

Nama : Destya Rosa Mardiana
NIM : 22/496395/PA/21338

Tugas Sinyal 1 Dimensi Pengenalan Pola

Klasifikasi Aritmia pada EKG menggunakan CNN

Pada tugas ini dilakukan untuk memprediksi perbedaan bentuk sinyal aritmia pada EKG. Proyek ini mengklasifikasikan bentuk sinyal EKG dalam 5 kelas. Pengelompokan kelasnya diantaranya :

- N : Non-ecotic beats (normal beat) - [0]
- S : Supraventricular ectopic beats - [1]
- V : Ventricular ectopic beats - [2]
- F : Fusion Beats - [3]
- Q : Unknown Beats - [4]

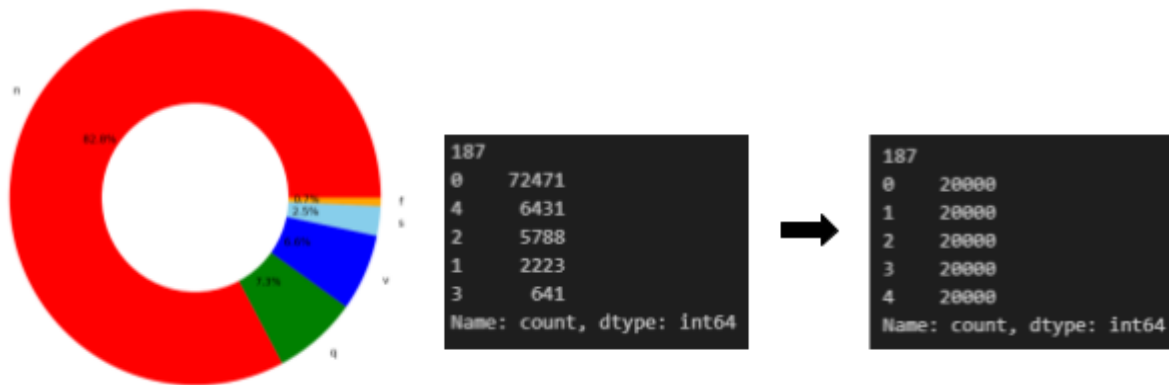
```
Number of Sample(train): 87554
Number of categories(train): 5
Class Distribution(train):
187
0      72471
4       6431
2       5788
1       2223
3        641
Name: count, dtype: int64
```

```
Number of Sample(test): 21892
Number of categories(test): 5
Class Distribution(test):
187
0.0     18118
4.0      1608
2.0      1448
1.0       556
3.0       162
Name: count, dtype: int64
```

Dataset yang digunakan adalah MIT-BIH (<https://physionet.org/content/mitdb/1.0.0/>). MIT-BIH Arrhythmia Dataset berisi 48 rekaman ECG yang diambil selama setengah jam. Sebanyak 23 rekaman dipilih secara acak dari kumpulan 4000 rekaman EKG 24 jam yang dikumpulkan dari pasien rawat inap(sekitar 60%) dan rawat jalan (sekitar 40%) di Beth Israel Hospital Boston. Sedangkan 25 rekaman lainnya dipilih secara khusus dari kumpulan yang sama untuk mencakup aritmia yang lebih jarang tetapi memiliki signifikansi klinis yang penting, karena jika dipilih secara acak, jenis aritmia ini mungkin tidak akan terwakili dengan baik.

Preprocessing

Pada proses preprocessing dalam proses ini digunakan untuk membersihkan, menormalkan, dan mengubah data agar lebih optimal untuk dimasukkan ke dalam model machine learning. Perbedaan jumlah sampel dalam beberapa kelas yang tidak seimbang perlu dilakukan resampling. Pada tahap resampling dilakukan untuk meyeimbangkan jumlah komposisi pada tiap kelasnya.

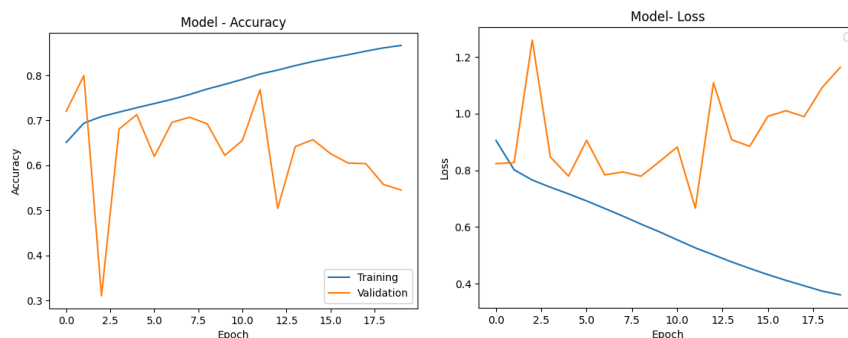


Metodologi

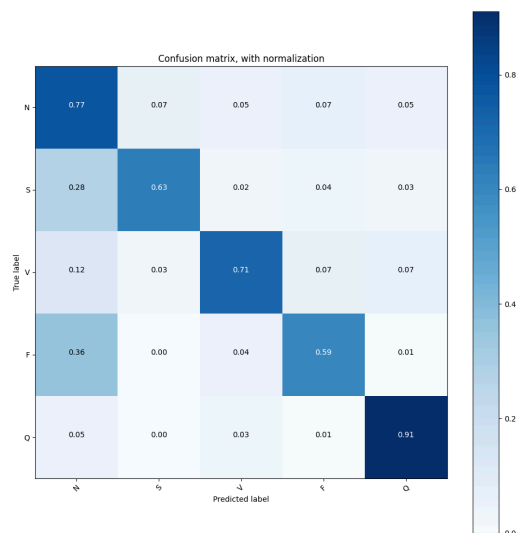
Proyek ini menggunakan model CNN dengan tiga lapis convolution berturut-turut dengan Batch Normalization untuk meningkatkan stabilitas dan performa model. Pada lapis convolution pertama menggunakan 64 filter dengan ukuran kernel 6. Kemudian untuk lapisan convolution 2 dan 3 masih menggunakan 64 filter dengan kernel yang lebih kecil yaitu 3. Pada proyek ini menggunakan **optimizer adam** untuk mempercepat konvergensi, **loss function yang digunakan adalah categorical cross entropy** karena ini adalah klasifikasi multi-kelas. Training ini dilakukan sebanyak **40 epoch** dengan jumlah **batch size 32**, dan **learning rate : 0.001**.

Hasil dan analisis

Accuracy: 76.78%



Berdasarkan grafik menunjukkan bahwa akurasi data training meningkat secara stabil, tetapi akurasi validasi menunjukkan fluktuasi yang signifikan. Ini mengindikasikan bahwa model mengalami ketidakstabilan saat validasi, bisa jadi tanda overfitting. Grafik Loss memperlihatkan tren loss yang menurun pada training set, sedangkan loss validasi mengalami fluktuasi yang cukup besar. Hal ini kembali mengindikasikan bahwa model mungkin tidak generalisasi dengan baik terhadap data baru.



Berdasarkan confusion matrix diatas menunjukkan bahwa kelas dengan performa prediksi terbaik adalah **kelas Q (Unknown Beats)** dengan akurasi mencapai **91%**. Kemudian kelas **N (Non-ecotic beats (normal beat))** dan **V (Ventricular ectopic beats)** memiliki akurasi prediksi yang cukup baik juga yaitu berada di angka **77%** dan **71%**. Sedangkan untuk kelas **S (Supraventricular ectopic beats)** dan **F (Fusion Beats)** memiliki akurasi yang rendah dengan banyak instance yang diklasifikasikan salah. Misalnya, 36% dari kelas F salah diklasifikasikan dengan N, yang menunjukkan bahwa model kesulitan membedakan kedua kelas ini. Akurasi total dari model ini adalah **76.78%** yang bisa digolongkan cukup baik, tetapi masih perlu dilakukan perbaikan. Perbaikan yang mungkin bisa dilakukan adalah melakukan data augmentasi, yang memungkinkan untuk menambah lebih banyak data untuk kelas yang kurang terwakili untuk meningkatkan keseimbangan dataset. Selain itu perlu juga dilakukan kombinasi learning, batch size, dan optimizer untuk mendapatkan model yang stabil.

Source Code Lengkap :

<https://github.com/Destyarosa12/Arrhythmia-on-ECG-Classification-using-CNN/tree/main>