensive model of macrophage morphological changes.

CHAPTER 3: FUTURE RESEARCH PLAN

This internship under the IROAST Young Internship Researcher program has provided a comprehensive learning experience in the field of biomedical engineering, particularly in the application of computer science for cell morphology analysis. Throughout this research, I have focused on analyzing macrophage cell structures by implementing advanced image processing techniques, deep learning models, and statistical evaluations. The key aspects covered include aspect ratio, complexity, and boundary length measurements, which have been crucial in understanding cell behavior, activation states, and structural adaptations under different experimental conditions. The internship has reinforced the importance of interdisciplinary

collaboration, where my expertise in computer science was applied to enhance the efficiency of image analysis and data processing for biological research. I have worked closely with members of the bioengineering laboratory under Prof. Yuta Nakashima to refine computational methods for segmenting, classifying, and predicting cell morphology. The integration of Python-based computational techniques allowed for the automation of image segmentation, statistical calculations, and data visualization, significantly reducing manual efforts in biological experiments.

3.1 Future Research Directions

A. Deep Learning Integration for Automated Cell Classification

- The next step in this research is to develop a deep learning-based classification system capable of distinguishing between activated and non-activated macrophages using morphological features such as aspect ratio, complexity, and boundary length.
- By utilizing Convolutional Neural Networks (CNNs) and Transformer-based models, we can improve cell detection accuracy and pattern recognition, enabling real-time analysis of cell behavior.

B. Expanding Image Dataset and Model Generalization

- A major limitation in biological image analysis is dataset diversity. Future research will focus on expanding the dataset to include macrophages cultured under different conditions, ensuring that machine learning models are robust and generalizable.
- Multi-modal learning will be explored by incorporating metadata from cell culture conditions along with image-based features, providing a more comprehensive understanding of cell behavior.

C. Optimized Feature Extraction for Bioengineering Applications

- The extracted parameters, such as aspect ratio, complexity, and boundary length, provide valuable insights into cell functionality and adaptation.
- Future research will focus on refining feature extraction methods to improve the correlation between computationally derived morphological features and biological phenomena.
- The goal is to enhance biomedical diagnostic systems by providing accurate, quantitative, and reproducible cell morphology assessments.

CHAPTER 4: CONCLUSION

The research conducted during this IROAST internship serves as a strong foundation for future advancements in biomedical engineering, especially in the automated analysis of cellular structures. By integrating deep learning, advanced image processing, and computational modeling, this research can contribute to more efficient, accurate, and scalable solutions for studying cell morphology and immune responses. Moving forward, expanding the dataset, optimizing computational methods, and enhancing interdisciplinary collaboration will be key steps toward applying this research in real-world medical and bioengineering applications.

No. 2-7-4	Evaluation of microscopic stress state in asperities on a rock discontinuity		
Name	Pimao LI	Title	Master Student
Affiliation	Shandong University of Science and Technology, China Email: lipimao9908@163.com		
Research Field	Environment-friendly technology		
Period of Internship	Online: September 23, 2024 - September 27, 2024 Onsite: October 7, 2024 - November 29, 2024		
Host Professor	Atsushi SAINOKI	Title	Associate Professor
Affiliation	Faculty of Advanced Science and Technology Email: atsushi_sainoki@kumamoto-u.ac.jp		

1. Tasks and Learning Outcomes

During this internship, my primary focus was evaluating the microscopic stress state at asperities on a rock discontinuity through 3DEC numerical simulations. The key tasks accomplished and insights gained are as follows:

1) Mesh Simplification and Model Preparation:

I utilized Rhino software to simplify the mesh of STL files, reducing mesh density while retaining critical geometric features required for numerical simulations. Through this process, I became proficient in various tools and functional modules within Rhino, such as mesh simplification, repair, and optimization.

Specimen STL file → Simplifying meshing with Rhino software → Use the Griddle plug-in to process the model into a model acceptable to the 3DEC software
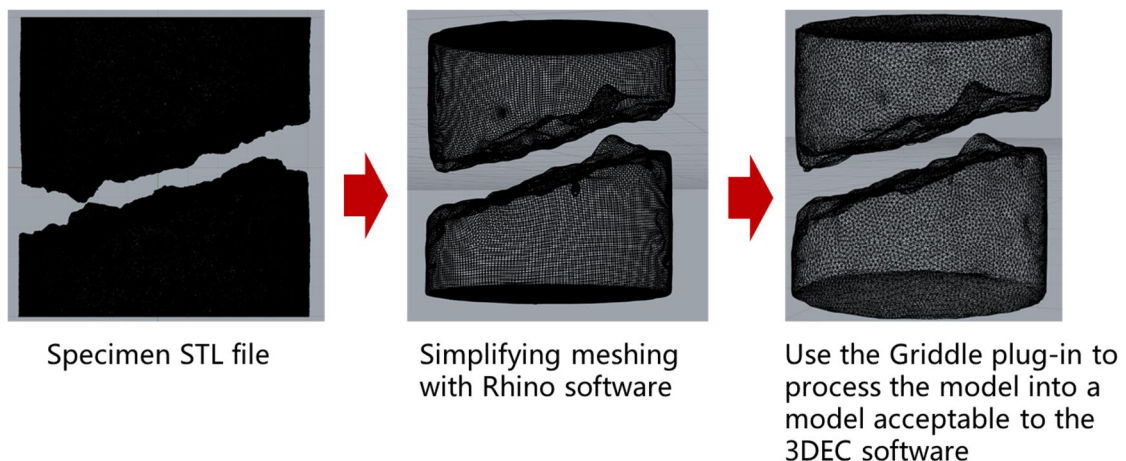
Fig. 1 Mesh Simplification

2) Model Generation and Parameter Configuration:

After exporting the simplified mesh in grid format, I prepared the model in 3DEC for uniaxial compression tests. This stage involved defining material parameters, boundary conditions, and loading configurations, with a focus on replicating the microscopic characteristics of the discontinuity. Through this task, I gained a deeper understanding of the importance of parameter calibration in accurately simulating the mechanical behavior of rock discontinuities.
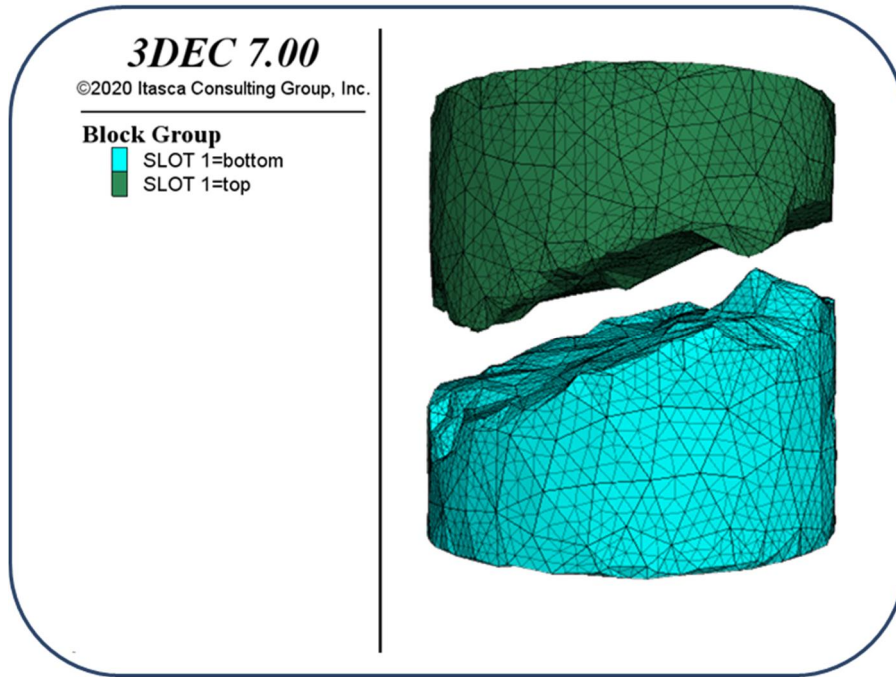
Fig. 2 Model Generation
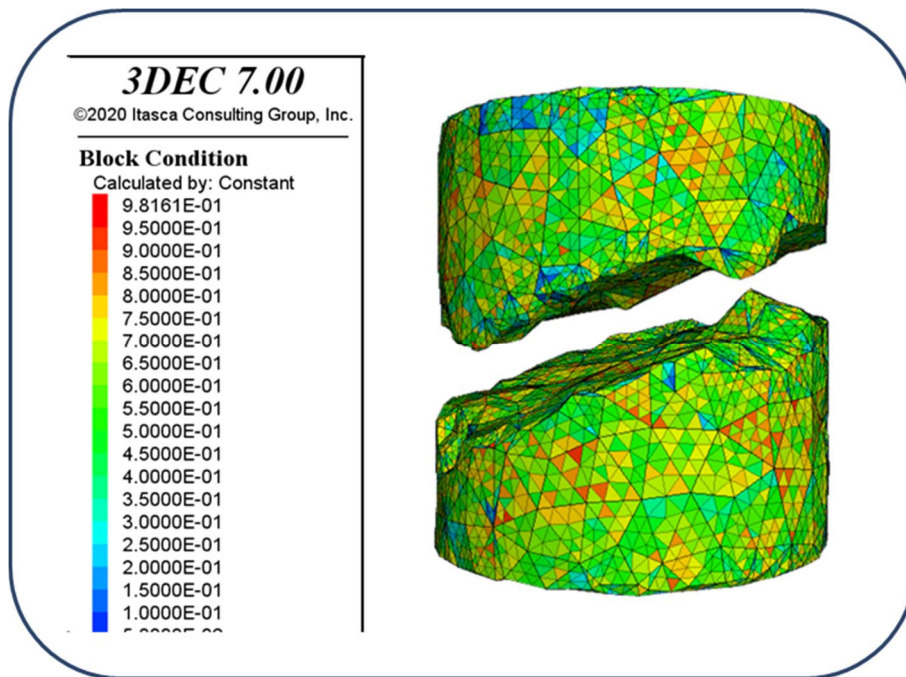
Fig. 3 Define material parameters

3) Uniaxial Compression Simulation:

I conducted uniaxial compression tests on cylindrical models using 3DEC, focusing on the stress distribution and deformation characteristics at asperities under loading conditions. This process significantly enhanced my skills in numerical simulation and stress analysis.
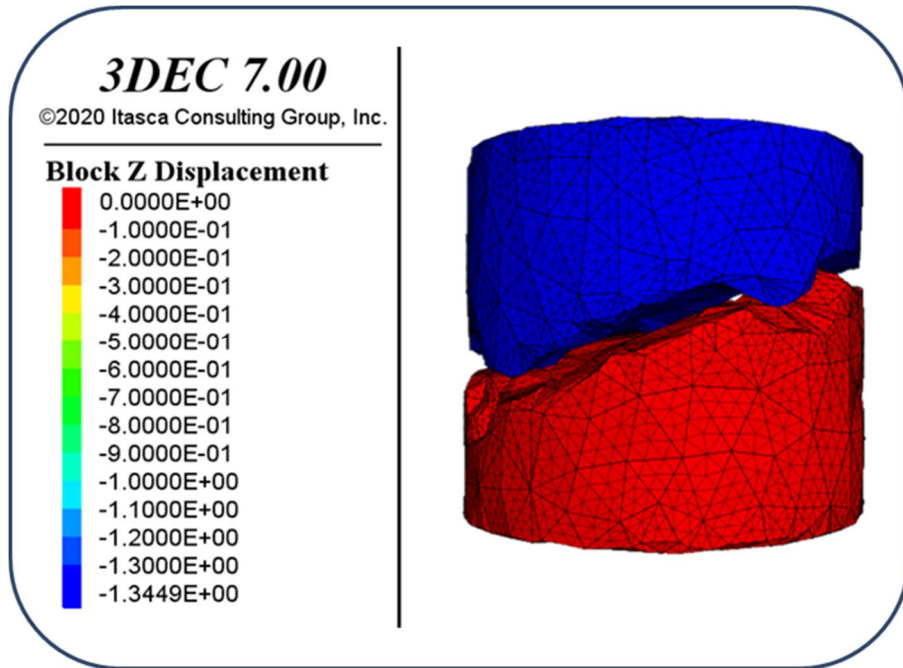
Fig. 4 Pre-compression contact
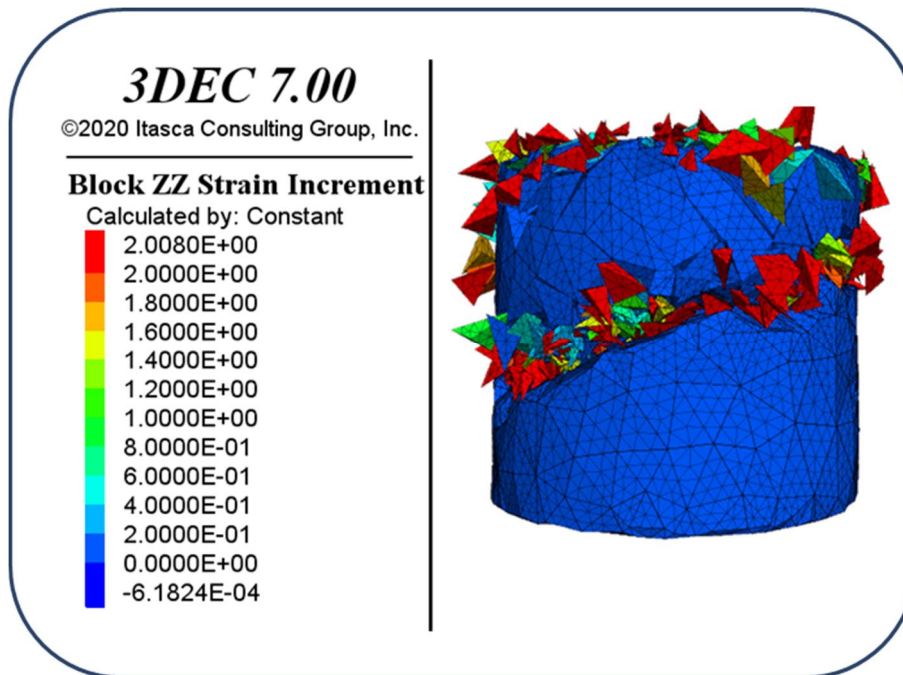
Fig. 5 uniaxial compression

4) Post-Processing and Analysis:
I utilized advanced post-processing techniques to analyze the simulation results, focusing on the strain distribution at asperities. Visualization of failure zones was generated to provide deeper insights into the deformation and failure mechanisms.
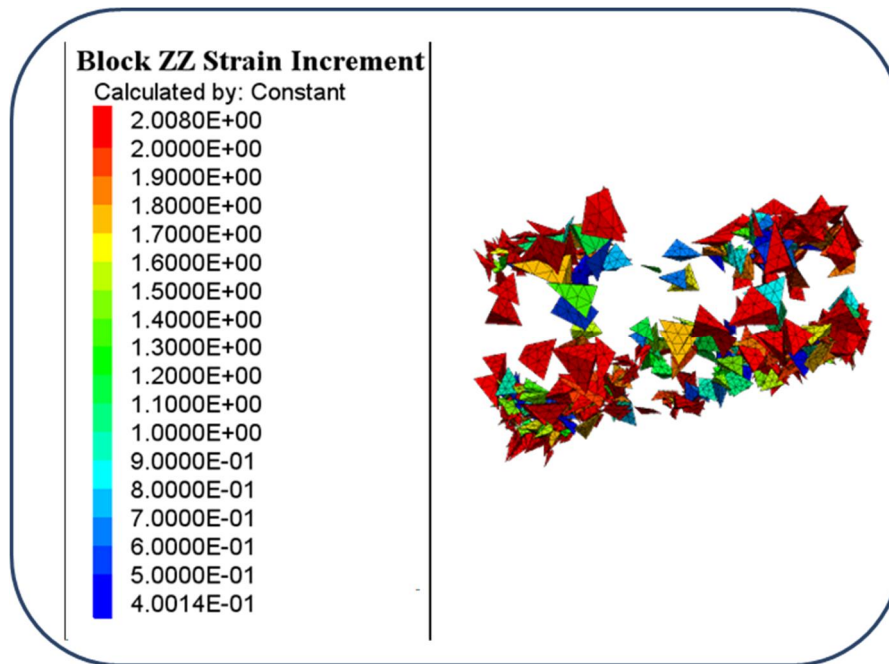
Fig. 6 Strain distribution at the bulge

Through the above tasks, I gained proficiency in key techniques such as numerical modeling, mesh optimization, and strain analysis of rock discontinuities. Additionally, I further enhanced my expertise in using software tools like Rhino and 3DEC, laying a solid foundation for future research in rock mechanics.

2. Future Research Plans

Building on the experience gained during this internship, I plan to further deepen my research in the following areas:

1) Incorporating Complex Geometric Features:

Future studies will focus on analyzing more complex asperity geometries and discontinuity patterns to better simulate real-world engineering conditions. This includes integrating natural surface roughness characteristics to explore their impact on stress states and failure mechanisms.

2) Experimental Validation:

I plan to conduct laboratory tests, such as CT scanning of asperity deformation, to validate the numerical simulation results and acquire additional data to improve model accuracy.

3) Sensitivity Analysis and Optimization:

I aim to systematically analyze the sensitivity of key parameters, such as asperity size, shape, and material properties, to identify the factors with the most significant influence on stress states. This will facilitate optimizing model designs for specific applications, such as stability assessments in rock engineering.

The knowledge and skills acquired during this internship have provided a strong foundation for my future research and have fueled my enthusiasm for addressing complex challenges in rock mechanics. I aspire to contribute to advancements in the field and its practical applications through continued investigation and innovation.

No. 2-7-5	Deep Learning Approaches for Rainfall-runoff Modeling and Flood Mitigation Strategies			
Name	Shofiyul Anam Ai MUBAROK	Title	Researcher	
Affiliation	Institut Teknologi Sepuluh Nopember, Indonesia Email: a1mubar0k@outlook.com			
Research Field	Environment-friendly technology / Strengthening resilience / Data science and AI			
Period of Internship	Online: November 18, 2024 – December 13, 2024 Onsite: December 17, 2024 – January 10, 2025			
Host Professor	Kei ISHIDA		Title	Associate Professor
Affiliation	Center for Water Cycle, Marine Environment and Disaster Management Email: keiishida@kumamoto-u.ac.jp			

1. Tasks you have engaged in and what you have learned

Hydrology is concerned with the interaction of water with the environment at each stage of the hydrologic cycle. The hydrological cycle contains several interrelated components; for example, streamflow is connected with precipitation. Surface runoff is a key component of water resource monitoring as well as resolving water quality and quantity issues like flood predictions and ecological and biological connections in the aquatic environment. Runoff is also a major contributor to contaminant transport because surplus nutrients and pesticides from agricultural lands are transported into rivers by runoff events. Therefore, estimating runoff is crucial for the management of water resources, the aquatic environment, and flood mitigation.

Runoff models depict what happens in water systems as a result of changes in impervious areas, vegetation, and weather events. A runoff model is a series of equations that help in the calculation of the quantity of rainfall that converts into runoff as a function of several parameters used to characterize the watershed. Based on the modeling structure, rainfall-runoff modeling can be characterized as empirical or data-driven black box models, conceptual models, and physically based models.

As a kind of data-driven approach, machine learning models are effective and usually outperform standard hydrologic models but they require high computational resources to train complex relations within input and output datasets. As a subfield of machine learning, deep learning is a highly discussed subject in various fields and has been applied to address numerous issues in hydrology, including rainfall-runoff modeling. Thus, developing accurate and efficient rainfall-runoff models using deep learning approaches would be beneficial for disaster prevention and mitigation efforts. In this research internship program, I employed deep learning approaches for runoff simulation, such as recurrent neural network (RNN) and convolutional neural network (CNN), and attention mechanism.
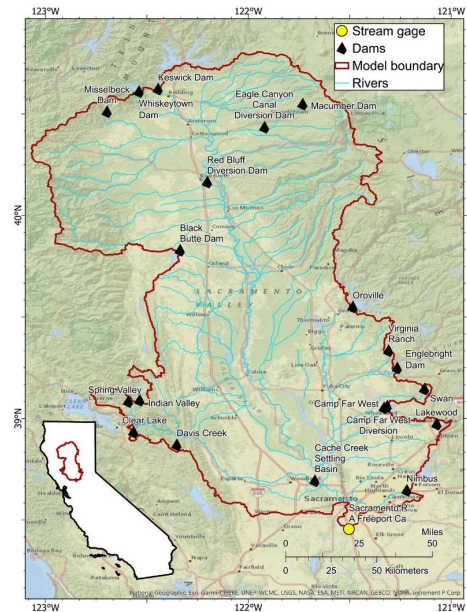
Figure 1. Sacramento River basin coverage area.
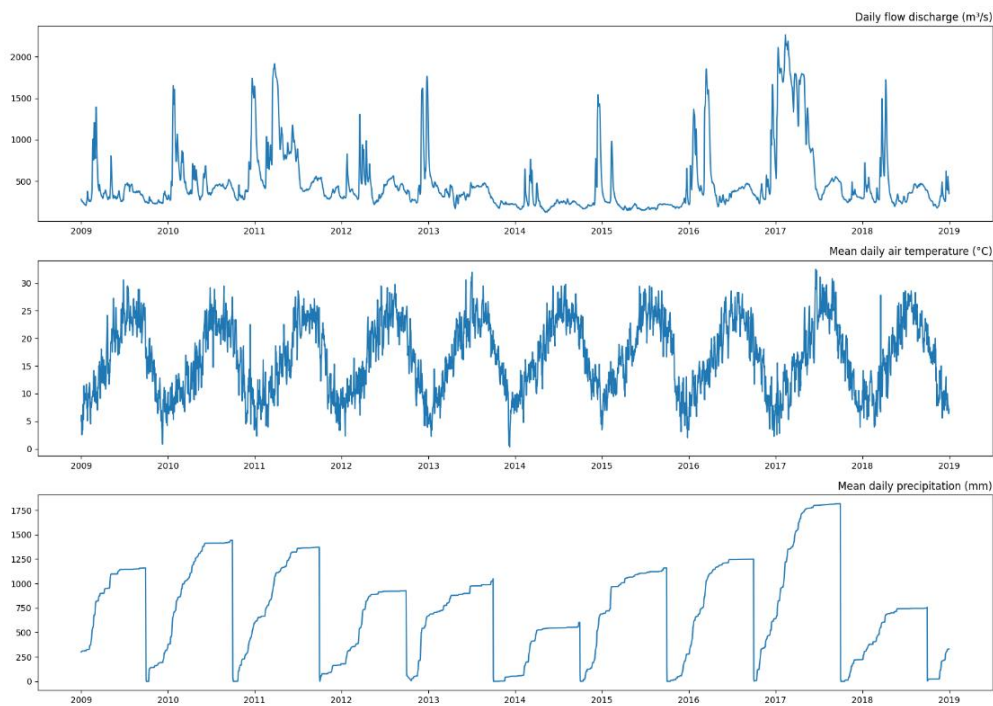
Figure 2. Visualization of the daily observation data.

The rainfall-runoff modeling was implemented to simulate flow discharge with the consideration of air temperature and precipitation data using several deep learning approaches. A simulation of flow discharge of Sacramento River in California, USA is selected to be the case study. The Sacramento River basin as illustrated in Figure 1, characterized by diverse land use, topography, and vegetation, offers unique insights into rainfall-runoff interactions across varied environments, such as urban, agricultural, and forested areas. Its significant rainfall patterns, including Northern California's rainy winters, allow for modeling runoff behaviors under both moderate and heavy precipitation events. The region's extensive historical hydrological data, encompassing streamflow, rainfall, and weather records, supports accurate calibration and validation of runoff models. Additionally, the area's susceptibility to flooding, exacerbated by Sierra Nevada snowmelt during the rainy season, underscores the importance of such models in flood risk management, infrastructure planning, and emergency response.

The dataset used was observational data of daily river flow discharge, air temperature, and precipitation collected over 10 years, from 2009 to 2018. The daily river flow discharge, air temperature, and precipitation were collected from the Global Runoff Data Centre (GRDC), National Centers for Environmental Information (NCEI), and California Data Exchange Center (CDEC), respectively. The dataset was segregated into three sets: training, validation, and test datasets. The periods of the training, validation, and test datasets are 80%, 10%, and 10%, respectively. Missing data points are handled by using linear interpolation.

In this study, three RNN architectures were used: LSTM, GRU, and original RNN. Multi-Head Attention (MHA) and a one-dimensional (1D) CNN architectures were used for comparation. The same hyperparameter configuration listed in Table 1. In addition, 1D CNN was used with the configuration listed in Table 2. Figure 2 illustrates the data daily flow discharge, air temperature, and precipitation assigned as input features. The task was to simulate flow discharge data for the next seven days based on the given data features from the past data within the given timesteps. Peak flows (flows ≥ 90th percentile) were measured to evaluate and enhance the model's performance in capturing extreme hydrological events, such as floods. The importance of these peak flows lies in their relevance to applications like flood management and hydraulic structure design, especially under climate change conditions. By focusing on these extreme conditions, reliability in predicting critical events was ensured, addressing the challenges essential for practical water resource and disaster management tasks.

Table 1. Tuning information used in this study.

Hyperparameter	Value or method
Hidden state length	50
Numbers of recurrent layer	1
Batch size	64
Embed size	64
Number of attention heads	4
Loss function	MSE
Optimization algorithm	Adam
Epoch	100
Patience of early stopping	50
Sequence length	30, 60, 120
NCHF	8, 16, 32
Evaluation metrics	RMSE, R, NSE

Table 2. Configuration information of the 1D CNN layers.

Layer	Input Channel	Output Channel	Kernel	Stride	Padding
Conv1	num_features	8/16/32	3	1	1
MaxPool1	8/16/32	8/16/32	2	2	-
Conv2	8/16/32	16/32/64	3	1	1
MaxPool2	16/32/64	16/32/64	2	2	-
FC	16/32/64 * (seq_length // 4)	output_size	-	-	-

The results of RNN statistically compared each model in Table 3. Figure 3 illustrates the relationships between variables and Figure 4 illustrates the observed and simulated flow discharge. Overall, the selected RNN models showed the capability to learn long-duration dependencies between input and target time-series variables with relatively higher accuracy than 1D CNN. It resulted in lower accuracy for the configuration set among the employed RNN models. NSE value of more than 0.5 is deemed acceptable. The comparison results may be different with other combinations of the hyperparameters. However, the MHA model with sequence length above 60 because longer sequences significantly increase the complexity of the task, requiring the model to learn relationships across a broader temporal window, which can exceed its representational capacity, especially if the embedding size or number of attention heads is insufficient.

Additionally, longer sequences may introduce irrelevant or noisy information, diluting the model's ability to focus on the most critical temporal patterns. The reduced predictive relevance of distant timesteps further complicates learning, and the model may struggle to maintain meaningful gradient flow across extended inputs.
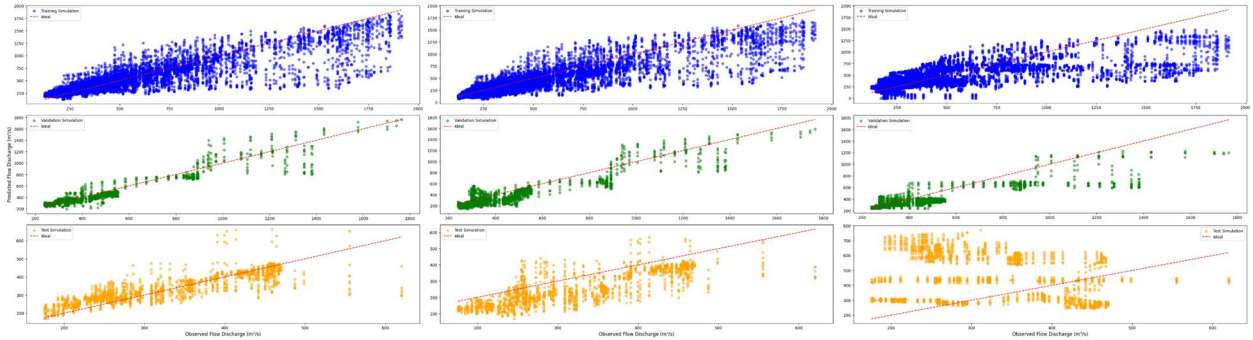

Figure 3. Example of scatter plot for LSTM, 1D CNN, and MHA from the training, validation, and testing phase with sequence of 120.
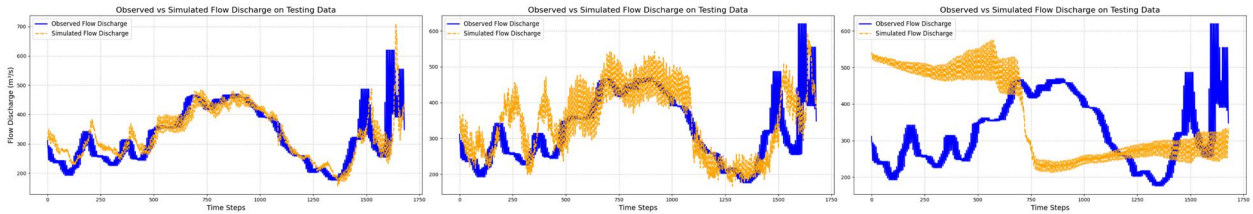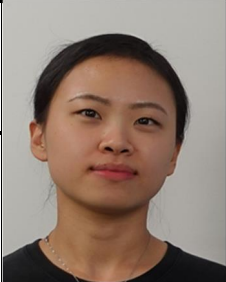

Figure 4. Example plot of observed and simulated flow discharge for LSTM, 1D CNN, and MHA from the testing phase with sequence of 120.

Table 3. Statistical comparison of each model.

Model	Seq.	Training			Validation			Testing			
		RMSE	R	NSE	RMSE	R	NSE	RMSE	RMSE (Peak)	R	NSE
LSTM	120	144.244	0.888	0.782	91.515	0.946	0.890	53.288	34.155	0.822	0.671
	60	136.928	0.899	0.803	145.145	0.972	0.926	152.030	305.045	0.806	0.637
	30	135.914	0.898	0.806	148.783	0.974	0.944	149.592	264.290	0.794	0.617
GRU	120	139.867	0.892	0.795	84.749	0.952	0.905	54.741	40.516	0.820	0.658
	60	136.357	0.897	0.803	133.771	0.971	0.936	153.032	330.009	0.799	0.629
	30	139.843	0.893	0.795	150.478	0.974	0.943	147.651	276.246	0.798	0.628
RNN	120	137.730	0.895	0.801	81.368	0.956	0.913	54.525	46.551	0.817	0.657
	60	137.751	0.894	0.799	136.158	0.972	0.934	153.779	353.263	0.795	0.625
	30	140.335	0.891	0.793	163.079	0.975	0.933	146.203	326.651	0.800	0.634
MHA	120	211.766	0.743	0.537	168.461	0.873	0.657	241.619	102.865	-0.305	-1.162
	60	167.941	0.852	0.707	185.117	0.953	0.878	200.319	315.444	0.700	0.433
	30	162.098	0.857	0.726	208.252	0.964	0.891	168.395	266.847	0.741	0.530
1D CNN, NCHF 32	120	163.336	0.871	0.727	136.673	0.916	0.772	83.136	83.408	0.744	0.391
	60	139.999	0.922	0.803	144.673	0.964	0.926	194.695	309.195	0.807	0.498
	30	117.398	0.925	0.855	157.333	0.970	0.938	153.810	351.859	0.788	0.597
1D CNN, NCHF 16	120	117.543	0.932	0.857	123.027	0.895	0.801	80.853	57.990	0.663	0.303
	60	134.600	0.912	0.812	159.614	0.962	0.910	175.999	352.801	0.778	0.547
	30	147.765	0.879	0.771	205.525	0.964	0.895	147.799	269.407	0.799	0.626
1D CNN, NCHF 8	120	159.654	0.878	0.741	133.823	0.912	0.779	93.848	46.481	0.796	0.348
	60	151.246	0.884	0.763	192.578	0.964	0.874	168.515	346.296	0.783	0.573
	30	141.685	0.890	0.790	216.514	0.968	0.885	150.080	281.206	0.799	0.618

2. Future research plan

During this research internship, I gained valuable insights into advanced modeling techniques and their applications in hydrological studies. I plan to continue collaborating with Dr. Ishida to refine the model and explore opportunities for joint publications. Dr. Ishida's research focus and the work of his laboratory align closely with my future research interests, particularly in leveraging deep learning to address critical challenges in hydrology. I am especially excited about the possibility of establishing a robust research collaboration with Dr. Ishida at IROAST, Kumamoto University. The integration of cutting-edge deep learning methods into hydrological modeling offers a promising avenue for tackling urgent issues such as flood mitigation, water resource management, and the assessment of climate change impacts. Dr. Ishida's expertise, coupled with the innovative resources and environment of his laboratory, will be pivotal in advancing these efforts. I am confident that this collaboration will not only enhance my academic development but also contribute significantly to impactful research in this field.

No. 2-7-6	The role of extrachromosomal DNA in cancer drug resistance acquisition			
Name	Yoonjoo NAM	Title	Ph.D. Student	
Affiliation	SungKyunkwan University, Korea Email: nyj9906@g.skku.edu			
Research Field	Data science and AI			
Period of Internship	Online: July 1, 2024 – July 5, 2024 Onsite: July 16, 2024 – July 29, 2024			
Host Professor	Ruda LEE	Title		Associate Professor
Affiliation	Institute of Industrial Nanomaterials Email: aeju-lee@kumamoto-u.ac.jp			

Details of activities
1. Tasks you have engaged in and what you have learnt
【16/07/24 ~ 22/07/24】
- **Identification of ecDNA breakpoint sequences**
Whole Genome Sequencing (WGS) data was utilized to identify ecDNA breakpoint sequences using SvABA and Amplicon Architect. Based on the identified sequences, primer sequences were designed.
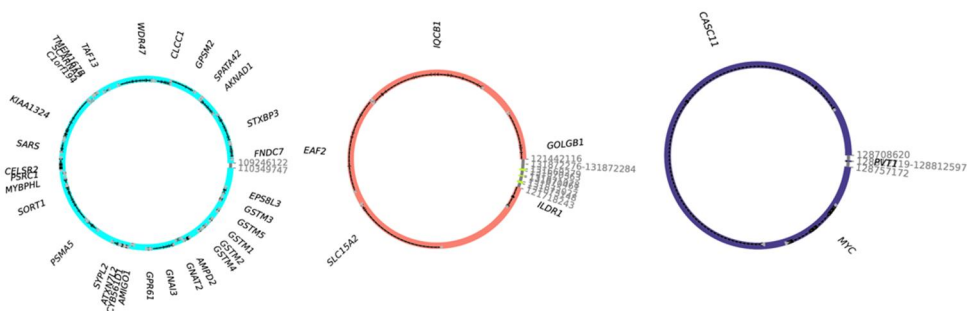

Figure 1. analyzed ecDNA structure

- **DNA extraction from parental and MDR cell-lines**
DNA was extracted from parental and MDR cell-lines using the 'QIAamp DNA Mini Kit' for subsequent PCR analysis. Since a single cell can contain multiple types of ecDNA, a sufficient amount of DNA is required.
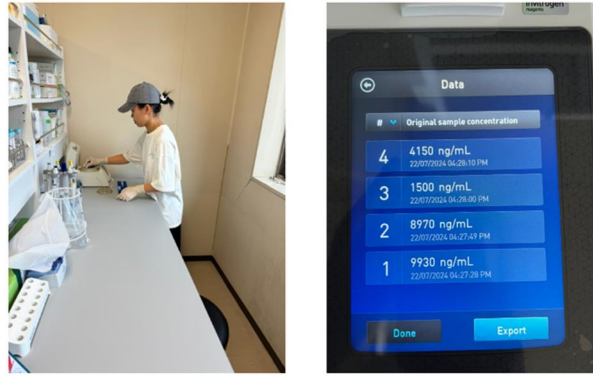
Figure 2. DNA Extraction Process and Quantification Results of Extracted DNA

【23/07/24 ~ 29/07/24】
- **PCR using extracted DNA**
Basic experimental knowledge necessary for PCR, such as the preparation of agarose gels, was acquired. PCR was then conducted using the previously extracted DNA and the identified breakpoint target primers. This experiment was conducted to cross-check the presence of ecDNA identified in silico at the in vitro level.

Figure 3. PCR Experiment Procedure and Results

Through the PCR results, we were able to confirm the actual presence of the predicted ecDNA structure.

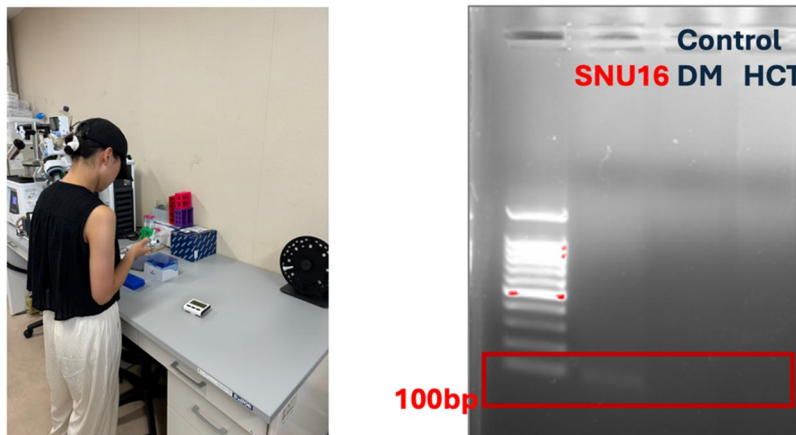| Project Title	Study on the Preparation and Properties of Alkali-Activated Fly Ash-Slag Foam Concrete			
Name	Zhichun WU	Title	Master Student	
Affiliation	Zhengzhou University Email: 15839192530@163.com			
Research Field	Environment-friendly technology / Strengthening resilience			
Period of Internship	Online: October 15 Onsite: November 1, 2024-March 24, 2025			
Host Professor	Gaochuang CAI			
Affiliation	IROAST Email: cai@kumamoto-u.ac.jp		Title	Associate Professor

During my internship at Kumamoto University, I gained a great deal and sincerely thank the university for providing this invaluable opportunity for learning and exchange. Here, I met many outstanding students and researchers who offered tremendous support and encouragement during my visit, allowing me to truly experience the rich academic atmosphere and caring human touch of the institution. Over the course of the five-month internship, I delved into research on the performance of alkali-activated fly ash and slag foam concrete, attempted to write a review, and processed some experimental data. Additionally, I was fortunate to participate in an international academic symposium, where I had in-depth discussions with distinguished scholars from around the world, which proved to be immensely beneficial. In the final two months, I actively joined Professor Cai's team in experimental research, engaging deeply in the project and gaining valuable hands-on experience. Next, I will provide a detailed account of these work activities.

1. Writing a Review of Background and Research objectives in the Field of Foamed Concrete

1.1 A Review of Foamed Concrete

Traditional foam concrete uses cement and mineral admixtures as the main raw materials, adds an appropriate amount of admixtures and water, and the gas generated by foaming through physical or chemical methods is evenly mixed into the concrete slurry. After mixing, pouring, molding and curing, etc. The process forms an inorganic porous lightweight concrete material that is rich in evenly distributed air bubbles. The material has good thermal insulation, sound absorption and earthquake resistance, and is easy to process. However, traditional foam concrete has significant shortcomings. First, due to the large amount of foam added and the high water-cement ratio used in its preparation process, when Portland cement is used as the gel material, the viscosity of the slurry is not enough. The initial setting time is long, and the foam cannot be fixed in the early slurry in time, resulting in internal foam defoaming. The foam concrete material is prone to collapse, low strength, and large drying shrinkage in the early stage. Secondly, using cement as the gel material will cause a lot of energy consumption and environmental pollution, and too much natural river sand will also lead to a waste of resources. In the context of sustainable urban development, finding cementitious materials that can replace cement, meet the needs of the foam concrete industry, and have the advantages of low price, green and environmental protection has become a hot research topic in the construction industry.

Compared with traditional silicate cement, alkali-activated foamed concrete has its unique advantages. Alkali-activated foamed concrete is an inorganic lightweight porous concrete formed

by mixing foam into alkali-activated materials and uniformly stirring and curing. Alkali-activated materials refer to materials with cementitious properties made by the reaction of alkaline activators (such as NaOH, Na_2O, $Ca(OH)_2$, KOH, etc.) with aluminum silicate solids with potential volcanic ash activity or potential hydraulic properties (such as fly ash, slag, recycled micropowder, etc.). The principle is that the ionic bonds and covalent bonds such as Si-O-Si bonds, Al-O-Al bonds, and Al-O-Si bonds in slag, fly ash, and recycled micropowder are easily broken in an alkaline environment, accelerating the disintegration of the structure and making itself hydraulically hard. The biggest difference between alkali-activated foamed concrete and traditional foamed concrete is that the former does not require any cement clinker, which can not only make full use of industrial products to save energy, but also minimize the emission of CO_2 during cement production and use. Secondly, compared with ordinary Portland cement, alkali-activated foamed concrete has better performance, with advantages such as high early strength, rapid coagulation and hardening, high temperature resistance, acid and alkali corrosion resistance, etc. Finally, the alkaline activator used in the preparation process can make the fresh slurry have a larger viscosity, which can better bind the bubbles inside the slurry, forming a large number of closed and uniform pores, so that the foamed concrete has better thermal insulation performance. Therefore, the use of alkali-activated cementitious materials instead of ordinary Portland cement to prepare foamed concrete can not only achieve or even exceed the lightweight, high strength and thermal insulation performance of cement-based foamed concrete, but also reasonably utilize solid waste to achieve the goals of low energy consumption and low carbon emissions, and has broad development prospects.

At present, some scholars at home and abroad have conducted a series of in-depth studies on alkali-activated foamed concrete. Joon-Woo et al. studied the changes in the strength characteristics of lightweight aggregate blast furnace slag concrete by changing the dosage of alkali activator and the replacement amount of kaolin. The results showed that the compressive strength of all specimens at 28 days of age was above 28 MPa, proving the feasibility of using alkali-activated lightweight concrete as lightweight concrete. Abdullah et al. studied alkali-activated fly ash foamed concrete with a density between 1200 kg/m^3 and 1700 kg/m^3. The experimental results showed that its compressive strength can reach up to 17.60 MPa. Duxson et al. showed that the cementitious material prepared by activating fly ash with a strong alkaline solution has the advantages of high strength and low drying shrinkage, and compared with traditional Portland cement, it can reduce CO_2 emissions by about 80% during the preparation process. Rezaei et al. poured alkali-activated fly ash/alumina silicate foamed concrete with a density between 600 and 1100 kg/m^3. The experiment found that the compressive strength of the prepared foamed concrete was positively correlated with the density grade, and as the density grade decreased, the drying shrinkage rate of the foamed concrete also decreased. Shubham et al. studied the effects of gel material composition, silica modulus, alkali content and water-binder ratio on the physical stability, hydrolysis stability and mechanical stability of alkali-activated fly ash-slag foamed concrete, and gave the optimal ratio of the above parameters to obtain the maximum benefit. Zhou et al. partially replaced fly ash with ground blast furnace slag to synthesize a high-strength alkali-activated foamed concrete, and conducted compressive strength and thermal conductivity tests. The results showed that the prepared alkali-activated fly ash/slag foamed concrete not only has a considerable thermal conductivity, but also has excellent compressive strength, which provides a basis for its wide application in thermal insulation materials. At present, drying shrinkage and frost resistance, as important research contents of alkali-activated foamed concrete, have not been thoroughly studied at present. In this paper, alkali-activated foamed concrete was prepared by physical foaming with slag powder and fly ash as raw materials and

water glass-NaOH aqueous solution as activator. The effects of different water-binder ratio, fly ash content and alkali equivalent on compressive strength, water absorption, drying shrinkage and frost resistance of foamed concrete were analyzed, which can provide help for promoting the application of alkali-activated foamed concrete in industry.

It is not difficult to find from the above research that most researchers' research on the macroscopic properties of alkali-activated foamed concrete is mainly focused on density, mechanical properties (such as compressive strength, flexural strength, etc.), water absorption, thermal conductivity, etc. To date, few studies have comprehensively investigated the macroscopic properties and microstructure of alkali-activated fly ash−slag foamed concrete or established the correlations between them. Therefore, this study employs fly ash and slag as raw materials, using water glass and NaOH as alkaline activators, to produce alkali-activated fly ash−slag foamed concrete via a physical foaming method. Polypropylene fiber and hydrophobic agent dosages are used as experimental variables to optimize the initial mix, yielding a high-performance material. The study systematically examines the effects of polypropylene fibers and hydrophobic agents on the mechanical properties, thermal properties, water absorption, hydration products, and microstructure, with the aim of establishing correlations between the macroscopic properties and the microstructure. The resulting alkali-activated foamed concrete not only fully exploits the inherent reactivity of fly ash and slag, but also meets the production requirements in terms of mechanical performance, stability, thermal insulation, and water absorption.
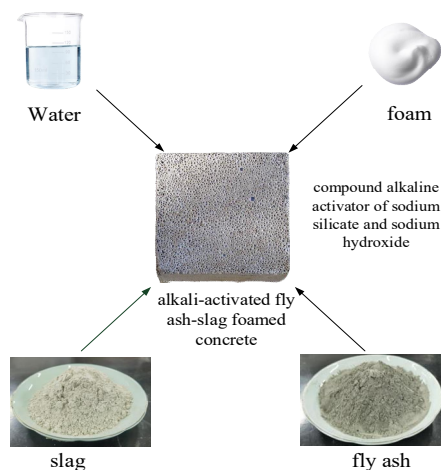

Figure 1. Composition of Alkali-Activated Fly Ash-Slag Foam Concrete

Figure 2. Cured Specimens of Alkali-Activated Fly Ash-Slag Foamed Concrete

1.2 Research objectives

(1) The study investigates the effects of varying polypropylene fiber and hydrophobic agent

contents on the early-stage stability and hardened properties—such as mechanical strength, thermal performance, water absorption, drying shrinkage, and frost resistance of alkali-activated foamed concrete.

(2) By adjusting the amounts of polypropylene fiber and hydrophobic agent, an optimized initial mix design for high-performance alkali-activated foamed concrete is developed. The resulting material not only effectively utilizes the inherent reactivity of raw materials like fly ash and slag but also meets the required standards for mechanical strength, stability, thermal insulation, and water resistance essential for practical applications. This optimization establishes a correlation between material composition, macroscopic properties, microstructure, and reaction products.

2. Experimental data processing and analysis

2.1 Design of Experiments

In this study, an orthogonal design method was applied to establish nine experimental groups, with polypropylene fiber and hydrophobic agent contents as independent variables. The alkali-activated foamed concrete was designed with a target dry density of 1100 kg/m³. A composite foaming agent was employed, and the mix proportion was designed with a water-to-binder ratio of 0.45, a silica modulus of the activator at 1.2, an alkali dosage (as Na_2O equivalent) of 4%, a superplasticizer dosage of 0.5%, and a binder proportion of 20% fly ash and 80% slag. For instance, AAFC-0.5-1% indicates a mix in which the hydrophobic agent content accounts for 0.5% of the binder mass, while the polypropylene fiber content is 1%. The detailed experimental design is presented in the following table.

Table 1. Experimental Design for Alkali-Activated Fly Ash-Slag Foamed Concrete

Group	Hydrophobic Agent Content (% of Binder Mass)	Polypropylene Fiber Content (% of Binder Mass)	Water-to-Binder Ratio	Activator Modulus	Alkali Dosage (Na_2O Equivalent, %)
AAFC-0-1%	0	1%	0.4	1.2	4%
AAFC-0.5-1%	0.5%	1%	0.4	1.2	4%
AAFC-1-1%	1%	1%	0.4	1.2	4%
AAFC-1.5-1%	1.5%	1%	0.4	1.2	4%
AAFC-2-1%	2%	1%	0.4	1.2	4%
AAFC-1.5-0%	1.5%	0%	0.4	1.2	4%
AAFC-1.5-0.5%	1.5%	0.5%	0.4	1.2	4%
AAFC-1.5-1.5%	1.5%	1.5%	0.4	1.2	4%
AAFC-1.5-2%	1.5%	2%	0.4	1.2	4%

2.2 Experimental Data Analysis (Compressive Strength)

The figure presents the effect of different polypropylene fiber dosages on the compressive strength of alkali-activated fly ash-slag foamed concrete at 7 and 28 days. The results indicate that compressive strength initially increases with fiber content, reaching a peak of 7.68 MPa at 1% fiber dosage, before subsequently decreasing. These findings suggest that an appropriate amount of polypropylene fibers enhances matrix strength through mechanisms such as bridging microcracks, inhibiting crack propagation, and refining the pore structure. At lower dosages ($\leqslant 1\%$ of the binder mass), the uniform fiber distribution within the matrix improves interfacial transition zone bonding, facilitates stress transfer, and reduces porosity, thereby enhancing material density and compressive strength. However, when the fiber content exceeds 1.5%, the compressive strength decreases, likely due to fiber agglomeration, which disrupts the uniformity of the pore structure, increases porosity, and reduces matrix density. This effect results in localized stress concentrations, weakening the overall mechanical performance. Additionally, excessive fiber incorporation may interfere with the alkali activation process, reducing interfacial bonding strength and further impairing material performance. Therefore, a polypropylene fiber dosage of 1% of the binder mass is optimal for this system, as it provides the best balance between fiber reinforcement and pore structure refinement, ensuring superior mechanical properties in alkali-activated foamed concrete.
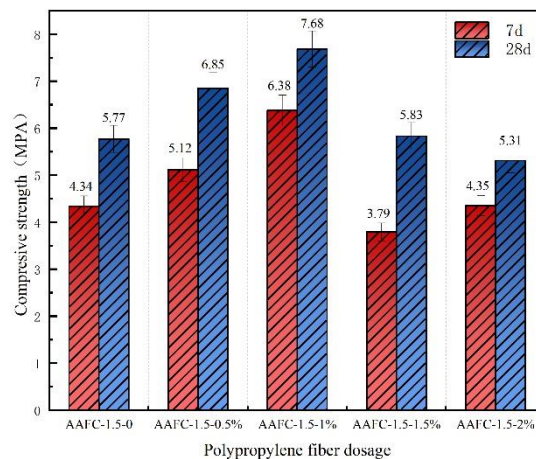

Figure 3. Compressive Strength of Alkali-Activated Fly Ash-Slag Foamed Concrete

3. Participate in an Academic Conference

The International Mini-Symposium on Sustainable and Carbon-Neutral Structures and Materials, held at Kumamoto University, provided a high-level academic exchange platform for scholars worldwide to explore cutting-edge technologies in sustainable building materials and structural engineering. The symposium brought together renowned international experts to engage in in-depth discussions on key topics, including bio-based materials, bamboo structural engineering, sustainable composite materials, modular connection systems, reusable low-carbon building components, as well as numerical analysis and AI-driven modeling of reinforced concrete (RC) structures. These discussions encompassed several critical technological directions within the fields of civil and structural engineering. A distinctive feature of this symposium was the emphasis on multidisciplinary integration and the application of advanced technologies as essential drivers for sustainable building development. The presentations not only covered material-level innovations, such as structural optimization of bamboo-based components and the

carbon-neutral development of bio-based construction materials, but also addressed structural system research, including sustainable steel-concrete composite flooring systems, reusable connection systems, and experimental and numerical studies on RC walls under cyclic loading. Moreover, the application of artificial intelligence and data-driven methods in seismic design and structural safety assessment emerged as one of the highlights of this symposium, offering novel insights into the optimization and performance prediction of future structural systems.This symposium clearly demonstrated that low-carbon, sustainable, and intelligent solutions are the prevailing trends in the future of civil and structural engineering. Experts at the event underscored that achieving global carbon neutrality requires greater emphasis on the development of innovative low-carbon materials, the optimization of efficient structural systems, the reduction of life-cycle carbon footprints in construction, and the integration of digital modeling with intelligent analytical tools. Throughout the discussions, scholars engaged in comprehensive dialogues that spanned theoretical analysis, experimental studies, and engineering practice, effectively showcasing the latest research advancements and technical challenges in the field.

Attending this symposium has significantly broadened my understanding of sustainable building materials and structures, reinforcing the idea that future research should not only focus on enhancing the performance of individual materials or structures but also consider multiple factors such as environmental sustainability, resource efficiency, durability, and structural safety. The opportunity for academic exchange was invaluable, deepening my insights into the evolution of structural engineering within the carbon-neutral framework and inspiring new directions for future research.

Figure 4. Participate in an Academic Conference

3. Conducting experiments within Dr. Cai's research group

During my participation in Dr. Cai's research team, I was involved in the preparatory work for the concrete wall loading test. My tasks included attaching strain gauges to collect strain data, marking measurement points and crack observation grids on the wall, and setting up displacement and load sensors. Additionally, I inspected the bonding quality of fiber-reinforced composite materials to ensure the effectiveness of reinforcement. Furthermore, I assisted in calibrating the loading equipment and checking the stability and safety of the experimental system to ensure the smooth execution of subsequent tests.
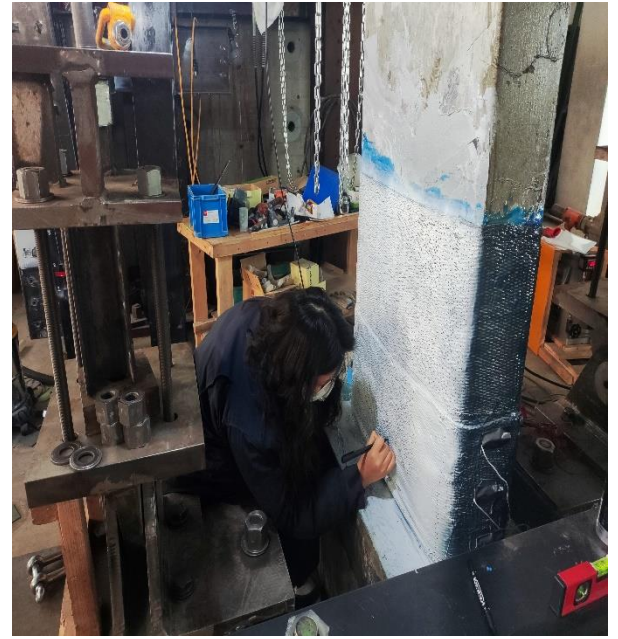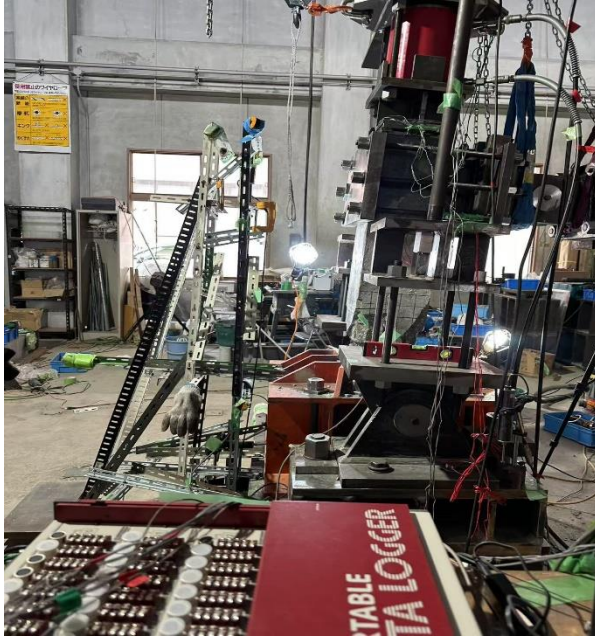
Figure 5. Conducting experiments

No. 2-7-8	Post-earthquake damage assessment of RC structures based on deep learning			
Name	Dan YUAN	Title	Master Student	
Affiliation	Xi'an University of Technology, China Email: yuandan1998@qq.com			
Research Field	Data science and AI			
Period of Internship	Online: June 15, 2024 - July 31, 2024 Onsite: August 1, 2024 - September 6, 2024			
Host Professor	Gaochuang CAI	Title		Associate Professor
Affiliation	IROAST Email: cai@kumamoto-u.ac.jp			

During my internship at IROAST of Kumamoto University, I gained a lot and received valuable guidance from Professor Cai. I also had the opportunity to meet many outstanding students and staff, all of whom provided great support for my studies and daily life. During my time at Kumamoto University, I mainly accomplished two tasks that were deeply connected to the topic of "Using Machine Learning Algorithms to Solve Key Problems in the Field of Earthquake Resistance". First, I completed the paper titled *Quantitative Evaluation of the Relationship between Damage Evolution and Seismic Capacity of RC Columns: A Deep Learning-Based Perspective*, which was supervised by Dr. Cai and worked with Mr. Wen (Cai lab member) before I came. Second, I utilized 3D reconstruction to assess the surface damage in concrete columns subjected to earthquake loads. I will now provide a detailed introduction to these two areas of work.

1. The first project
1.1 Background and data preparation
The first part of my work involved completing the experiments and writing for a paper, which has already been submitted to the journal Advanced Engineering Informatics. At the beginning of my internship, I received experimental data from previous experiments conducted in the lab. In these experiments, a lateral offset angle was applied to concrete columns, and photos were taken at different offset angles. The raw image data from the experiment is shown in Figure 1.

Figure1. Original image data

With this data in hand, my goal was to use machine learning methods to assess the damage to the concrete columns. To achieve this, I employed deep-learning semantic segmentation algorithms to identify the cracks and spalling in the images. First, I constructed a dataset to train and optimize the model's parameters. I used Labelme annotation software for labeling, carefully outlining all the cracks in the images with polygons. This process had to be as detailed as possible because the more precise the annotations, the better the model's performance would be. After completing the crack annotations, I created a "crack" label and then proceeded to annotate the spalling areas, creating a "spalling" label after finishing. I repeated this process until all the photos, approximately 200 images, were annotated. Once the annotations were complete, I split the data into training and testing sets, using 80% of the data for training and 20% for testing. At this point, the dataset was ready for use.

1.2 Algorithm

To address this problem, I proposed the KU-net architecture, which integrates both deep learning and machine learning components. The deep learning component focuses on extracting semantic segmentation information from the images, specifically identifying and delineating cracks and spalling areas. Once these defects are extracted, their characteristics, such as crack width and length and the area of spalling, are quantified. The machine learning component then correlates these quantified damage metrics with structural mechanics indicators, such as Degradation of Applied Force, Degradation of Stiffness, and the Equivalent Viscous Damping Coefficient, enabling a precise evaluation of the damage. This combined approach leverages the strengths of both deep learning for image analysis and machine learning for structural assessment to accurately determine the extent of damage to the concrete columns.
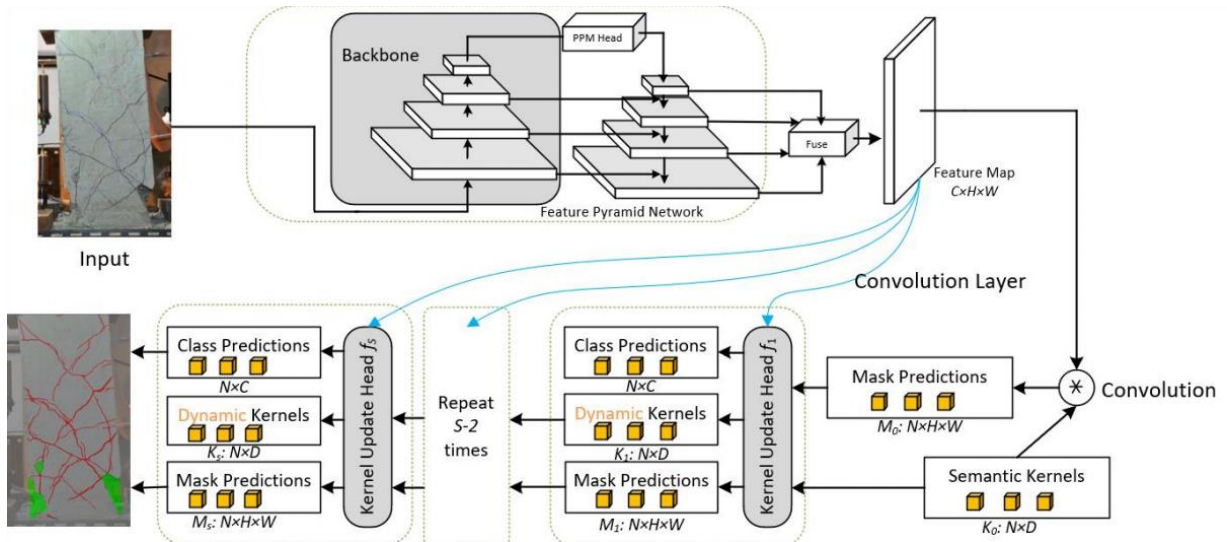
Figure2. The overall architecture of the feature extractor network

As shown in the figure, the model utilizes a PPM (Pyramid Pooling Module) combined with an FPN (Feature Pyramid Network) architecture to generate a more effective feature map. This feature map integrates multi-level semantic features. To further enhance this feature map, dynamic convolution is introduced. Dynamic convolution kernels automatically update the weight parameters of the convolutional sets based on the content of the images, allowing for better extraction of cracks and spalling. The semantic segmentation results are illustrated in Figure 3. After extracting the information related to cracks and spalling, this data is quantified. To correlate these quantified metrics with structural mechanics parameters, three machine learning algorithms─MLP (Multi-Layer Perceptron), XGBoost, and SVM (Support Vector Machine)─are employed to learn the mapping relationships.

1.3 Experiment result

The experimental results of the deep learning component are shown in Figure 4. To demonstrate the superiority of our proposed method, I compared it with three other semantic segmentation algorithms. The experimental results indicate that KU-net achieved the best performance among the compared methods.
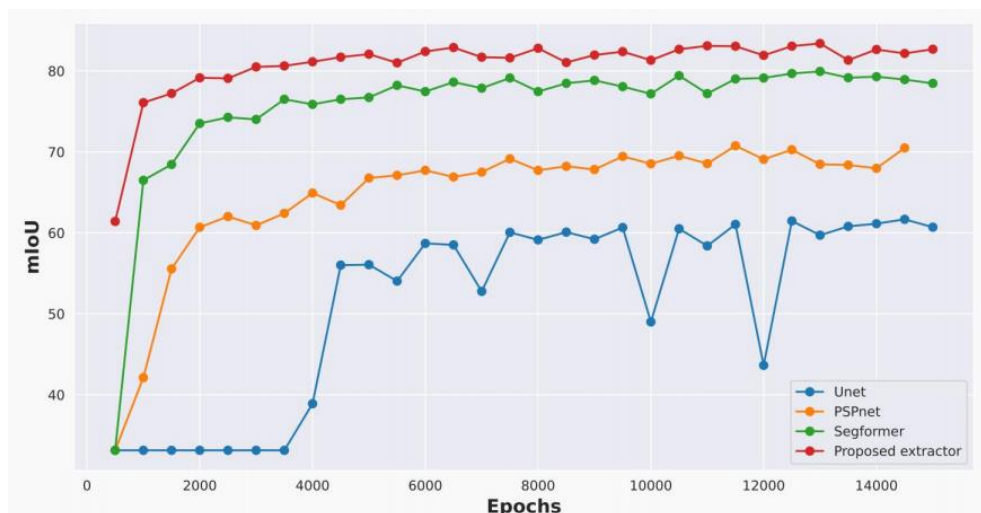

Figure 4. Comparison of mIoU results with different algorithms

The experimental results of the machine learning component are presented in Table 1. The SB-set and SBD-set represent the testing results on two different test datasets. The experiments show that the Mean Squared Error (MSE), which indicates the error between the predicted values and the ground truth, was the lowest for SVM. This indicates that SVM outperformed the other methods, and as a result, I chose to retain SVM as the algorithm for the machine learning component of KU-net.

Table 1. Comparison of the metrics of the machine learning module

Algorithm	SB-set		SBD-set	
	MSE	MAE	MSE	MAE
SVM	0.0075	0.0578	0.0333	0.1216
MLP	0.0078	0.0633	0.0378	0.1295
XGBoost	0.0090	0.0670	0.0419	0.1450

2. The second project
2.1 Background and data preparation

The second project during my internship aimed to more comprehensively assess the damage to concrete columns. In the previous project, I used photos taken from a single angle for evaluation, which was not sufficient for a complete analysis. Therefore, I explored the possibility of taking multiple photos of a concrete column from different angles, as shown in Figure 5. In the experiment, I captured photos from four different angles and used deep learning algorithms to reconstruct a 3D model of the column.

Figure 5. The data of the experimental subjects captured from different angles is used for 3D reconstruction.

Specifically, during the experiment, We prepared eight different concrete columns, including two NC (normal concrete) columns and six MRC (modified reinforced concrete) columns. In the experimental process, we first fixed the first concrete column and then applied horizontal push-pull forces to it, causing it to tilt at a specific angle. This simulates the conditions during an earthquake. As the tilt angle increases, the damage to the concrete becomes more severe. At each different loading point, we took four photos of the concrete column from four angles. These four photos provide visual information from all angles of the concrete column, ensuring that no side is left unphotographed or obstructed. The images mentioned above are used for model training. Additionally, at each loading point, we recorded a video around the concrete column. The purpose of this is to simulate real-world scenarios: due to earthquake damage to buildings, people may not

be able to approach the concrete columns inside the building to observe the damage, so a remote-controlled drone can be used to record a video. Using this video, we can then employ the trained model to reconstruct a 3D model for more precise subsequent calculations. The data collection scene is shown in Figure 6.
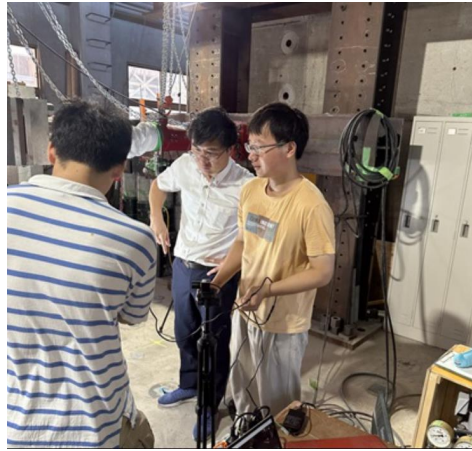

Figure 6. Data collection under the guidance of Professor Cai

The reconstructed 3D image includes cracks and spalling information, allowing for a more detailed quantification of the damage to the concrete column. This necessitated building a deep learning network for 3D reconstruction. Using multi-angle photos for 3D reconstruction enables a more comprehensive evaluation of the damage and condition of the concrete columns. In the previous project, I did not perform camera calibration, which is essential because the inherent distortion in photos caused by the pinhole camera model could lead to inaccurate experimental data. Therefore, before collecting experimental data for this project, I used OpenCV to correct the original images, ensuring the accuracy of the results. Specifically, I first prepared the calibration board and completed the code, I ran the code to extract the camera's internal and external parameters, distortion coefficients, and other parameters. Then, I used the image correction method provided by OpenCV to correct the original distorted image.

2.2 Experiment result

In this project, I first developed a 3D reconstruction algorithm capable of creating a 3D model from an input image. Specifically, the 3D results generated by this model are explicit mesh representations rather than implicit probability representations, which facilitates more accurate damage calculation. The model utilizes Zero123 to generate multi-view images and then performs 3D reconstruction using a signed distance function (SDF). This approach allows for better geometry, more consistent 3D results, and a closer adherence to the input images within a shorter time frame. Since the experiment with concrete columns has not been completed yet, I tested the network using pre-trained model parameters. The experimental results are shown in Figure 7.
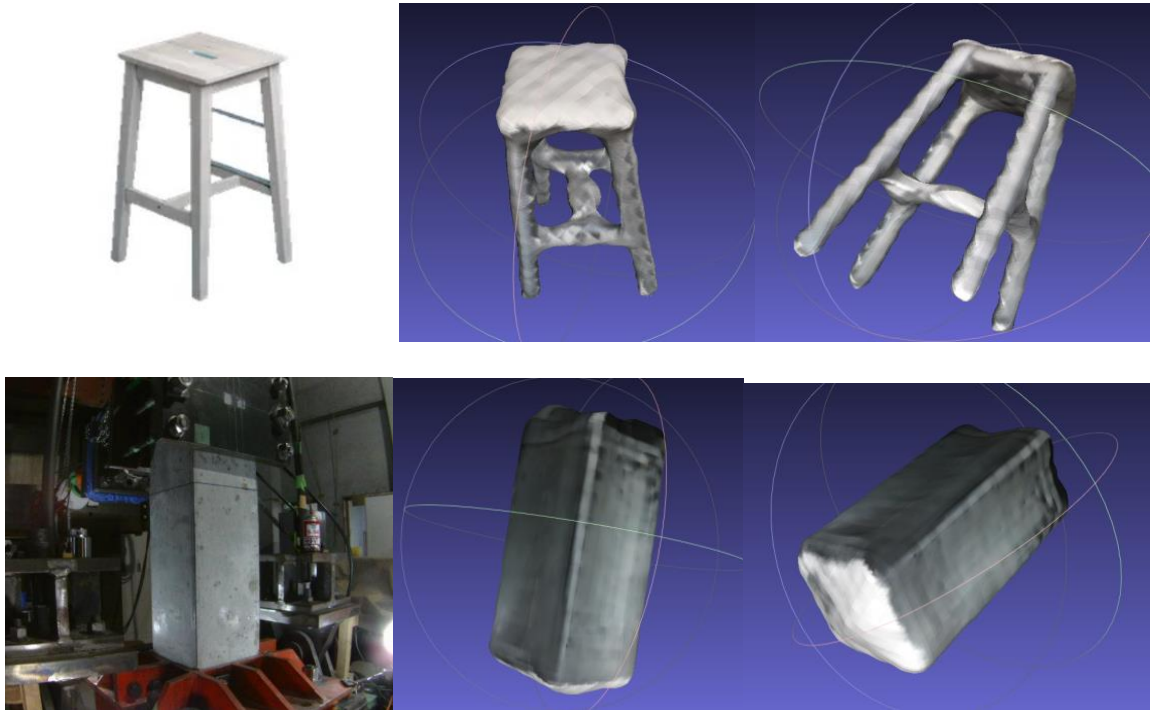
Figure 7. Experiment result of 3D reconstruction

3. Other experiments

For the previously completed crack extraction task, the analysis involved annotating several metrics from an image, including the damage area rate, concrete spalling area rate, hinge height, number of diagonal cracks, maximum transverse crack width and its coordinates, and the instantaneous failure model. I first input the images into the feature extraction network of our proposed KU-net to obtain a semantic segmentation map of cracks and spalling (as shown in Figure 8).

Figure 8. The semantic segmentation result

When annotating key points, I first preprocessed the segmented images by setting the pixel values for cracks to 2, spalling to 1, and background information to 0. Then I performed morphological operations on the semantic images, specifically applying closing operations to eliminate small crack holes and gaps. For the maximum transverse crack, I first extracted pixels with values 1 and 2, then calculated the contours for cracks and spalling separately. Each crack's information was

stored in a list, and then each crack was analyzed to determine its type. If a crack's transverse length is greater than its longitudinal length, it is classified as a transverse crack. For the extracted transverse cracks, I calculated the area and coordinates and displayed the one with the largest area. For the damage area ratio analysis, My approach was to apply U-Net for semantic segmentation to segment only the concrete column and the annotated horizontal lines. After re-annotating the dataset, I trained the model. The trained model can precisely identify the horizontal lines and the entire concrete column in the input images. Based on the segmentation of these horizontal lines, I calculated the area of the concrete column at 250 mm and 125 mm. Then, using the segmentation maps for cracks and spalling, I computed the areas of cracks and spalling. By overlaying the results of the two semantic segmentation maps, I obtained the area ratios. The experimental results are shown in Figure 8.
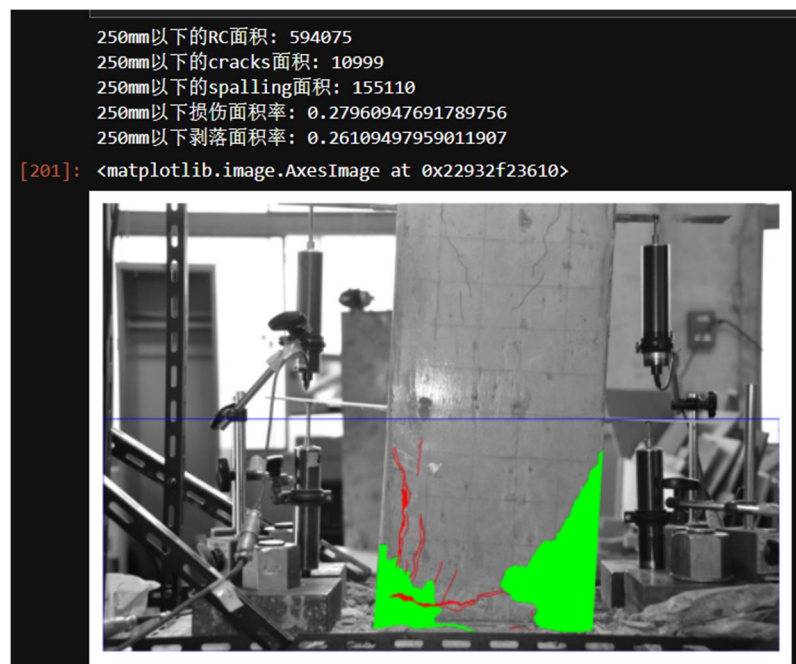

Figure 9. Analysis results of the damage area ratio

4. Future plan

After finishing this internship, I will participate in a six-month exchange program at Kumamoto University. I plan to complete the remaining experiments of Project 2 and write a new paper. Over the next six months, I also aim to complete the experiments and write two additional papers. In the longer term, I plan to pursue research at the intersection of machine learning and civil engineering, pursue a Ph.D., and stay in academia to continue my research.

5. Outcomes

Y. Wen, D. Yuan, G. Cai*. Quantitative Evaluation of the Relationship Between Damage Evolution and Seismic Capacity of RC Columns: A Deep Learning-Based Perspective Method. *Advanced Engineering Informatics* (IF:8.0, Submitted)