**IMPLEMENTASI ALGORITMA GENERATIVE ADVERSARIAL NETWORK (GAN) UNTUK MENAMBAH DATA CITRA MULTISPEKTRAL LAHAN SAWAH DALAM PEMETAAN KESUBURAN LAHAN**

**ANDI MUHAMMAD ALIFIKRI**

**G64190005**

Draft Proposal Kolokium

**DEPARTEMEN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA dan ILMU PENGETAHUAN ALAM INSTITUT PERTANIAN BOGOR**

**BOGOR 2022**

1. **PENDAHULUAN**

## Latar Belakang

Citra multispektral adalah pemisahan objek dalam hal fitur spesifik yang unik untuk setiap objek berdasarkan spektrum elektromagnetik (Kadyanan 2022). Dengan kata lain, citra multispektral dapat diartikan sebagai citra yang berisi data gelombang frekuensi secara detail yang direpresentasikan dalam spektrum elektromagnetik. Frekuensi penglihatan manusia hanya terletak pada rentang panjang gelombang 400-700 nm, dan akibatnya manusia hanya dapat melihat dan membedakan warna mulai dari ungu hingga merah (Santoso 2009). Sedangkan citra multispektral menangkap data dari berbagai panjang gelombang tertentu pada seluruh spektrum elektromagnetik sehingga memberikan lebih banyak informasi yang diperlukan dalam menganalisis karakteristik dan identifikasi komponen yang menyusun pada citra yang ditangkap (ElMasry *et al.* 2019). Oleh karena itu citra multispektral dapat menampilkan spektrum cahaya lain yang tidak tampak oleh penglihatan manusia yang sangat berguna untuk bermacam kebutuhan penelitian.

Dalam dunia pertanian, citra multispektral dapat digunakan untuk memetakan kesuburan lahan persawahan padi yang diukur menggunakan Bagan Warna Daun (BWD) atau Leaf Color Chart (LCC) dimana warna daun padi menjadi parameter kebutuhan pupuk tanaman padi (Nasution *et al.* 2022). Data pada citra multispektral ini dapat digunakan untuk melatih sebuah model *machine learning* yang nantinya digunakan untuk mengklasifikasikan tingkat kesuburan lahan pertanian, tetapi tentunya hal ini membutuhkan record data yang cukup banyak untuk melatih model tersebut (Wijayanto *et al.* 2020).

Data citra multispektral pada lahan persawahan dapat diperoleh melalui alat UAV multispektral. UAV (*Unmanned Aerial Vehicle*) sendiri merupakan sebuah pesawat tanpa awak yang dapat dikendalikan oleh manusia menggunakan remot kontrol (Fahlstrom *et al.* 2022). UAV memiliki berbagai sensor yang bisa menyamai sensor yang ada pada satelit sehingga dapat digunakan pada bidang pertanian dengan efisiensi yang tinggi untuk skala lapangan dalam pengambilan data (Shofiyanti 2011). UAV multispectral merupakan sebuah alat yang sangat efisien dalam melakukan monitoring tanaman padi di sawah (Wang *et al.* 2021). Dengan efisiensi yang ditawarkan maka UAV multispectral banyak digunakan dalam pemetaan dan menganalisis lahan persawahan.

Telah ada beberapa penelitian dengan memanfaatkan citra multispektral pada bidang pertanian yang sebelumnya telah dilakukan oleh para peneliti. Misalnya pada penelitian yang dilakukan oleh wang et al pada tahun 2021, mereka meneliti tentang status ketersediaan nitrogen tanaman padi pada fase vegetatif menggunakan UAV multispectral. didalam penelitian ini menyimpulkan bahwa UAV multispektral sebagai sebuah alat yang cukup efisien ketika digunakan untuk memantau status nitrogen tanaman padi di sawah. Namun, gangguan latar belakang (air, tanah gundul, dan ganggang) menciptakan masalah ketika mencoba mengolah informasi status nitrogen untuk tanaman padi pada data citra multispektral (Wang *et al.* 2021). Masalah lainnya yaitu persebaran yang tidak merata pada tanaman padi yang memiliki kebutuhan pupuk berbeda pada sawah yang diambil datanya sehingga ketika mengambil data dengan UAV menyebabkan jumlah record data kelas yang tidak seimbang. Sebagai contoh, pada lahan persawahan di kecamatan Dramaga, persebaran area lahan tanaman padi yang memiliki kebutuhan pupuk tinggi dan kebutuhan pupuk rendah tidak sama banyak. Akibatnya ketika melakukan pengambilan citra multispektral pada area sawah tersebut didapati jumlah record data yang tidak seimbang antara data kelas dengan kebutuhan pupuk tinggi dan data kelas dengan kebutuhan pupuk rendah. Sehingga peneliti harus melakukan pengambilan data secara berulang pada lokasi dan kondisi lahan sawah yang berbeda untuk mengumpulkan record data hingga jumlah record data setiap kelasnya sama banyak dan dianggap cukup. Tentunya hal ini akan sangat sulit dilakukan karena membutuhkan pengambilan data yang sangat banyak dan memakan waktu yang lama.

Data yang tidak seimbang (imbalance) bisa menjadi masalah pada saat peneliti ingin melakukan training model untuk mendapatkan model klasifikasi pemetaan kesuburan lahan menggunakan machine learning, karena setiap kelas data tidak memiliki jumlah yang sama, sehingga membuat tingkat akurasi klasifikasi pada setiap kelas menjadi tidak maksimal (Zhang *et al.* 2021).  Ketika melakukan klasifikasi data dengan distribusi kelas yang tidak seimbang akan sangat mempengaruhi kinerja machine learning dalam membuat model klasifikasi. sehingga akibatnya adalah pengklasifikasian menjadi bias terhadap record data yang jumlah kelasnya lebih banyak (Patel *et al.* 2020).

Ada beberapa pendekatan dalam mengatasi data imbalance yaitu pendekatan pada tingkat data dan pendekatan pada tingkat algoritma. Pada pendekatan tingkat data menggunakan oversampling dan undersampling. Dikatakan pendekatan tingkat data karena mereka langsung memanipulasi data yang ada pada dataset demi menyeimbangkan sampel data dengan cara mengurangi sampel data mayoritas ataupun menghapus kelas minoritas. Kekurangan dari teknik oversampling ini yaitu bisa terjadi over fitting ketika dilakukan pelatihan model sedangkan untuk teknik undersampling dapat menyebabkan kehilangan beberapa sampel data yang (Rout *et al.* 2018). Lalu tipe yang kedua adalah pendekatan tingkat algoritma atau biasa disebut sebagai pendekatan internal karena memanfaatkan desain algoritma klasifikasi baru atau memodifikasi algoritma yang ada demi mengatasi bias akibat ketidakseimbangan data (Spelmen dan Porkodi 2018).

Oleh karena fokus penelitian ini adalah ingin menambah sampel data, maka pendekatan yang akan digunakan adalah pendekatan tingkat data (*Data-Level Approaches*) dengan teknik oversampling. SMOTE (*Synthetic Minority Oversampling Technique*) merupakan teknik oversampling yang sangat populer saat ini (Douzas *et al.* 2019). Algoritma SMOTE telah terverifikasi efektif dan valid dalam memproduksi sampel data sintetis dan telah banyak variasi SMOTE yang dibuat oleh beberapa peneliti (Xie dan Zhang 2018). Teknik lainnya yaitu dengan menggunakan algoritma GAN (*Generative Adversarial Networks).* GAN pada awalnya diciptakan untuk menghasilkan gambar buatan dari hasil pelatihan dua model algoritma yang saling bersaing (*adversarial*). Berbeda dengan SMOTE yang melakukan pendekatan pembelajaran berdasarkan informasi lokal dari dataset yang ada, GAN melakukan proses pembelajaran pada keseluruhan distribusi sampel data (Xie dan Zhang 2018). Perbedaan lainnya yaitu pada SMOTE memproduksi sampel data sintetis di sepanjang garis yang menghubungkan antar sampel data minoritas, sedangkan GAN bekerja dengan memperkirakan distribusi data aslinya lalu memproduksi data sintetis untuk kelas minor (Douzas dan Bacao 2018). Oleh karena itu penulis tertarik untuk meneliti apakah dengan menerapkan algoritma GAN dapat menghasilkan data sintetis yang valid dan dapat digunakan untuk menambah data imbalance pada citra multispektral lahan sawah dalam pemetaan kesuburan lahan.

## Rumusan Masalah

Data citra lahan sawah umumnya imbalance sehingga tidak cukup untuk digunakan dalam pemodelan pemetaan kesuburan lahan sawah, maka penulis merumuskan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma GAN *(Generative Adversarial Network)* untuk memproduksi data sintetis demi mengatasi *Imbalance Data* Citra Multispektral Lahan Sawah dalam pemetaan Kesuburan Lahan?
2. Apakah data sintetis yang dihasilkan oleh model algoritma GAN ini dapat digunakan sebagai data training yang valid?

## Tujuan

Berdasarkan latar belakang dan rumusan masalah, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma GAN *(Generative Adversarial Network)* untuk membangkitkan data tambahan/sintesis citra multispektral lahan sawah.
2. Membuat model klasifikasi kesuburan lahan sawah menggunakan data citra multispektral yang sudah ditambah dengan data sintetis.
3. Menganalisis pengaruh penambahan data sintetis terhadap akurasi model klasifikasi.

## Manfaat

Dengan menggunakan model dari GAN tersebut dapat membantu para peneliti dalam menambah data citra multispektral lahan sawah yang tidak seimbang sehingga para peneliti tidak mengulang pengambilan data ketika mengalami ketidakseimbangan data.

## Ruang Lingkup

Ruang lingkup pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data sekunder citra multispektral yang diambil dari kamera UAV multispectral di atas lahan persawahan di Kecamatan Dramaga yang telah diambil dari penelitian sebelumnya.
2. Masing-masing record data menunjukkan rata-rata fitur multispectral pada suatu grid sawah berukuran 4 x 4 meter.
3. Label data adalah level kebutuhan pupuk pada grid-grid sawah tersebut, yang diukur menggunakan Bagan Warna Daun (BWD) atau Leaf Color Chart (LCC), di mana warna daun padi menjadi parameter kebutuhan pupuk tanaman padi.
4. Implementasi algoritma GAN (Generative Adversarial Network) menggunakan bahasa pemrograman python dengan memanfaatkan library tensorflow.
5. **TINJAUAN PUSTAKA**
   1. **Citra Multispektral untuk Pemetaan Lahan Sawah**

Citra multispektral dapat menampilkan spektrum cahaya baik yang tampak maupun tidak tampak oleh penglihatan manusia, spektrum ini berguna untuk bermacam kebutuhan penelitian. Citra multispektral menangkap data dari berbagai panjang gelombang tertentu pada seluruh spektrum elektromagnetik sehingga memberikan lebih banyak informasi yang diperlukan dalam menganalisis karakteristik dan identifikasi komponen penyusun objek pada citra yang ditangkap (ElMasry *et al.* 2019).

Saat ini, teknologi citra multispektral juga telah banyak diimplementasikan dalam dunia pertanian seperti penentuan kualitas pangan dan tanaman seperti sereal, kacang-kacangan, buah-buahan, dan sayur-sayuran. Dengan memanfaatkan citra multispektral, kita dapat mendeteksi kualitas, mengevaluasi parameter fisiokimia, mendeteksi mikrobiologi, mendeteksi varietas, tingkat kontaminasi, hingga kecacaran pada pangan dan tanaman (Su dan Sun 2018).

Citra multispektral untuk pemetaan lahan sawah dapat diperoleh melalui alat yang bernama UAV multispektral. UAV (*Unmanned Aerial Vehicle*) sendiri merupakan sebuah pesawat tanpa awak yang dapat dikendalikan oleh manusia menggunakan remot kontrol (Fahlstrom *et al.* 2022). UAV memiliki berbagai sensor yang bisa menyamai sensor yang ada pada satelit sehingga dapat digunakan pada bidang pertanian dengan efisiensi yang tinggi untuk skala lapangan dalam pengambilan data (Shofiyanti 2011). UAV multispektral merupakan sebuah alat yang sangat efisien dalam melakukan monitoring tanaman padi di sawah (Wang *et al.* 2021). Dengan efisiensi yang ditawarkan, UAV multispectral banyak digunakan dalam pemetaan dan menganalisis lahan persawahan.

Didalam sebuah penelitian yang dilakukan oleh Rokhmatuloh *et al.* (2019) dengan judul “Pemetaan Sawah Padi menggunakan citra UAV Multispektral”. Penelitian ini ingin meneliti seberapa efektif sensor citra multispektral yang ada pada UAV untuk menghasilkan nilai NDVI (Normalized Difference Vegetation Index) dalam memetakan persawahan. UAV Multispektral yang digunakan memiliki sensor multispectral empat bands yaitu green, red, red edge, dan near-infrared. Dalam melakukan pemetaan sawah padi mereka menggunakan nilai NDVI sebagai alat untuk membedakan berbagai jenis dan karakteristik tanaman. Untuk menghitung nilai NDVI ini digunakan formula perhitungan sebagai berikut:

(1)

dimana red dan NIR merupakan nilai radiasi red (cahaya tampak) dan near-infrared yang ditangkap oleh sensor.

Dari hasil penelitian yang dilakukan oleh Rokhmatuloh *et al.* (2019) didapatakan kesimpulan bahwa sensor multispektral pada UAV telah berhasil digunakan untuk menghasilkan nilai NDVI untuk membantu memetakan lahan sawah dengan dapat membedakan jenis jenis tanaman yang ada pada lahan sawah yang diteliti. Selain itu nilai NDVI yang dihasilkan pula dapat membedakan umur tanaman padi. Tetapi nilai NDVI ini masih bervariasi dan tidak jarang beberapa tanaman yang sama memiliki nilai NDVI yang berbeda.

* 1. **Masalah dalam Imbalance Data**

Ketidakseimbangan data kelas pada sebuah dataset terjadi karena adanya kelas yang jumlah datanya sedikit (kelas minoritas) dibandingkan kelas lain (kelas mayoritas). Sebagai contoh dari 1000 jumlah sampel terdapat 99% data yang tergolong kelas mayoritas dan hanya terdapat 1% data yang tergolong kelas minoritas, lalu ketika melakukan pelatihan model didapatkan akurasi 99% yang sebenarnya akurasi itu hanya berasal dari kelas mayoritas sedangkan kelas minoritas didominasi oleh kelas mayoritas, bisa saja 1% data minoritas tersebut mengandung inforamsi yang penting. Sehingga ketika menjalankan model klasifikasi, sebagian besar akan salah mengklasifikasikan *(missclassification)* sampel data kelas minoritas, sedangkan sampel data pada kelas mayoritas akan jarang untuk terjadi kesalahan klasifikasi (Ali *et al.* 2019). Kesalahan klasifikasi pada model seperti ini memberikan dampak pada kerugian biaya dan waktu (Longadge *et al.* 2013).

Data yang tidak seimbang *(imbalance)* dapat menjadi masalah pada saat peneliti ingin melakukan training model untuk mendapatkan model klasifikasi menggunakan machine learning, karena setiap kelas data tidak memiliki jumlah yang sama, sehingga membuat tingkat akurasi klasifikasi pada setiap kelas menjadi tidak maksimal (Zhang *et al.* 2021).  Ketika melakukan klasifikasi data dengan distribusi kelas yang tidak seimbang akan sangat mempengaruhi kinerja machine learning dalam membuat model klasifikasi. sehingga akibatnya adalah pengklasifikasian menjadi bias terhadap record data yang jumlah kelasnya lebih banyak (Patel *et al.* 2020)

Pendekatan yang dapat dilakukan dalam mengatasi data imbalance yaitu pendekatan pada tingkat data dan pendekatan pada tingkat algoritma. Pada pendekatan tingkat data menggunakan oversampling dan undersampling. Dikatakan pendekatan tingkat data karena mereka langsung memanipulasi data yang ada pada dataset demi menyeimbangkan sampel data dengan cara mengurangi sampel data mayoritas ataupun menghapus kelas minoritas. Kekurangan dari teknik oversampling ini yaitu bisa terjadi over fitting ketika dilakukan pelatihan model sedangkan untuk teknik undersampling dapat menyebabkan kehilangan beberapa sampel data yang (Rout *et al.* 2018). Lalu tipe yang kedua adalah pendekatan tingkat algoritma atau biasa disebut sebagai pendekatan internal karena memanfaatkan desain algoritma klasifikasi baru atau memodifikasi algoritma yang ada demi mengatasi bias akibat ketidakseimbangan data (Spelmen dan Porkodi 2018).

* 1. **Generative Adversarial Networks**

Algoritma Generative Adversarial Networks (GAN) merupakan salahsatu jenis algoritma oversampling. Dikatakan oversampling karena algoritma ini berusaha membuat atau *generate* data sintetis baru menyerupai data yang dimiliki dengan tujuan menambah jumlah sampel data. GAN bekerja dengan mempelajari distribusi data asli baik data mayoritas maupun data minoritas yang jumlahnya terbatas, lalu memanfaatkan distribusi yang telah dipelajari untuk menghasilkan data sintetis (Sampath *et al.* 2021).

Algoritma gan pertamakali diperkenalkan oleh Goodfellow *et al.* (2014). Sesuai namanya, GAN merupakan dua jaringan machine learning yang saling bersaing *(adversarial)* untuk mengoptimalkan kinerja masing masing. Kedua jaringan itu ialah *Generator* dan *Discriminator. Generator* akan membuat data sintetis dengan tujuan untuk menipu diskriminator sedangkan diskriminator akan mencoba membedakan antara dara asli dengan data yang dibuat oleh diskriminator. Kedua jaringan ini memiliki kemampuan untuk mengoptimalkan kinerja masing-masing dengan konsep backpropagation. Ketika *discriminator* berhasil membedakan antara data asli dengan data sintetis maka dikembalikan sebagai feedback ke *generator* dan generator akan berusaha meningkatkan kinerjanya. Begitu pula sebaliknya diskriminator akan terus belajar untuk membedakan data palsu dan data asli secara akurat. Proses pelatihan selesai ketika diskriminator tidak lagi dapat membedakan antara data asli dan data sintetis. Secara matematis Goodfellow *et al.* (2014) menggambarkan GAN dalam persamaan (2).

(2)

dimana z adalah variabel input dengan sampel data berupa random noise untuk melatih jaringan *Generator,*  merepresentasikan probabilitas x merupakan data asli bukan merupakan data hasil dari *Generator* . Diskriminator (D) dilatih untuk memaksimalkan probabilitas penentuan sampel data asli dan data sintetis. Secara simultan, Generator (G) dilatih untuk meminimalkan nilai agar D dapat terkecoh. Dengan kata lain, D dan G memainkan permainan minimax-game.

Di dalam paper tersebut pula disebutkan metode *Conditional Generative model* untuk mendapatkan data sintetis sesuai kelas yang diinginkan. Dengan secara sederhana menambahkan variabel y sebagai input tambahan pada D dan G.

(3)

*Conditional model* ini mampu mengarahkan proses generasi data sehingga output dari G dapat dikontrol sesuai apa yang diinginkan. Nilai y ini merupakan informasi tambahan seperti kelas label (Mirza dan Osindero 2014).

**DAFTAR PUSTAKA**

Ali H, Salleh M, Saedudin R, Hussain K, Mushtaq M. 2019. Imbalance class problems in data mining: A review. *Indonesian Journal of Electrical Engineering and Computer Science*. 14.doi:10.11591/ijeecs.v14.i3.pp1552-1563.

Douzas G, Bacao F. 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Applications*. 91:464–471.doi:10.1016/j.eswa.2017.09.030.

Douzas G, Bacao F, Fonseca J, Khudinyan M. 2019. Imbalanced Learning in Land Cover Classification: Improving Minority Classes’ Prediction Accuracy Using the Geometric SMOTE Algorithm. *Remote Sensing*. 11(24):3040.doi:10.3390/rs11243040.

ElMasry G, Mandour N, Al-Rejaie S, Belin E, Rousseau D. 2019. Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview. *Sensors*. 19(5):1090.doi:10.3390/s19051090.

Fahlstrom PG, Gleason TJ, Sadraey MH. 2022. *Introduction to UAV Systems*. John Wiley & Sons. Ed ke-Google-Books-ID: s8Z6EAAAQBAJ.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. 2014. Generative Adversarial Nets. Di dalam: *Advances in Neural Information Processing Systems*. [internet] Vol. 27. Curran Associates, Inc. [diunduh 2022 Okt 14]. Tersedia pada: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

Kadyanan IGAGA. 2022. Deteksi Perubahan Penggunaan Lahan dan Penutup Lahan Pada Citra Multispectral Berdasarkan Pendekatan Codebook dan Keyblock. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*. 11(2):145–155.doi:10.23887/janapati.v11i2.49102.

Longadge MR, Dongre SS, Malik DL. 2013. Class Imbalance Problem in Data Mining: Review. 2(1):6.

Mirza M, Osindero S. 2014. Conditional Generative Adversarial Nets. [diunduh 2022 Okt 8]. Tersedia pada: http://arxiv.org/abs/1411.1784

Patel H, Singh Rajput D, Thippa Reddy G, Iwendi C, Kashif Bashir A, Jo O. 2020. A review on classification of imbalanced data for wireless sensor networks. *International Journal of Distributed Sensor Networks*. 16(4):1550147720916404.doi:10.1177/1550147720916404.

Rokhmatuloh, Supriatna, Giok Pin T, Hernina R, Ardhianto R, Putut Ash Shidiq I, Adi W. 2019. Paddy Field Mapping Using Uav Multi-Spectral Imagery. *International Journal of GEOMATE*. 17(61):242–247.doi:https://doi.org/10.21660/2019.61.icee408.

Rout N, Mishra D, Mallick MK. 2018. Handling Imbalanced Data: A Survey. Di dalam: Reddy MS, Viswanath K, K.M. SP, editor. *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications*. Singapore. Singapore: Springer. (Advances in Intelligent Systems and Computing). hlm. 431–443.

Sampath V, Maurtua I, Aguilar Martín JJ, Gutierrez A. 2021. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data*. 8(1):27.doi:10.1186/s40537-021-00414-0.

Santoso I. 2009. *Interaksi Manusia dan Komputer*. Ed ke-2. C.V. Andi Offset. Ed ke-Google-Books-ID: \_pXa7CvwTC0C.

Shofiyanti R. 2011. TEKNOLOGI PESAWAT TANPA AWAK UNTUK PEMETAAN DAN PEMANTAUAN TANAMAN DAN LAHAN PERTANIAN. *Informatika Pertanian*. 20(2):58–64.

Spelmen VS, Porkodi R. 2018. A Review on Handling Imbalanced Data. Di dalam: *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*;. hlm. 1–11.

Su W-H, Sun D-W. 2018. Multispectral Imaging for Plant Food Quality Analysis and Visualization. *Comprehensive Reviews in Food Science and Food Safety*. 17(1):220–239.doi:10.1111/1541-4337.12317.

Wang Y-P, Chang Y-C, Shen Y. 2021. Estimation of nitrogen status of paddy rice at vegetative phase using unmanned aerial vehicle based multispectral imagery. *Precision Agric*. 23(1):1–17.doi:10.1007/s11119-021-09823-w.

Wijayanto AW, Wahyu Triscowati D, Marsuhandi AH. 2020. Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach. Di dalam: *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*. *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*;. hlm. 168–173.

Xie Y, Zhang T. 2018. Imbalanced Learning for Fault Diagnosis Problem of Rotating Machinery Based on Generative Adversarial Networks. Di dalam: *2018 37th Chinese Control Conference (CCC)*. *2018 37th Chinese Control Conference (CCC)*;. hlm. 6017–6022.

Zhang C, Li J, Zhao Y, Li T, Chen Q, Zhang X, Qiu W. 2021. Problem of data imbalance in building energy load prediction: Concept, influence, and solution. *Applied Energy*. 297:117139.doi:10.1016/j.apenergy.2021.117139.