

Evaluasi Pemodelan Augmentasi Data Sifat Mekanik Aluminium Menggunakan *Generative Adversarial Networks*

¹⁾Desmarita Leni, ²⁾Ade Usra Berli, ³⁾Dytchia Septi Kesuma, ⁴⁾Haris*, ⁵⁾Ruzita sumiati
(^{1,2,3})Teknik Mesin, Fakultas Teknik, Universitas Muhammadiyah Sumatera Barat
(^{4,5}) Teknik Mesin, Politeknik Negeri Padang
*Email:haris@pnp.ac.id

Diterima: 14.10.2023, Disetujui: 06.04.2024, Diterbitkan: 18.04.2024

ABSTRACT

Materials informatics is a new approach in material science that integrates information technology and material science to optimize the discovery of new materials more efficiently and innovatively. In materials informatics, experimental and simulation data are combined with data-driven methods such as big data, data augmentation, and machine learning to gain a deeper understanding of material properties. However, limitations in the availability of samples with desired characteristics and the lack of accurate experimental data pose challenges in materials informatics. In this study, we attempt to address these challenges by modeling the augmentation of mechanical properties of aluminum using Generative Adversarial Networks (GAN). GAN is used to generate synthetic data of aluminum's mechanical properties that closely resemble experimental data. This modeling is trained using experimental testing data consisting of aluminum's mechanical properties and chemical elements in the alloy, obtained from the material database. The dataset comprises 9 chemical element variables in the aluminum alloy and 2 mechanical property variables. The synthetic data generated from the modeling is evaluated using descriptive statistics, Pearson correlation, and Kolmogorov-Smirnov (KS) test to assess the extent to which the synthetic data resembles the original data. The evaluation results indicate that the distribution of synthetic data is similar to the original data. The Pearson correlation results show that most variables of chemical elements and mechanical properties of aluminum in the synthetic data have a correlation that is quite similar to the original data. The KS test results also indicate that the distribution of synthetic data does not significantly differ from the distribution of the original data. This indicates that the synthetic data generated has a high resemblance to the experimental data, enabling its use in materials informatics research. Thus, modeling the augmentation of aluminum's mechanical property data using GAN provides a significant contribution to expanding data availability in material science.

Keywords: *Modeling, Augmentation, Mechanical properties, Aluminum, Generative Adversarial Networks*

ABSTRAK

Materials informatics adalah pendekatan baru dalam ilmu material yang mengintegrasikan teknologi informasi dan ilmu material untuk mengoptimalkan proses penemuan material baru yang lebih efisien dan inovatif. Dalam materials informatics, data eksperimen dan simulasi digabungkan dengan metode berbasis data seperti big data, augmentasi data, dan machine learning untuk mendapatkan pemahaman yang lebih mendalam tentang sifat material. Namun, keterbatasan dan ketersediaan sampel dengan karakteristik yang diinginkan, serta minimnya data eksperimen yang akurat, menjadi tantangan tersendiri dalam ilmu materials informatics. Padapenelitian ini, kami mencoba mengatasi tantangan tersebut dengan melakukan pemodelan augmentasi data sifat mekanik aluminium menggunakan Generative Adversarial Networks (GAN). GAN digunakan untuk menghasilkan data sintesis sifat mekanik aluminium yang sangat mirip dengan data eksperimen. Pemodelan ini dilatih menggunakan data hasil pengujian eksperimen yang terdiri dari sifat mekanik aluminium dan unsur kimia pada paduan, yang diperoleh dari database material. Dataset terdiri dari 9 variabel unsur kimia pada paduan aluminium dan 2 variabel sifat mekanik. Data sintesis yang dihasilkan dari pemodelan dievaluasi menggunakan statistik deskriptif, korelasi Pearson, dan uji Kolmogorov-Smirnov (KS)

untuk menilai sejauh mana data sintetis mirip dengan data asli. Hasil evaluasi menunjukkan bahwa distribusi data sintetis memiliki kesamaan dengan data asli. Hasil korelasi Pearson menunjukkan bahwa sebagian besar variabel unsur kimia dan sifat mekanik aluminium pada data sintetis memiliki korelasi yang cukup mirip dengan data asli. Hasil dari uji KS juga menunjukkan bahwa distribusi data sintetis tidak berbeda secara signifikan dengan distribusi data asli. Hal ini mengindikasikan bahwa data sintetis yang dihasilkan memiliki kemiripan yang tinggi dengan data eksperimental, yang memungkinkan penggunaan data sintetis dalam penelitian materials informatics. Dengan demikian, pemodelan augmentasi data sifat mekanik aluminium menggunakan GAN memberikan kontribusi penting dalam memperluas ketersediaan data dalam ilmu material.

Kata Kunci: *Pemodelan, Augmentasi, Sifat mekanik, aluminium, Generative Adversarial Networks*

I. Pendahuluan

Aluminium (Al) adalah salah satu logam paling banyak digunakan dalam industri modern. Logam ini memiliki berbagai sifat mekanik yang sangat diperlukan dalam berbagai aplikasi teknik, seperti dalam pembuatan kendaraan, pesawat terbang, dan peralatan rumah tangga. Aluminium memiliki bobot yang ringan, sifat tahan korosi, dan konduktivitas panas yang baik, sehingga menjadikannya pilihan yang sangat baik dalam banyak aplikasi (Ashkenazi, 2019). Sifat mekanik aluminium, seperti *Yield Strength* (kekuatan luluh) dan *Tensile Strength* (kekuatan tarik), sangat penting dalam proses perancangan dan pengembangan produk (Sert et al., 2019). Pemahaman yang baik tentang sifat mekanik aluminium, dapat membantu pelaku industri dalam perancangan produk yang lebih optimal seperti memaksimalkan kekuatan, elastisitas, keuletan, dan kekerasan. Hal ini dapat menghindari kegagalan struktural, dan memaksimalkan umur pakai produk (Arif et al., 2003). Sifat mekanik aluminium dipengaruhi oleh berbagai faktor seperti unsur kimia, struktur mikro, dan perlakuan panas (Morini et al., 2019). Penambahan kandungan unsur kimia dalam paduan aluminium dapat mempengaruhi sifat mekanik aluminium, seperti penambahan unsur Magnesium (Mg) pada paduan aluminium dapat meningkatkan kekuatan dan kekerasan (Tsai et al., 2009). Pada studi lain (Hwang et al., 2009) juga ditemukan bahwa penambahan unsur (Mg) juga dapat meningkatkan ketahanan korosi dan kemampuan bentuk. Hal ini

mengindikasikan bahwa penambahan unsur kimia yang berbeda pada paduan aluminium dapat mempengaruhi sifat mekanik aluminium. Pengujian aluminium pada umumnya masih dilakukan secara manual menggunakan mesin uji tarik yang membutuhkan biaya yang cukup besar, waktu yang lama dan tidak ramah lingkungan sebab menghasilkan limbah dari sampel aluminium. Kendala lainnya yang dihadapi dalam pemahaman sifat mekanik aluminium adalah keterbatasan tersedianya sampel dengan karakteristik yang diinginkan dan minimnya data eksperimental yang akurat (Leni, Yermadona, et al., 2023).

Materials informatics merupakan suatu pendekatan baru dalam ilmu material yang mengintegrasikan teknologi informasi dan ilmu material untuk mengoptimalkan proses penemuan material baru yang lebih efisien dan inovatif (Frydrych et al., 2021; Rajan, 2015). Dalam *materials informatics*, data eksperimental dan simulasi digabungkan dengan metode berbasis data seperti big data, augmentasi data, dan *machine learning* untuk mendapatkan pemahaman yang lebih mendalam tentang sifat material (Blaiszik et al., 2019). Augmentasi data adalah teknik yang digunakan untuk memperbanyak data dengan cara memodifikasi data yang telah ada menggunakan algoritma machine learning (Yu et al., 2021). *Generative Adversarial Networks (GAN)* merupakan salah satu jenis arsitektur jaringan saraf tiruan (*neural network*) yang digunakan untuk menghasilkan data sintetis yang realistis. GAN terdiri dari dua model utama yaitu, generator dan *discriminator*.

Generator menghasilkan data sintetis yang menyerupai data asli, sedangkan discriminator membedakan antara data asli dan data sintetis. Kedua model ini saling bersaing dalam proses pelatihan, di mana generator berusaha menghasilkan data sintetis yang semakin mirip dengan data asli, sedangkan *discriminator* berusaha membedakan antara data asli dengan data sintetis dengan semakin baik. Proses pelatihan berakhir ketika generator menghasilkan data sintetis yang sangat mirip dengan data asli, sehingga discriminator tidak dapat membedakan keduanya (Wang et al., 2017).

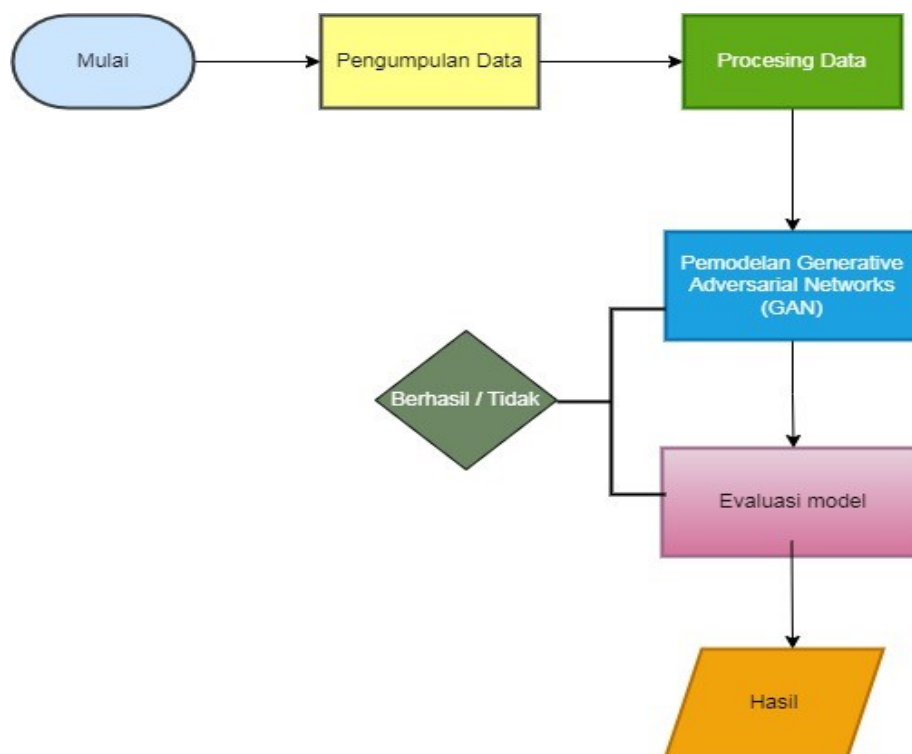
Penelitian terdahulu telah membahas penggunaan GAN dalam pembuatan data sintetis dalam bentuk tabel terkait dengan umur leleh multiaxial pada berbagai jenis material, seperti aluminium, kuningan, dan baja tahan karat. Temuan dari penelitian tersebut mengindikasikan bahwa penggunaan data sintetis ini dapat menghasilkan peningkatan dalam akurasi model *Machine Learning* dalam melakukan prediksi terkait umur leleh material-material tersebut (He et al., 2022). Dalam studi (Marani et al., 2020) tentang prediksi kekuatan beton berkekuatan tinggi, ditemukan bahwa GAN dapat meningkatkan jumlah data experimental yang awalnya 810 sampel dengan 15 fitur masukan dan satu output yaitu, kekuatan tekan beton. Hasil penelitian tersebut menyatakan bahwa GAN dapat meningkatkan volume data asli menjadi 8 kali lipat dan data ini dapat memprediksi 810 sampel data experimental yang tidak ada pada data pelatihan.

Berdasarkan permasalahan yang telah diuraikan diatas dan dampak positif dari

penelitian terdahulu tentang penerapan GAN sebagai pemodelan untuk memperoleh data sintetis sifat material. Penelitian ini bertujuan untuk mengevaluasi pemodelan augmentasi data sifat mekanik aluminium menggunakan GAN. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknik augmentasi data untuk sifat mekanik material, terutama untuk mendukung perkembangan teknologi dan pemahaman yang lebih baik tentang sifat mekanik aluminium.

II. Bahan dan Metode

Penelitian ini merupakan jenis penelitian experimental yang bertujuan untuk menghasilkan data sintetis unsur kimia paduan dan sifat mekanik aluminium dalam bentuk tabel (*tabular*) menggunakan metode *Generative Adversarial Networks (GAN)*. Pemodelan augmentasi data sifat mekanik aluminium ini dibuat menggunakan bahasa pemrograman Python dengan *library Tensor Flow* yang dijalankan pada *google colab*. Pada penelitian ini, data experimental diperoleh dari database material Matmach.com, yang merupakan sebuah website gratis yang menyediakan 70.000 lebih data material termasuk aluminium (Matmach, n.d.). Data yang dikumpulkan ini dijadikan sebagai dataset pelatihan GAN dalam memperoleh data sintetis. Data sintetis yang dihasilkan dari pemodelan GAN dievaluasi menggunakan statistik deskriptif, korelasi pearson, dan Uji KS, untuk mengetahui seberapa mirip data sintetis yang diperoleh dengan data experimental. Tahap penelitian ini dapat dilihat dengan lebih jelas pada gambar 1.



Gambar 1 . Skema penelitian

1. Pengumpulan data

Pada tahap ini, data diambil dari website Matmatch.com yang terdiri dari 9 variabel unsur kimia seperti Aluminium (Al), Magnesium (Mg), Zinc (Zn), Titanium (Ti), Copper (Cu), Manganese (Mn), Chromium (Cr), Iron (Fe), dan Silicon (Si), dengan dua perlakuan panas yaitu, O (*Anil*) dan H (*Strain Hardened*), serta sifat mekanik *Yield Strength (YS)* dan *Tensile Strength (TS)*. Data aluminium yang dikumpulkan memiliki komposisi kimia yang sama dan memiliki kadar persentase yang berbeda, kemudian data ini disimpan dalam bentuk format tabel.

2. Preprocessing data

Preprocessing data adalah tahap persiapan data dalam sebuah penelitian yang meliputi penanganan missing values, pembersihan data dari outlier, encoding variabel kategorikal dan normalisasi data untuk memastikan data siap dianalisis (Agrawal et al., 2014; Leni, 2023). Dalam penelitian ini, normalisasi data dilakukan menggunakan metode *MinMax Scaler*. Metode normalisasi ini adalah yang paling umum digunakan dalam *machine*

learning, dan dapat dihitung menggunakan persamaan 1.

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \dots\dots\dots(1)$$

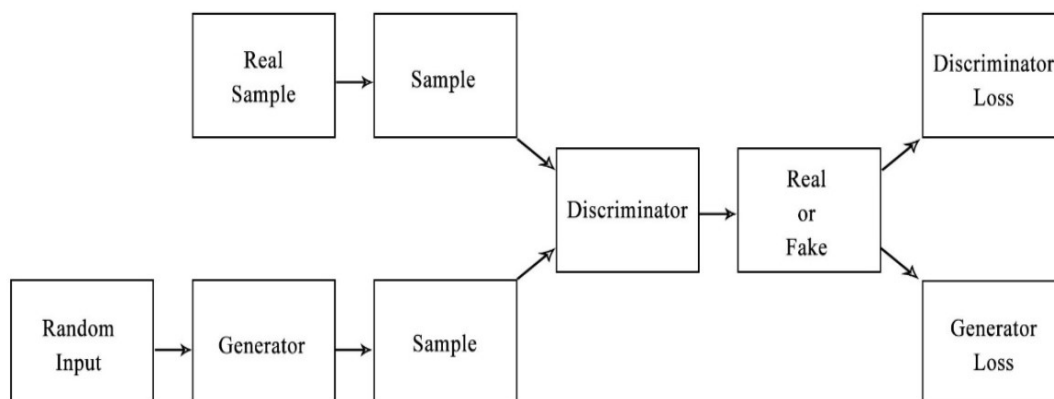
Dimana X adalah data asli, sedangkan Xsc adalah data yang dinormalisasi, tujuan dari normalisasi data adalah untuk menghilangkan bias yang mungkin ada dalam data yang dihasilkan oleh skala yang berbeda pada setiap variabel (Koc & Gurgun, 2022). Seluruh tahapan *preprocessing* data sifat mekanik aluminium dapat dilakukan menggunakan berbagai library dalam Python seperti Pandas, NumPy, dan Scikit-learn.

3. Pemodelan Generative Adversarial Networks (GAN)

Pada tahap ini digunakan *Generative Adversarial Networks (GAN)* untuk memperoleh data sintesis sifat mekanik aluminium. GAN merupakan *algoritma deep learning* yang terdiri dari dua jaringan, yaitu generator dan *discriminator* (Wang et al., 2017). Data sifat mekanik aluminium dan

unsur kimia dijadikan sebagai dataset pelatihan model, dimana generator mencoba untuk menghasilkan data sintesis yang semirip mungkin dengan data asli, sementara

discriminator belajar untuk membedakan antara data asli dan data sintesis. Arsitektur GAN dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur Generative Adversarial Networks (GAN) (Li et al., 2021).

4. Evaluasi model

Data sintesis yang dihasilkan akan dievaluasi menggunakan korelasi pearson, statistik deskriptif, dan Uji Kolmogorov-Smirnov (KS) untuk melihat seberapa mirip data sintesis sifat mekanik aluminium dengan data sebenarnya. Dengan menggunakan kombinasi metode evaluasi ini, dapat dilihat gambaran yang lebih lengkap dan memadai tentang seberapa mirip data sintesis dengan data asli. Korelasi Pearson mengukur hubungan linier antar unsur kimia dengan sifat mekanik aluminium, statistik deskriptif memberikan wawasan tentang statistik keseluruhan variabel data, dan Uji KS mengukur kesamaan distribusi setiap variabel dalam data. Dengan demikian, penggunaan tiga metode evaluasi ini secara bersama-sama dapat memberikan pemahaman yang lebih baik tentang kualitas data sintesis yang dihasilkan oleh GAN. Korelasi pearson dapat dapat dihitung menggunakan persamaan 2 dan Uji Kolmogorov-Smirnov (KS) dapat dihitung menggunakan persamaan 3.

$$r_{xy} = \frac{\sum xy}{(n-1)s_x.s_y} \dots\dots\dots(2)$$

Dimana r_{xy} adalah koefisien korelasi Pearson, $\sum xy$ adalah jumlah perkalian x dan y , n adalah ukuran sampel, x adalah variabel independen, y adalah variabel dependen, dan S adalah standar deviasi (Leni, Earnestly, et al., 2023). Nilai koefisien korelasi berkisar dari -1 hingga 1. Nilai -1 menunjukkan korelasi negatif yang kuat antara kedua variabel, nilai 0 menunjukkan tidak ada korelasi, dan nilai 1 menunjukkan korelasi positif yang kuat (Leni, Muchlisinalahuddin, et al., 2023).

$$D = \max(F1_{(x)} - F2_{(x)}) \dots\dots\dots(3)$$

Dimana D adalah Uji KS, $F1(x)$ adalah fungsi distribusi empiris (ECDF) dari sampel pertama, dan $F2(x)$ adalah fungsi distribusi empiris (ECDF) dari sampel kedua (Berger & Zhou, 2014). Semakin besar nilai D , semakin berbeda kedua distribusi data tersebut. Pada umumnya, uji KS membandingkan distribusi data dengan distribusi teoritis, sehingga distribusi kedua harus dijabarkan. Dalam konteks perbandingan dengan distribusi empiris (data asli), D digunakan untuk mengukur kesamaan atau perbedaan antara distribusi data sintesis dengan distribusi data asli. Semakin kecil nilai D , semakin mirip distribusi kedua data tersebut.

III. Hasil dan Pembahasan

1. Pengumpulan data

Dataset pelatihan model augmentasi data sifat mekanik aluminium diambil dari *website* *matmatch.com*, yang merupakan perpustakaan bahan *online* akses terbuka yang terdiri dari ribuan lembar sifat material. Namun, sebagian besar data sifat material masih tercampur dengan sifat material lainnya sehingga perlu dilakukan penyaringan untuk memperoleh kumpulan data yang sesuai dengan kebutuhan pemodelan augmentasi data.

Dataset aluminium yang diambil terdiri dari 9 variabel unsur kimia dan 2 sifat mekanik aluminium seperti yang dapat dilihat pada tabel 1.

Tabel 1. Statistik dataset Aluminium

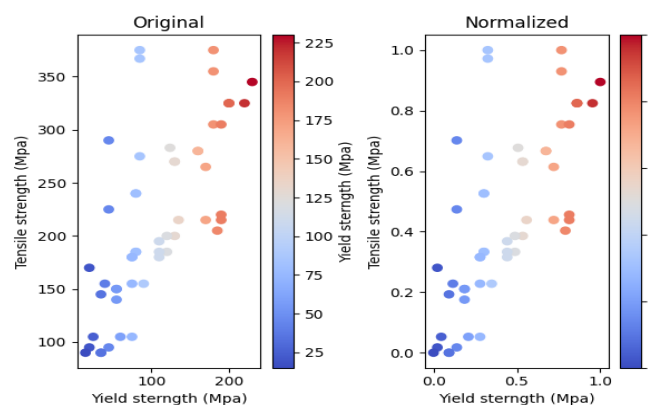
Variabel	Min	Max	Mean
Mg (%)	0	3.1	0.978
Zn (%)	0.05	0.5	0.207
Ti (%)	0	0.25	0.069
Cu (%)	0	0.5	0.130
Mn (%)	0.02	1	0.404
Cr (%)	0	0.3	0.11
Fe (%)	0.150	1	0.576
Si (%)	0.15	1	0.436
Al (%)	94.65	99.58	97.089
Yield strength (Mpa)	55	230	112.408
Tensile strength (Mpa)	90	375	232.95

2. Preprocessing data

Setelah mengumpulkan data, langkah berikutnya adalah membersihkan dan menyaring data. Proses ini melibatkan penghapusan data ganda, data yang tidak sesuai dengan format variabel unsur kimia yang telah ditentukan, dan data yang memiliki rentang nilai yang signifikan berbeda dengan data lainnya. Setelah proses preprocessing data selesai, dataset terdiri dari 53 sampel yang mewakili dua perlakuan panas, yaitu O (Anil) dan H (Strain Hardened).

Langkah selanjutnya dilakukan normalisasi data, dataset yang terdiri dari unsur kimiadan sifat mekanik memiliki rentang nilai yang

berbeda-beda. Untuk mengatasi permasalahan ini, digunakan metode *MinMaxScaler*. Metode ini bekerja dengan cara memetakan nilai setiap fitur pada rentang 0 hingga 1, sehingga rentang nilai setiap fitur pada dataset akan disesuaikan dan memiliki rentang nilai yang seragam. Hal ini bertujuan untuk mempermudah pelatihan GAN, karena *minmaxscaler* membantu dalam menyelaraskan skala dari data asli sehingga kompatibel dengan model GAN yang pada gilirannya membantu dalam pelatihan yang lebih baik. Selain itu, dengan normalisasi menggunakan *MinMaxScaler*, data sintetis yang dihasilkan oleh GAN cenderung memiliki distribusi yang serupa dengan data asli (He et al., 2022). Ini berarti, data sintetis yang dihasilkan akan lebih mirip dengan data asli, dan dapat membantu dalam menciptakan data yang berkualitas tinggi. Penerapan normalisasi dengan metode *minmaxscaler* dapat dilihat pada gambar 3.



Gambar 3. Ilustrasi normalisasi data sifat mekanik aluminium

3. Pemodelan Generative Adversarial Networks (GAN)

Pemodelan ini bertujuan untuk menghasilkan data sintetis yang berkualitas tinggi dan memiliki karakteristik yang serupa dengan data asli. Pemodelan augmentasi data sifat mekanik aluminium dijalankan pada *framework* google colab menggunakan *Generative Adversarial Networks (GAN)*. Pada studi ini, data sintetis yang akan dibuat sebanyak jumlah data asli yaitu 53 sampel, hal ini bertujuan untuk mempermudah dalam

proses evaluasi data sintetis dengan data asli. Parameter GAN yang digunakan pada

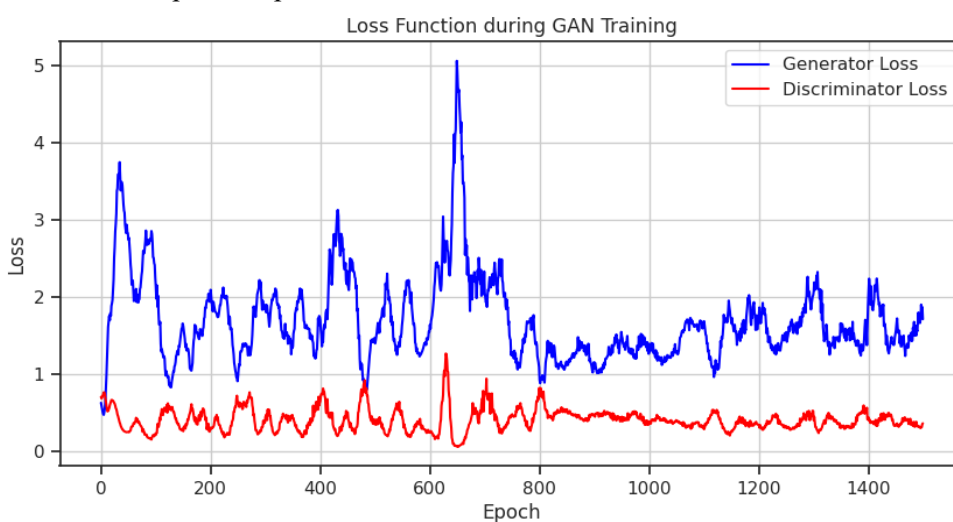
penelitian ini dapat dilihat pada tabel 2.

Tabel 2. Parameter Generative Adversarial Networks

No	Parameter	Nilai
1	Epochs	1500
2	Batch	128
3	Vektor Noise	100
4	Output Gen	11
5	Neuron Layer Dense Pert	128
6	Neuron Layer Dense Ked	256
7	Aktivasi Output Gen	11
8	Neuron Layer Dense Pert	256
9	Neuron Layer Dense Ked	128
10	Aktivasi Output Disc	Sigmoid
11	Loss Function Disc	Binary Crossentropy
12	Loss Function GAN	Binary Crossentropy

Pada saat pelatihan model, generator berusaha untuk menghasilkan data sintetis yang semakin mirip dengan data asli. Dimana pada awalnya, generator menghasilkan data sintetis yang acak dan belum sesuai dengan bentuk data asli. Sedangkan di sisi lain, discriminator bertugas untuk membedakan antara data asli dan data sintetis. *Discriminator* melakukan evaluasi terhadap kualitas data yang diberikan. Selama proses pelatihan,

generator dan discriminator berada dalam persaingan. Generator berusaha untuk memperbaiki kualitas data sintetisnya agar sulit dibedakan oleh discriminator, sementara *discriminator* berupaya untuk tetap dapat membedakan data asli dari data sintetis. Kualitas pelatihan model GAN dapat dilihat dari grafik *lossfunction* (fungsi kerugian), seperti yang dapat dilihat pada gambar 4.



Gambar 3. Loss pelatihan model

Berdasarkan gambar 4 dapat dilihat bahwa loss function dari generator dan discriminator

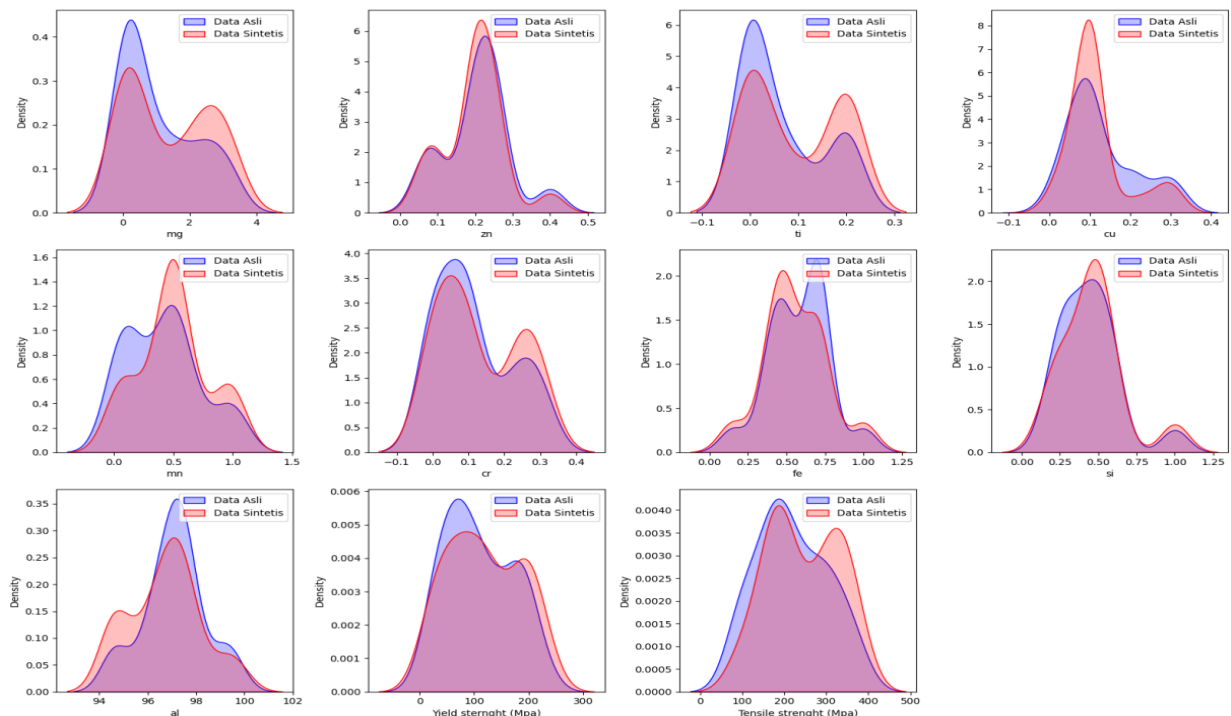
berubah seiring waktu selama pelatihan. Pada awal pelatihan, generator menghasilkan data

yang sangat buruk, hal ini dapat dilihat pada grafik awal generator yg cukup tinggi. Namun, seiring berjalannya waktu, generator belajar menghasilkan data yang semakin mirip dengan data asli, sehingga loss function generator berkurang. Di sisi lain, loss function discriminator awalnya juga tinggi, tetapi dapat turun seiring dengan berjalannya pelatihan karena discriminator harus menghadapi generator yang semakin baik. Tujuan akhir dari pelatihan GAN adalah mencapai keseimbangan di mana generator menghasilkan data yang sangat mirip dengan data asli, dan discriminator memiliki kesulitan untuk membedakan keduanya. Itu ditandai dengan kedua loss function generator dan discriminator yang stabil pada tingkat yang relatif rendah.

4. Evaluasi Model

Sebanyak 53 sampel data sintetis dihasilkan melalui pemodelan augmentasi data menggunakan GAN. Data ini dievaluasi untuk menilai sejauh mana kesamaannya dengan data

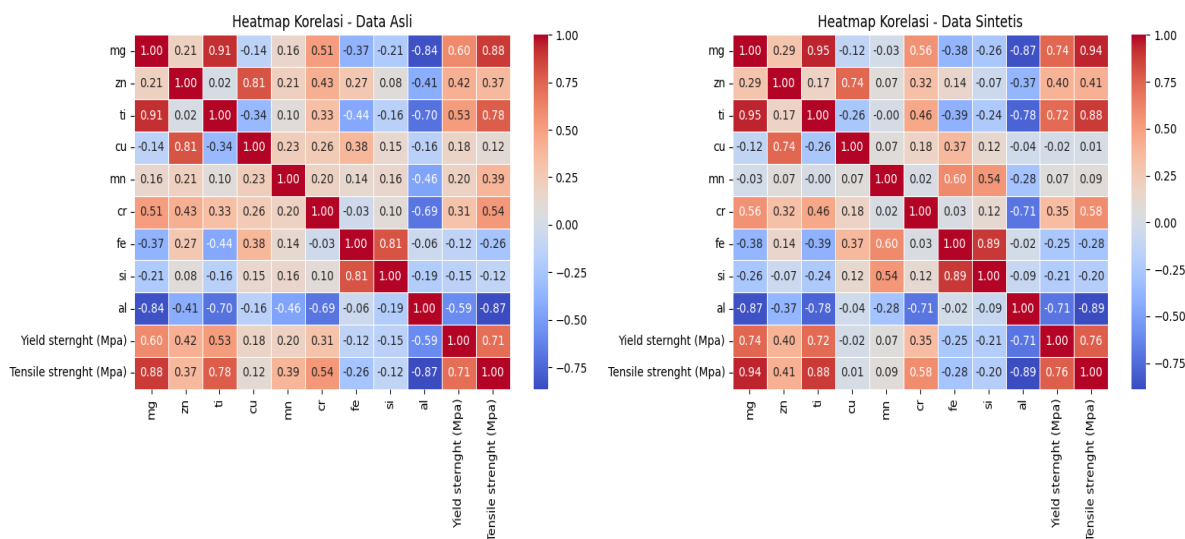
asli. Untuk memudahkan perbandingan antara masing-masing variabel unsur kimia dan sifat mekanik aluminium, kedua dataset divisualisasikan dengan menggunakan kernel density, seperti yang terlihat pada Gambar 5. Hasil evaluasi menunjukkan bahwa setiap variabel unsur kimia dan sifat mekanik memiliki bentuk kernel density yang serupa. Hal ini mengindikasikan bahwa setiap variabel dalam data sintetis memiliki kepadatan distribusi yang hampir serupa dengan data asli dari aluminium. Data sintetis yang dihasilkan oleh GAN dapat menjadi sangat mirip dengan data asli karena GAN belajar dari distribusi data asli. Proses pembelajaran ini memungkinkan GAN untuk memahami pola dan struktur data asli, sehingga data sintetis yang dihasilkan memiliki karakteristik yang serupa (Fonseca & Bacao, 2023). Dengan demikian, GAN dapat digunakan untuk menghasilkan data sintetis dengan kualitas tinggi yang dapat memperkaya dataset atau mengatasi keterbatasan data asli.



Gambar 4. Perbandingan data asli dengan data sintetis

Evaluasi berikutnya dilakukan dengan menganalisis korelasi antara unsur kimia dan sifat mekanik aluminium pada data asli dan data sintetis. Korelasi Pearson adalah sebuah metode statistik yang digunakan untuk mengukur hubungan linier antara dua variabel numerik. Nilai korelasi berkisar antara -1 hingga 1, di mana nilai 1 menunjukkan korelasi positif sempurna. Metode ini mengukur sejauh mana dua variabel tersebut berkorelasi atau bergerak bersamaan dalam suatu hubungan linear. Dengan kata lain, data sifat mekanik aluminium yang diperoleh dari

GAN dapat dianggap mirip dengan data asli jika korelasi pada data sintetis memiliki korelasi yang serupa atau mendekati data asli. Sebaliknya, jika korelasi antara unsur kimia dan sifat mekanik pada data sintetis sangat berbeda dari data asli, maka data sintetis dianggap memiliki tingkat kesamaan yang rendah dan tidak mewakili data sebenarnya. Perbandingan korelasi antar variabel unsur kimia dan sifat mekanik aluminium pada data asli dan data sintetis dapat dilihat pada Gambar 5.



Gambar 5. Perbandingan korelasi data asli (a) dengan data sintetis (b)

Pada data asli, terlihat bahwa unsur kimia seperti Mg memiliki korelasi yang sangat kuat dengan Yield Strength (YS) dan Tensile Strength (TS) sebesar 0.6 dan 0.8, secara berturut-turut. Titanium (Ti) menunjukkan korelasi positif yang cukup kuat dengan YS sebesar 0.53 dan 0.78 dengan TS. Sebaliknya, Aluminium (Al) memiliki korelasi negatif yang sangat kuat dengan sifat mekanik, yaitu -0.59 terhadap YS dan -0.87 terhadap TS. Pada data sintetis, ditemukan hasil korelasi yang hampir sama dengan data asli, menunjukkan hubungan positif yang sangat kuat antara Mg dengan YS dan TS, dengan nilai korelasi masing-masing sebesar 0.74 dan 0.94. Hal ini mengalami peningkatan sebesar 0.14 pada YS dan 0.06 pada TS jika dibandingkan dengan korelasi pada data asli. Unsur Ti juga

menunjukkan korelasi positif yang cukup kuat dengan sifat mekanik, dan mengalami peningkatan korelasi dibandingkan dengan data asli, yaitu 0.19 pada YS dan 0.1 pada TS. Sementara itu, unsur Aluminium tetap memiliki korelasi negatif yang kuat terhadap kedua sifat mekanik, dengan nilai korelasi sebesar 0.71 untuk YS dan 0.89 untuk TS. Perbedaan ini tidak terlalu signifikan dibandingkan dengan korelasi pada data asli.

Hasil korelasi unsur kimia dengan sifat mekanik aluminium yang diperoleh pada penelitian ini, sejalan dengan penelitian terdahulu yang menyatakan bahwa penambahan Mg pada paduan aluminium meningkatkan YS dan TS secara signifikan. Dalam penelitian tersebut, penambahan Mg

berkontribusi pada pembentukan struktur mikro yang kuat, menghasilkan peningkatan sifat mekanik seperti YS dan TS (Rawal et al., 2012). Dalam penelitian (Ye et al., 2022) mengenai komposit matriks logam, penambahan unsur Titanium (Ti) dan Magnesium (Mg) dapat memainkan peran penting dalam meningkatkan sifat mekanik. Hasil penelitian tersebut menunjukkan bahwa komposit Mg-3Al-1Zn yang diperkuat dengan partikel Ti memiliki peningkatan pada *yield strength*, *elongation*, dan modulus elastis seiring dengan peningkatan kandungan partikel Ti. Peningkatan kekuatan utama disebabkan oleh pepadatan butir (*grain refinement*) dan ikatan antar muka yang kuat antara partikel Ti dan matriks Mg.

Pada penelitian ini juga dilakukan uji Kolmogorov-Smirnov (KS), yang merupakan suatu metode statistik untuk menguji kesamaan distribusi antara dua kelompok data atau sampel (Berger & Zhou, 2014). Dalam konteks penelitian ini, KS digunakan untuk membandingkan distribusi data sintetis dengan distribusi data asli. Proses pengujian KS ini berjalan dengan membandingkan dua fungsi

distribusi kumulatif (CDF) dari dua kelompok data yang akan dibandingkan, yaitu data asli dan data sintetis. Uji ini menghasilkan dua statistik, yaitu nilai KS Statistic dan P-Value. Nilai KS Statistic mengukur seberapa besar deviasi atau perbedaan antara kedua distribusi CDF. Semakin besar nilai KS Statistic, semakin besar perbedaan antara kedua distribusi tersebut. Sedangkan P-Value adalah ukuran signifikansi statistik. Nilai P-Value mencerminkan seberapa signifikan perbedaan antara dua distribusi. Nilai P-Value yang rendah menunjukkan adanya perbedaan yang signifikan antara kedua distribusi, sementara nilai P-Value yang tinggi menunjukkan bahwa perbedaan tersebut tidak signifikan (Lall, 2015). Nilai KS Statistic berkisar antara 0 hingga 1, di mana 0 berarti kedua distribusi identik (sama persis) dan 1 berarti distribusi sepenuhnya berbeda. Nilai P-Value dibandingkan dengan tingkat signifikansi yang telah ditentukan, pada penelitian ini digunakan 0,05. Jika P-Value lebih besar dari tingkat signifikansi, maka dapat diterima bahwa kedua distribusi tersebut mirip. Hasil uji KS data sintetis dengan data asli dapat dilihat pada tabel 3.

Tabel 3. Hasil uji Kolmogorov-Smirnov (KS) data sintetis dengan data asli

Variabel	KS Statistic	P-Value
Mg	0,151	0,586
Zn	0,094	0,975
Ti	0,151	0,586
Cu	0,132	0,75
Mn	0,17	0,433
Cr	0,094	0,975
Fe	0,113	0,891
Si	0,075	0,999
Al	0,151	0,586
Yield strength (Mpa)	0,132	0,75
Tensile strength (Mpa)	0,151	0,586

Berdasarkan tabel 1, dapat dilihat bahwa KS Statistic untuk semua variabel unsur kimia berada pada rentang yang cukup rendah, mendekati 0. Nilai-nilai ini menunjukkan bahwa distribusi unsur kimia pada data sintetis

cenderung mirip dengan distribusi pada data asli. Hal ini juga didukung oleh nilai P-Value yang tinggi pada semua variabel unsur kimia, yaitu mendekati 1, mengindikasikan tingkat signifikansi yang tinggi. Dengan kata lain,

hasil ini menunjukkan bahwa distribusi unsur kimia pada data sintetis sangat mirip dengan distribusi pada data asli. Pada sifat mekanik *Yield Strength (YS)* dan *Tensile Strength (TS)* juga diperoleh hasil yang serupa yaitu nilai KS yang lebih kecil dan nilai P-Value yang lebih besar. Hasil ini menyimpulkan bahwa data sintetis yang dihasilkan dari pemodelan augmentasi data menggunakan GAN memiliki tingkat kemiripan distribusi yang tinggi dengan data asli. Ini mengindikasikan bahwa GAN mampu menciptakan data sintetis yang berkualitas tinggi, yang dapat digunakan untuk pengembangan model dan analisis material lebih lanjut

IV. Kesimpulan

Berdasarkan hasil pemodelan augmentasi data sifat mekanik aluminium menggunakan *Generative Adversarial Networks (GAN)* dapat disimpulkan, bahwa pemodelan augmentasi data sifat mekanik aluminium menggunakan *Generative Adversarial Networks (GAN)* memiliki potensi untuk menghasilkan data sintetis yang mirip dengan data asli. Hal ini dapat dilihat melalui evaluasi data sintetis menggunakan beberapa metode, yaitu korelasi Pearson, statistik deskriptif, dan Uji Kolmogorov-Smirnov (KS). Hasil korelasi Pearson menunjukkan bahwa sebagian besar variabel unsur kimia dan sifat mekanik aluminium pada data sintetis memiliki korelasi yang cukup mirip dengan data asli. Dalam analisis statistik deskriptif, distribusi data sintetis memiliki kesamaan dengan data asli, mengindikasikan bahwa parameter statistik seperti mean, median, dan deviasi standar relatif serupa. Ini menunjukkan bahwa karakteristik data sintetis memiliki kemiripan dengan data asli. Selain itu, melalui Uji Kolmogorov-Smirnov (KS), data sintetis diuji untuk kesesuaian distribusi dengan data asli. Hasil menunjukkan bahwa distribusi data sintetis dan data asli memiliki kemiripan yang signifikan, sehingga data sintetis dapat mewakili data asli dengan baik. Hasil temuan ini memberikan implikasi positif dalam konteks penggunaan GAN untuk menghasilkan

data sintetis sifat mekanik aluminium. Data sintetis yang berkualitas tinggi dan mirip dengan data asli dapat digunakan dalam berbagai aplikasi, seperti pemodelan dan analisis, tanpa harus bergantung pada ketersediaan data asli yang terbatas. Dengan demikian, penggunaan GAN dalam augmentasi data material menjadi alternatif yang menjanjikan untuk mendukung penelitian dan pengembangan material di masa yang akan mendatang.

Daftar Pustaka

- Agrawal, A., Deshpande, P. D., Cecen, A., Basavarsu, G. P., Choudhary, A. N., & Kalidindi, S. R. (2014). Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integrating Materials and Manufacturing Innovation*, 3(1), 90–108. <https://doi.org/10.1186/2193-9772-3-8>
- Arif, A. F. ., Sheikh, A. ., & Qamar, S. . (2003). A study of die failure mechanisms in aluminum extrusion. *Journal of Materials Processing Technology*, 134(3), 318–328. [https://doi.org/10.1016/S0924-0136\(02\)01116-0](https://doi.org/10.1016/S0924-0136(02)01116-0)
- Ashkenazi, D. (2019). How aluminum changed the world: A metallurgical revolution through technological and cultural perspectives. *Technological Forecasting and Social Change*, 143, 101–113. <https://doi.org/10.1016/j.techfore.2019.03.011>
- Berger, V. W., & Zhou, Y. (2014). Kolmogorov–Smirnov Test: Overview. In *Wiley StatsRef: Statistics Reference Online*. Wiley. <https://doi.org/10.1002/9781118445112.stat06558>
- Blaiszik, B., Ward, L., Schwarting, M., Gaff, J., Chard, R., Pike, D., Chard, K., & Foster, I. (2019). A data ecosystem to support machine learning in materials science. *MRS Communications*, 9(4), 1125–1133. <https://doi.org/10.1557/mrc.2019.118>
- Fonseca, J., & Bacao, F. (2023). Tabular and latent space synthetic data generation: a literature review. *Journal of Big Data*, 10(1), 115.

- <https://doi.org/10.1186/s40537-023-00792-7>
- Frydrych, K., Karimi, K., Pecelerowicz, M., Alvarez, R., Dominguez-Gutiérrez, F. J., Rovaris, F., & Papanikolaou, S. (2021). Materials Informatics for Mechanical Deformation: A Review of Applications and Challenges. *Materials*, *14*(19), 5764. <https://doi.org/10.3390/ma14195764>
- He, G., Zhao, Y., & Yan, C. (2022). Application of tabular data synthesis using generative adversarial networks on machine learning-based multiaxial fatigue life prediction. *International Journal of Pressure Vessels and Piping*, *199*, 104779. <https://doi.org/10.1016/j.ijpvp.2022.104779>
- Hwang, J. Y., Banerjee, R., Doty, H. W., & Kaufman, M. J. (2009). The effect of Mg on the structure and properties of Type 319 aluminum casting alloys. *Acta Materialia*, *57*(4), 1308–1317. <https://doi.org/10.1016/j.actamat.2008.11.021>
- Koc, K., & Gurgun, A. P. (2022). Scenario-based automated data preprocessing to predict severity of construction accidents. *Automation in Construction*, *140*, 104351. <https://doi.org/10.1016/j.autcon.2022.104351>
- Lall, A. (2015). Data streaming algorithms for the Kolmogorov-Smirnov test. *2015 IEEE International Conference on Big Data (Big Data)*, 95–104. <https://doi.org/10.1109/BigData.2015.7363746>
- Leni, D. (2023). Pemilihan Algoritma Machine Learning Yang Optimal Untuk Prediksi Sifat Mekanik Aluminium. *Jurnal Engine: Energi, Manufaktur, Dan Material*, *7*(1), 35–44.
- Leni, D., Earnestly, F., Sumiati, R., Adriansyah, A., & Kusuma, Y. P. (2023). Evaluasi sifat mekanik baja paduan rendah berdasarkan komposisi kimia dan suhu perlakuan panas menggunakan teknik exploratory data analysis (EDA). *Dinamika Teknik Mesin*, *13*(1), 74. <https://doi.org/10.29303/dtm.v13i1.624>
- Leni, D., Muchlisinalahuddin, M., Maimuzar, M., Haris, H., & Hendra, H. (2023). Analisis Heatmap Korelasi dan Scatterplot untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Pelabelan AC efisiensi Energi. *Jurnal Rekayasa Material, Manufaktur Dan Energi*, *6*(1). <https://doi.org/10.30596/rmme.v6i1.13133>
- Leni, D., Yermadona, H., Usra Berli, A., Sumiati, R., & Haris, H. (2023). Pemodelan Machine Learning untuk Memprediksi Tensile Strength Aluminium Menggunakan Algoritma Artificial Neural Network (ANN). *Jurnal Surya Teknik*, *10*(1), 625–632. <https://doi.org/10.37859/jst.v10i1.4843>
- Li, D.-C., Chen, S.-C., Lin, Y.-S., & Huang, K.-C. (2021). A Generative Adversarial Network Structure for Learning with Small Numerical Data Sets. *Applied Sciences*, *11*(22), 10823. <https://doi.org/10.3390/app112210823>
- Marani, A., Jamali, A., & Nehdi, M. L. (2020). Predicting Ultra-High-Performance Concrete Compressive Strength Using Tabular Generative Adversarial Networks. *Materials*, *13*(21), 4757. <https://doi.org/10.3390/ma13214757>
- Matmach. (n.d.). *matmach*. Retrieved July 15, 2022, from <https://matmatch.com/advanced-search?categories=aluminium&properties=tensile-strength,yield-strength&header=tensile-strength,yield-strength&units-system=imperial>
- Morini, A. A., Ribeiro, M. J., & Hotza, D. (2019). Early-stage materials selection based on embodied energy and carbon footprint. *Materials & Design*, *178*, 107861. <https://doi.org/10.1016/j.matdes.2019.107861>
- Rajan, K. (2015). Materials Informatics: The Materials “Gene” and Big Data. *Annual Review of Materials Research*, *45*(1), 153–169. <https://doi.org/10.1146/annurev-matsci-070214-021132>
- Rawal, A., Kumar, R., & Saraswat, H. (2012). Tensile mechanics of braided sutures. *Textile Research Journal*, *82*(16), 1703–1710. <https://doi.org/10.1177/0040517512445340>
- Sert, E., Öchsner, A., Hitzler, L., Werner, E., & Merkel, M. (2019). *Additive Manufacturing: A Review of the*

Influence of Building Orientation and Post Heat Treatment on the Mechanical Properties of Aluminium Alloys (pp. 349–366). https://doi.org/10.1007/978-3-030-30355-6_14

Tsai, Y.-C., Chou, C.-Y., Lee, S.-L., Lin, C.-K., Lin, J.-C., & Lim, S. W. (2009). Effect of trace La addition on the microstructures and mechanical properties of A356 (Al–7Si–0.35Mg) aluminum alloys. *Journal of Alloys and Compounds*, 487(1–2), 157–162. <https://doi.org/10.1016/j.jallcom.2009.07.183>

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F.-Y. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4), 588–598. <https://doi.org/10.1109/JAS.2017.7510583>

Ye, J., Li, J., Luo, H., Tan, J., Chen, X., Feng, B., Zheng, K., & Pan, F. (2022). Effect of micron-Ti particles on microstructure and mechanical properties of Mg–3Al–1Zn based composites. *Materials Science and Engineering: A*, 833, 142526. <https://doi.org/10.1016/j.msea.2021.142526>

Yu, D., Zhang, H., Chen, W., Yin, J., & Liu, T.-Y. (2021). How Does Data Augmentation Affect Privacy in Machine Learning? *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 10746–10753. <https://doi.org/10.1609/aaai.v35i12.17284>