Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 8, Nomor 2, Agustus 2023, pp 696-704
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Klasifikasi Penyakit Skizofrenia menggunakan Algoritma Logistic Regresion

Khoirun Nisa[1], Sony Kartika Wibisono[2]
[1,2]Program Studi Informatika, Universitas Harapan Bangsa, Purwokerto, Indonesia
Email: khoirunnisa@uhb.ac.id[1], sonykartika@uhb.ac.id[2]

Abstract

The World Health Organization (WHO) reports that 20 million people worldwide are affected by this mental disorder. Therefore, it is necessary to develop an automated model to diagnose patients that can help doctors to start medical treatment early. This research aims to compare machine learning algorithms in the classification of schizophrenia, a mental condition that is often difficult to diagnose, this research is expected to help improve accuracy in the schizophrenia diagnosis process. The research approach involves applying a Logistic Regression model trained with medical and psychological data from schizophrenia patients. The developed model was then tested to evaluate its accuracy in detecting schizophrenia. The accuracy result obtained using the Logistic Regression model was 93.6%.

Keywords*: Schizophrenia, Classification, EEG, Logistic Regression*

Abstrak

World Health Organization (WHO) melaporkan bahwa 20 juta orang di seluruh dunia terkena gangguan mental ini. Oleh karena itu, perlu dikembangkan model otomatis untuk mendiagnosis pasien yang berguna membantu dokter untuk memulai perawatan medis secara dini. Penelitian ini bertujuan untuk membandingkan algoritma machine learning dalam klasifikasi penyakit skizofrenia, sebuah kondisi mental yang kerap kali sulit dalam melakukan diagnosis, penelitian ini diharapkan dapat membantu meningkatkan akurasi dalam proses diagnosis skizofrenia. Pendekatan penelitian ini melibatkan penerapan model Logistic Regression yang dilatih dengan data medis dan psikologis dari pasien skizofrenia. Model yang dikembangkan kemudian diuji untuk mengevaluasi keakuratannya dalam mendeteksi skizofrenia. Hasil akurasi yang diperoleh menggunakan model Logistic Regression sebesar 93.6%.

Kata kunci*: Penyakit Skizofrenia, Klasifikasi, EEG, Logistic Regression*

1. PENDAHULUAN

Skizofernia (SZ) merupakan gangguan mental berat yang mempunyai gejala seperti delusi, halusinasi, bicara tidak teratur, kesulitan berfikir, kurangnya motivasi dan masalah sosio-psikologis lainnya. SZ dapat disembuhkan dengan prognosis dan waktu yang tepat untuk membatu pengobatan yang lebih balik. Oleh karena itu, perlu dikembangkan model otomatis untuk mendiagnosis pasien yang berguna membantu dokter untuk memulai perawata medis secara dini [1]. Sinyal EEG berfungsi untuk melihat aktivitas aliran listrik pada otak. Selain itu akuisis sinyal ini ekonomis, non-invasif dan nonradioaktif karena keuntungan tersebut maka peneliti menggunakan sinyal EEG untuk mendeteksi SZ. Rekaman sinyal EEG berisi data dinamis dalam jumlah besar untuk mempelajari fungsi otak. Biasanya, data dalam jumlah besar ini dinilai dengan inspeksi visual, yang membutuhkan waktu lama, rentan terhadap kesalahan manusia, dan mengurangi keandalan pengambilan keputusan. Hingga saat ini, belum ada cara yang dapat

mengoptimalkan dalam mengidentifikasi skizofrenia dari data EEG secara otomatis, cepat, dan akurat [2].

Deteksi gangguan mental seperti skizofrenia (SZ) dengan menyelidiki aktivitas otak yang direkam melalui sinyal Electroencephalogram (EEG) adalah bidang yang menjanjikan dalam ilmu saraf. Penelitian ini menyajikan konektivitas efektif otak hibrida dan kerangka kerja deep learning untuk deteksi SZ pada sinyal EEG multichannel. Pertama, matriks konektivitas efektif diukur berdasarkan metode Transfer Entropy (TE) yang memperkirakan hubungan sebab-akibat yang terarah dalam hal aliran informasi otak dari 19 channel EEG untuk setiap subjek. Kemudian, elemen konektivitas efektif TE diwakili oleh warna dan membentuk Gambar konektivitas 19 × 19 yang, secara bersamaan, mewakili informasi waktu dan spasial sinyal EEG [3].

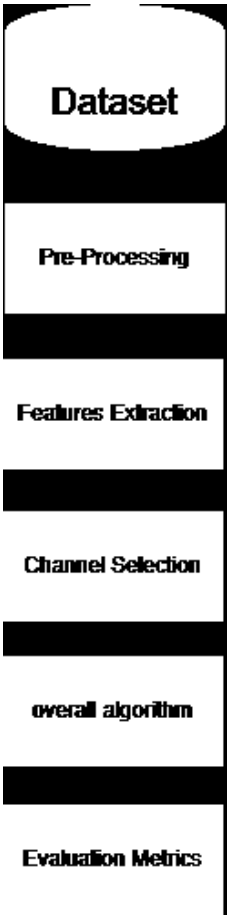
Pendekatan berbasis deskriptor lokal yang diusulkan memperoleh akurasi klasifikasi rata-rata 92,85% [2]. Pengklasifikasi SVM dengan set fitur mendalam yang diperoleh menghasilkan akurasi tertinggi 91,23% [4]. Penggabungan statistic dan fitur pattern memberikan akurasi 92,62% mengingat pengklasifikasi BT untuk dataset yang terdiri dari 28 subjek. Selanjutnya, akurasi meningkat menjadi 99,24% untuk kumpulan data yang lebih besar dengan 81 subjek [5]. Pendekatan hibrida yang menggabungkan teknik robust variational mode decomposition (RVMD) dan optimized extreme learning machine (OELM) classifier untuk mengembangkan sistem pendukung keputusan otomatis dalam mendeteksi SCZ menggunakan sinyal elektroensefalogram (EEG). RVMD digunakan untuk menganalisis dan mensintesis sinyal EEG, sedangkan OELM classifier digunakan untuk klasifikasi NHC dan SCZ. Selain itu, fitur-fitur statistik juga digunakan dalam proses klasifikasi. Akurasi yang dihasilkan menggunakan metode RVMD & OELM : 92,93% dengan enam fitur [6].

Oleh karena itu, penelitian berikutnya melakukan pengembangan untuk mengembangkan skema deteksi skizofrenia otomatis menggunakan data sinyal EEG. CWT digunakan untuk menguraikan sinyal EEG yang difilter menjadi komponen frekuensinya dan fitur statistik dari koefisien CWT dihitung dalam domain waktu. Menerapkan CNN yang dapat secara otomatis mendeteksi kelas yang dimiliki setiap skalogram dan mengklasifikasikannya. Model CNN menghasilkan akurasi yang seimbang pada kasus pertama sebesar 63,7% dan 81,5% pada kasus kedua, pada uji validasi eksternal sebesar 64,5% dan 83,2% [7]. Penelitian ahmad shalbaf melakukan Kombinasi continuous wavelet transform (CWT), transfer learning dengan empat CNN (AlexNet, ResNet-18, VGG-19 dan Inception-v3) dan SVM sebagai pendekatan baru untuk diagnosis otomatis pasien SZ dari sinyal EEG. akurasi sebesar 92,60%±2,29 dicapai untuk arsitektur ResNet-18-SVM pada citra skalogram kombinasi daerah frontal, sentral, pari etal, dan oksipital sinyal EEG [8]. Ekstraksi fitur data EEG dan arsitektur Neural Network menggunakan model klasifikasi otomatis untuk skizofrenia berdasarkan pemetaan spasial-temporal EEG dan LeViT digunakan untuk mengklasifikasikan berdasarkan struktur biara dan transformator. Model tersebut menghasilkan akurasi rata-rata subjek dependen sebesar 85,04%[9]. Hasil percobaan kinerja dengan menggunakan (a) Ada-Boost Classifier dengan

elektroda Tengah adalah 85,71%, (b) Gradient-Boosting Classifier dengan elektroda Parietal Occipital adalah 80,00%, (c) Decision-Tree Classifier & Ada-Boost Classifier dengan Frontal- Elektroda prefrontal adalah 76,67%, dan (d) XGBoost Classifier dengan kombinasi elektroda Central, Parietal-Occipital, dan Frontal-Prefrontal adalah 78,75% [10]. Berdasarkan uraian diatas, maka penelitian ini akan berfokus pada penerapan metode Linier Regression dalam mengklasifikasi penyakit skizofrenia.

2. METODOLOGI PENELITIAN

Penelitian ini akan menggunakan beberapa tahapan dan alur dalam proses penyelesaian masalahan yang ditunjukkan pada Gambar 1.



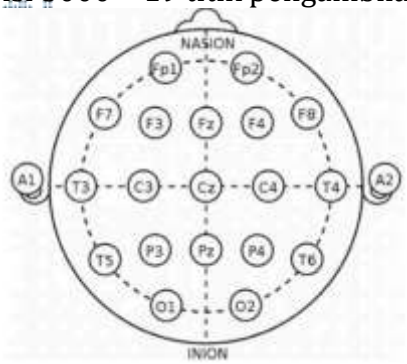
Gambar 1. Diagram Alur Penelitian

2.1. Dataset

Tahap awal pada penelitian ini adalah pengumpulan dataset yang diambil dari Kaggle. Dataset terdiri dari 22 kontrol dan 36 pasien skizofrenia telah digabungkan dengan 10 kontrol dan 13 pasien. Sinyal EEG direkam dalam keadaan istirahat selama mata tertutup dan diukur berdasarkan standar 10-20 system pada Gambar 2. Semua 19 saluran EEG terdiri dari Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2 dan elektroda FCz dianggap

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 8, Nomor 2, Agustus 2023, pp 696-704
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

sebagai elektroda referensi [11]. Sinyal dari rekaman EEG dibagi menjadi 15 detik segmen, masing-masing berisi 5000 × 19 titik pengambilan sampel.



Gambar 2. Standar 10 -20 System

2.2. Pre – Processing

Sinyal EEG diperoleh dari kulit kepala yang sangat dipengaruhi oleh artefak fisiologis seperti mata yang berkedip, gerakan otot, dan lain-lain, sehingga representasi akurat masih kurang dari sinyal yang diperoleh dari otak. Maka perlu memisahkan sinyal otak yang penting dan aktivitas saraf acak yang diamati selama rekaman EEG. Biasanya, frekuensi yang lebih tinggi dari 50 Hz dianggap sebagai noise. Penelitian ini telah menggunakan filter Butterworth band-pass fase-nol orde 2 untuk prapemrosesan sinyal EEG. Respons besaran filter Butterworth orde 2 diberikan oleh Persamaan (1).

$$|H(\mathrm{jw})^2| = \frac{G_o^2}{1+(\frac{jw}{jw_c})^{2M}} \quad (1)$$

Dimana M merupakan order of filter, w adalah passband, w_c representasi dari batas akhir frekuensi dan G_0 merupakan frekuensi nol. Filter Butterworth membatasi nilai frekuensi sinyal dalam kisaran [0,5, 50] Hz dan menghasilkan sinyal EEG yang lebih bersih. Sebelum memasukkan sinyal ini ke CNN untuk pelatihan, sinyal tersebut diskalakan dalam kisaran 0 hingga 1 menggunakan normalisasi z-score.

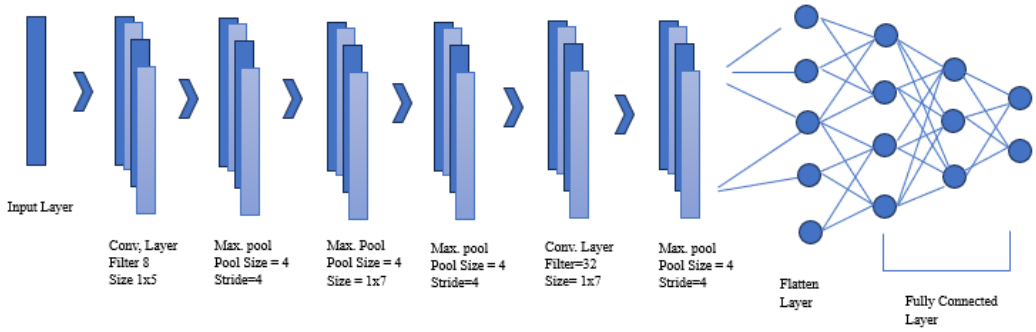
2.3. Feature Extraction

Sebagian besar pengklasifikasi pembelajaran mesin konvensional, seperti SVM, regresi logistik, pengklasifikasi penguat gradien, dll., memerlukan fitur buatan tangan yang diekstrak dari sinyal sebelumnya. Fitur-fitur ini biasanya didasarkan pada metrik statistik seperti varians, power, dll [12]. Kumpulan fitur yang dipilih selanjutnya dikurangi menggunakan algoritma pemilihan fitur karena tidak semua fitur sama pentingnya atau memiliki efek positif pada akurasi klasifikasi. Di sisi lain, CNN mengekstrak dan mempelajari fitur secara otomatis tanpa memerlukan pakar domain atau mekanisme pemilihan fitur eksternal. Oleh karena itu, alih-alih mengekstrak fitur secara manual dan kemudian mengidentifikasi fitur yang diskriminatif di antara fitur-fitur tersebut, penelitian ini menggunakan CNN sebagai ekstraksi fitur.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 8, Nomor 2, Agustus 2023, pp 696-704
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

2.4. Channel Selection

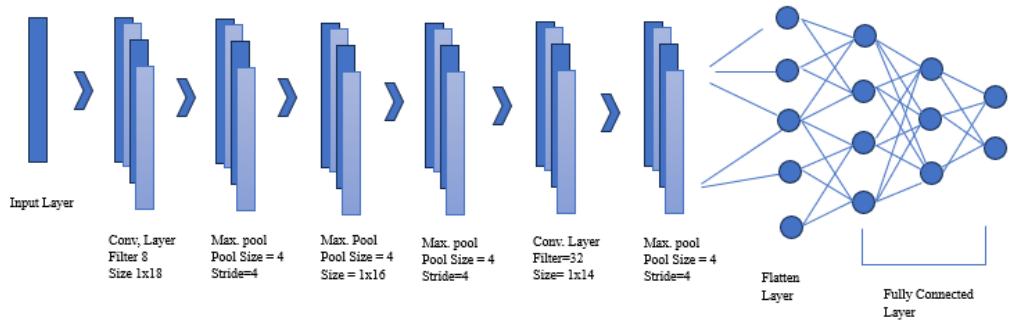
Beberapa percobaan dilakukan untuk menemukan saluran yang paling relevan. Sebagai percobaan pertama, kami menggabungkan data dari semua 19 saluran dan membangun Dataset Full_Channels di mana model CNN yang ditunjukkan pada Gambar 3 dilatih. Parameter network disesuaikan dengan mengamati pengaruhnya terhadap akurasi klasifikasi. Jumlah kernel yang berbeda (yaitu, 4, 8, 12, 16 dan 20) dengan ukuran yang berbeda dicoba dan nilai terbaik dipilih dengan menggunakan *cross-validation*.



Gambar 3. Arsitektur CNN atau percobaan dengan seluruh saluran

Demikian pula, laju pembelajaran divariasikan sebagai 0,01, 0,001 dan 0,0001. Akurasi maksimum dicapai dengan menggunakan tingkat pembelajaran 0,001 dengan *Adam Optimizer*. Fungsi RELU digunakan sebagai fungsi aktivasi bersama dengan *mean squared error* sebagai *loss function*. Lebih lanjut, kami juga mengamati efek dari daerah frontal, temporal, parietal, oksipital, dan pusat otak pada deteksi SZ. Hanya ukuran kernel dari lapisan konvolusi 1 dan lapisan konvolusi 2 yang diubah menjadi 7 dan 10.

Setelah itu, saluran ketiga digabungkan, dan proses dilanjutkan sampai akurasi yang diinginkan tercapai. sedangkan arsitektur CNN yang digunakan diberikan pada Gambar 4 Jumlah kernel disesuaikan dengan meningkatkan jumlahnya mulai dari 3 dan ukuran kernel disesuaikan dengan mempertimbangkan nilai pada rentang 3 hingga 20.



Gambar 4. Arsitektur Seleksi Channel menggunakan CNN

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 8, Nomor 2, Agustus 2023, pp 696-704
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Ekstraksi menggunakan arsitektur CNN, tugas selanjutnya adalah mengklasifikasikan sinyal EEG. Seperti yang telah dibahas sebelumnya, arsitektur CNN secara logis dapat dibagi menjadi dua bagian - konvolusi + *pooling layers* yang bertanggung jawab untuk ekstraksi fitur, dan lapisan yang sepenuhnya terhubung yang bertanggung jawab untuk klasifikasi menggunakan regresi logistik.

2.5. Overall Algorithm

Keseluruhan algoritma terdiri dari pemrosesan sinyal, normalisasi, ekstraksi fitur dan klasifikasi. Pseudocode dari model yang diusulkan diberikan pada tabel 1.

Tabel 1. Pseudocode Algoritma Usulan

Pseudocode Algoritma Usulan
1. Input : Poin sampel pelatihan S_{train} Output Label O_{train} dan tes sampel S_{test}
2. Output: prediksi label-label di tes sampel P_{test}
3. Applay 2nd order Butterworth filter di fitur S_{train} and S_{test}
4. Apply Z-score normalisasi di S_{train} dan S_{test}
5. f_{train} = memilih fitur setelah melewati proses S_{train} dan O_{train} to M
6. Set S'_{train} = training dataset menggunakan fitur f_{train} dan S'_{test} setelah itu melewati tes dataset ke M'
7. Output P_{test} menggunakan CL (S'_{train}, O_{train}, S'_{test})

Fitur-fitur yang diekstraksi, setelah diratakan dimasukkan ke dalam model Regresi Logistik [41]. Regresi logistik adalah metode analisis statistik yang memodelkan variabel target dikotomis dan memperkirakan probabilitasnya berdasarkan fitur individu. Metode ini banyak digunakan dalam masalah di mana hasilnya bersifat kategorikal seperti halnya data Skizofrenia yang digunakan dalam penelitian ini.

2.6. Evaluation Metrics

Berbagai metrik telah digunakan dalam literatur untuk evaluasi klasifikasi, seperti akurasi, sensitivitas, dll. Metrik-metrik ini menggunakan label yang diprediksi dan label yang sebenarnya untuk menilai kinerja model. Akurasi adalah ukuran yang paling umum digunakan untuk membandingkan antar model karena mudah untuk diinterpretasikan. Namun, ukuran ini juga rentan terhadap bias ketika ukuran kelas tidak sama.

Akurasi memberikan persentase total prediksi yang benar dari semua label yang diprediksi. Akurasi dihitung dengan menggunakan persamaan 2.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

Dimana TP, TN, FN dan FN merupakan *true positives, true negatives, false positives, dan false negatives*.

3. HASIL DAN PEMBAHASAN

Hasil akurasi menunjukan bahwa Logistik regresi sebesar 93,6 % dan k-NN sebesar 65.77%. terdapat pada tabel 1.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 8, Nomor 2, Agustus 2023, pp 696-704
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Tabel 2. Hasil Akurasi

Algoritma	Akurasi
K-NN	65,77%
Linier Regresion	93,6%

Ditunjukkan dengan pengolahan data pada Gambar 5 untuk KNN dan Gambar 6 untuk Logistic Regression.

```
from sklearn.neighbors import KNeighborsClassifier
clf = KNeighborsClassifier()
clf.fit(X_train_norm, Y_train_norm)
y_pred_knn = clf.predict(X_test_norm)
acc_knn = round(clf.score(X_train_norm, Y_train_norm) * 100, 2)
print (str(acc_knn)+'%')
```

65.77%

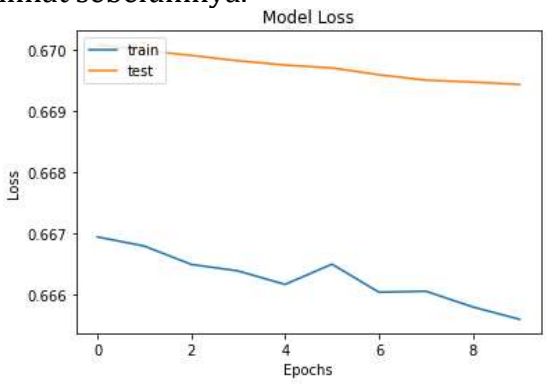
Gambar 5. Linier Regression

```
from sklearn.linear_model import LogisticRegression
clf = LogisticRegression()
clf.fit(X_train_norm, Y_train_norm)
y_pred_log_reg = clf.predict(X_test_norm)
acc_log_reg = round(clf.score(X_train_norm, Y_train_norm) * 100, 2)
print (str(acc_log_reg) + ' %')
```

93.6 %

Gambar 6. K-NN

Dalam evaluasi performa model terhadap data tes dengan epoch 10, ditemukan bahwa nilai kerugian (loss) model terlihat pada Gambar 3 adalah 0,67. Ini mengindikasikan bahwa, rata-rata, selisih antara prediksi model dan nilai sebenarnya adalah 0,697. Nilai ini memberikan Gambaran tentang sejauh mana kesalahan prediksi yang dihasilkan model ketika dihadapkan dengan data baru yang belum pernah dilihat sebelumnya.



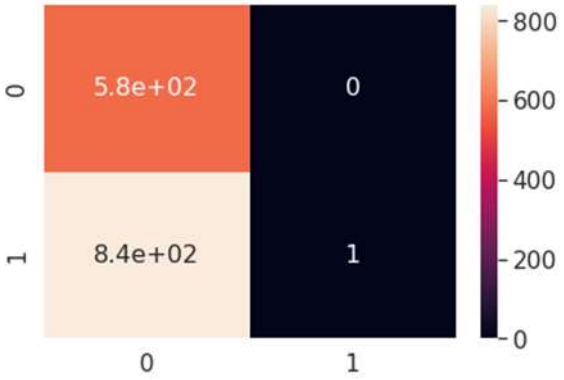
Gambar 7. Nilai Loss

Selanjutnya, terkait dengan akurasi, model mampu memprediksi label kelas dengan benar dalam sekitar 40% dari kesempatan pada data tes. Mengingat konteks dan kesulitan tugas klasifikasi yang dihadapi, ini dapat diartikan sebagai indikasi awal dari performa model. Namun, perlu dicatat bahwa angka ini harus ditafsirkan dengan hati-hati. Sebagai contoh, jika data sangat tidak seimbang, maka

akurasi yang relatif rendah ini mungkin lebih mencerminkan performa model pada kelas minoritas daripada kemampuannya untuk memprediksi secara umum. Oleh karena itu, penelitian lebih lanjut dan penyesuaian model mungkin diperlukan untuk meningkatkan performa ini.

Hasil klasifikasi model pada Gambar 4 menunjukkan bahwa ada 578 True Positives untuk skizofrenia. Ini berarti bahwa dalam 578 kasus, model kita berhasil memprediksi dengan benar bahwa subjek memiliki skizofrenia. Di sisi lain, model tidak menghasilkan False Positives untuk skizofrenia, yang berarti tidak ada subjek yang sebenarnya sehat namun diprediksi oleh model sebagai memiliki skizofrenia. Namun, model memiliki 840 False Negatives untuk subjek yang sehat. Ini berarti ada 840 subjek yang sebenarnya sehat, namun model memprediksi mereka memiliki skizofrenia. Ini menunjukkan bahwa model kita memiliki kecenderungan untuk memprediksi subjek sebagai memiliki skizofrenia, meski sebenarnya mereka sehat.

Terakhir, ada 1 True Negative untuk subjek yang sehat. Ini berarti hanya ada satu kasus di mana model berhasil memprediksi subjek yang sehat dengan benar. Ini menunjukkan bahwa performa model kita dalam mengidentifikasi subjek yang sehat sangat tinggi.



Gambar 8. Matrix Confusion

4. SIMPULAN

Hasil diatas menunjukkan bahwa model memiliki bias yang kuat terhadap prediksi skizofrenia dan mengalami kesulitan dalam membedakan subjek yang sehat. Penyesuaian lebih lanjut pada model mungkin diperlukan untuk memperbaiki ini dan mencapai kinerja yang lebih seimbang antara memprediksi subjek yang memiliki skizofrenia dan subjek yang sehat.

DAFTAR PUSTAKA

[1] F. Hassan, S. F. Hussain, and S. M. Qaisar, "Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques," *Inf. Fusion*, vol. 92, no. December 2022, pp. 466–478, 2023, doi: 10.1016/j.inffus.2022.12.019.

[2] T. S. Kumar, K. N. V. P. S. Rajesh, S. Maheswari, V. Kanhangad, and U. R. Acharya, "Automated Schizophrenia detection using local descriptors with EEG signals," *Eng.*

Appl. Artif. Intell., vol. 117, no. April 2022, p. 105602, 2023, doi: 10.1016/j.engappai.2022.105602.

[3] S. Bagherzadeh, M. S. Shahabi, and A. Shalbaf, "Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal," *Comput. Biol. Med.*, vol. 146, no. April, p. 105570, 2022, doi: 10.1016/j.compbiomed.2022.105570.

[4] S. Siuly, Y. Guo, O. F. Alcin, Y. Li, P. Wen, and H. Wang, "Exploring deep residual network based features for automatic schizophrenia detection from EEG," *Phys. Eng. Sci. Med.*, 2023, doi: 10.1007/s13246-023-01225-8.

[5] M. Agarwal and A. Singhal, "Fusion of pattern-based and statistical features for Schizophrenia detection from EEG signals," *Med. Eng. Phys.*, vol. 112, no. December 2022, 2023, doi: 10.1016/j.medengphy.2023.103949.

[6] S. K. Khare and V. Bajaj, "A hybrid decision support system for automatic detection of Schizophrenia using EEG signals," *Comput. Biol. Med.*, vol. 141, no. May 2021, p. 105028, 2022, doi: 10.1016/j.compbiomed.2021.105028.

[7] A. I. Korda, E. Ventouras, P. Asvestas, M. Toumaian, G. K. Matsopoulos, and N. Smyrnis, "Convolutional neural network propagation on electroencephalographic scalograms for detection of schizophrenia," *Clin. Neurophysiol.*, vol. 139, pp. 90–105, 2022, doi: 10.1016/j.clinph.2022.04.010.

[8] A. Shalbaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," *Phys. Eng. Sci. Med.*, vol. 43, no. 4, pp. 1229–1239, 2020, doi: 10.1007/s13246-020-00925-9.

[9] B. Li, J. Wang, Z. Guo, and Y. Li, "Automatic detection of schizophrenia based on spatial – temporal feature mapping and LeViT with EEG signals," *Expert Syst. Appl.*, vol. 224, no. April 2022, 2023.

[10] P. K. Sahu, "Artificial intelligence system for verification of schizophrenia via theta-EEG rhythm," *Biomed. Signal Process. Control*, vol. 81, no. December 2022, p. 104485, 2023, doi: 10.1016/j.bspc.2022.104485.

[11] V. J. ; S. L. O. V. R. E. J. C. K. H. C. N. A. U. R. Acharya;, "Automated detection of schizophrenia using nonlinear signal processing methods," *Artif Intell Med*, vol. 20, 2019, doi: doi: 10.1016/j.artmed.2019.07.006.

[12] S. Mian Qaisar and S. Fawad Hussain, "Arrhythmia Diagnosis by Using Level-Crossing ECG Sampling and Sub-Bands Features Extraction for Mobile Healthcare," *Sensors (Basel).*, vol. 20, no. 8, 2020, doi: 10.3390/s20082252.